

Spring 2017

A Human Factors Approach for Identifying Latent Failures in Healthcare Settings

Tara N. Cohen

Follow this and additional works at: <https://commons.erau.edu/edt>



Part of the [Health and Medical Administration Commons](#)

Scholarly Commons Citation

Cohen, Tara N., "A Human Factors Approach for Identifying Latent Failures in Healthcare Settings" (2017). *Dissertations and Theses*. 290.

<https://commons.erau.edu/edt/290>

This Dissertation - Open Access is brought to you for free and open access by Scholarly Commons. It has been accepted for inclusion in Dissertations and Theses by an authorized administrator of Scholarly Commons. For more information, please contact commons@erau.edu.

A HUMAN FACTORS APPROACH FOR IDENTIFYING LATENT FAILURES IN
HEALTHCARE SETTINGS

By

TARA N. COHEN

B.A., Psychology, University of Southern California, 2013

M.S., Human Factors, Embry-Riddle Aeronautical University, 2015

A Dissertation submitted to the

Department of Human Factors

In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy in Human Factors

Embry-Riddle Aeronautical University

Daytona Beach, Florida

Spring 2017

SIGNATURE PAGE

A HUMAN FACTORS APPROACH FOR IDENTIFYING LATENT FAILURES IN HEALTHCARE SETTINGS

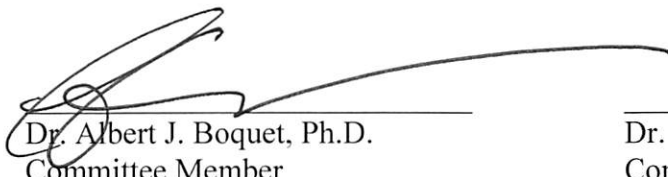
By

Tara N. Cohen

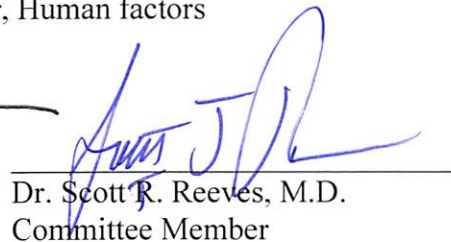
This dissertation was prepared under the direction of the candidate's Dissertation Committee Chair, Dr. Scott Shappell and has been approved by the members of the dissertation committee. It was submitted to the College of Arts and Sciences and was accepted in partial fulfillment of the requirements for the Degree of Doctor of Philosophy in Human Factors



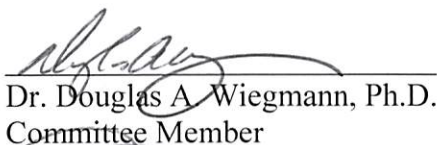
Dr. Scott A. Shappell, Ph.D.
Committee Chair
Department Chair, Human factors



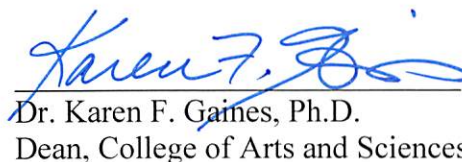
Dr. Albert J. Boquet, Ph.D.
Committee Member



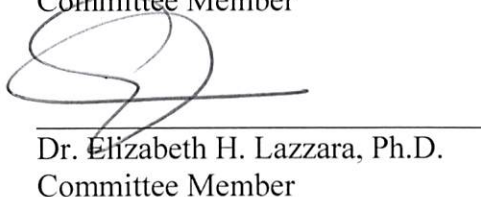
Dr. Scott R. Reeves, M.D.
Committee Member



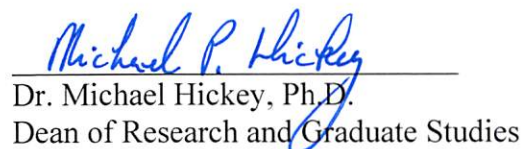
Dr. Douglas A. Wiegmann, Ph.D.
Committee Member



Dr. Karen F. Gaines, Ph.D.
Dean, College of Arts and Sciences



Dr. Elizabeth H. Lazzara, Ph.D.
Committee Member



Dr. Michael Hickey, Ph.D.
Dean of Research and Graduate Studies

01-24-17

Date

ACKNOWLEDGMENTS

While this process has certainly been challenging, several individuals have supported, encouraged, and guided me during this journey. I would first like to express my deepest appreciation to my dissertation advisor and chair Professor Scott Shappell, who has taught me more than I can list in one section. Not only has he mentored me in terms of academic writing, professional presentations, and teaching, but working with him has shaped who I am. He continues to inspire me to work hard and take on any project with confidence. I would also like to extend my sincerest gratitude for the other members of my committee – Dr. Albert Boquet, for providing statistical instruction and guidance before, during, and after the completion of my dissertation, and for being both a great friend and mentor, Dr. Doug Wiegmann, for his support not only throughout this project but on manuscripts and industry projects outside of academia, Dr. Elizabeth Lazzara, for her input, direction, and advice, especially with regards to “connecting the dots”, and Dr. Scott Reeves, for his continued support on this project amongst others as well as his medical expertise. Without these individuals, as well as many others in the Human Factors Department at Embry-Riddle, this experience would be vastly different.

I would also like to thank my graduate school colleges and close friends Emily Anania, Victoria Lew, and Austin Yazel for not only coding all of my data, but for supporting and encouraging me along the way. I would especially like to thank Tracy Litzinger for not only being a great teacher, mentor, and best friend, but also for uplifting me during what felt like the toughest times in my life. I know I could not have done this without her, and I am eternally grateful for everything she has done. A very special

thanks to my boyfriend Josh, for his patience, care, and encouragement while tolerating me during this process.

Finally, my ultimate thanks must be extended to my family – my sister Ashley for her love and support, my mother Ronni for always listening, building me up when I needed encouragement, and teaching me to be the very best version of myself, and for my father Jeff for pushing me to achieve more than I ever thought I could. None of this would be possible without you.

TABLE OF CONTENTS

Signature Page	ii
Acknowledgments.....	iii
List of Tables	vii
List of Figures	viii
Definitions or Nomenclature.....	x
Abstract.....	xi
Chapter 1: Introduction	13
Chapter 2: Literature Review	21
Reason's Swiss cheese model of accident causation.....	21
The Human Factors Analysis and Classification System (HFACS).....	30
Research Questions	69
Chapter 3: Study I (CVOR - Academic hospital)	76
Chapter 4: Study II (CVOR-Non-Academic hospital).....	85
Chapter 5: Study III (Trauma Center).....	93
Results	97
Chapter 6: Exploratory Analyses	102
Chapter 7: Discussion	105
Bibliography	125
Appendices.....	141

Appendix A: Related publication - Cohen et al., 2015	141
Appendix B: Related publication - Ergai et al., 2016	142
Appendix C: Related publication - Cohen et al., 2016.....	143
Appendix D: Related publication - Cohen et al., 2016	144

LIST OF TABLES

Table 1. Description of the HFACS Categories.....	30
Table 2. Measures of reliability and associated values.....	57
Table 3. Articles that reported reliability indices of HFACS	58
Table 4. Comparison of HFACS reliability-specific studies reported in the literature	60
Table 5. Overall α for each HFACS Tier and Category; Inter-rater Reliability	66
Table 6. Average α for each HFACS Tier and Category; Intra-rater reliability.....	67
Table 7. Data-set I methodology comparisons at the tier and category level	83
Table 8. Data-set II methodology comparisons at the tier and category levels	89
Table 9. Data-set III methodology comparisons at the tier and category level	100

LIST OF FIGURES

Figure 1. Heinrich's Triangle (adapted from Heinrich, 1931)	16
Figure 2. SHELL Model (adapted from Hawkins, 1993)	23
Figure 3. Domino Theory (adapted from Heinrich, 1931).....	23
Figure 4. Reason's Swiss cheese model of accident causation (adapted from Reason, 1990)	26
Figure 5. The Human Factors Analysis and Classification System (HFACS)	32
Figure 6. X-ray CT (from Drew, Vo & Wolfe, 2013)	38
Figure 7. Left to right - Pancuronium (paralytic), Metoclopramide (antacid).....	42
Figure 8. Layout of the Cardiovascular Operating Room (CVOR).....	77
Figure 9. Data-set I line diagram (unanimous method)	81
Figure 10. Data-set I line diagram (majority method)	82
Figure 11. Data-set I line diagram (reconciled method)	82
Figure 12. Data-set I unanimous vs. majority vs. reconciled preconditions for unsafe acts	84
Figure 13. Data-set II line diagram (unanimous method).....	87
Figure 14. Data-set II line diagram (majority method).....	88
Figure 15. Data-set II line diagram (reconciled method).....	88
Figure 16. Data-set II unanimous vs. majority vs. reconciled preconditions for unsafe acts	90
Figure 17. CVOR academic and non-academic comparison of reconciled findings for preconditions for unsafe acts	91
Figure 18. Level II trauma center (Study III)	95

Figure 19. Data-set III line diagram (unanimous method)	98
Figure 20. Data-set III line diagram (majority method)	99
Figure 21. Data-set III line diagram (reconciled method)	99
Figure 22. Data-set III unanimous vs. majority vs. reconciled preconditions for unsafe acts	101
Figure 23. Comparison of preconditions for unsafe acts categories by hospital	104
Figure 24. (a) image of a typical cardiovascular operating room (b) image of a typical trauma resuscitation room.....	116
Figure 25. The HFIX Cube (Adapted from Shappell & Wiegmann, 2006)	123

DEFINITIONS OR NOMENCLATURE

Active Failure	Faults that are directly linked to an accident.
CVOR	Cardiovascular Operating Room
Fleiss' Kappa	A test for inter-rater reliability and measures nominal data on a scale of 0.0-1.0
HFACS	Human Factors Analysis and Classification System
HFACS Categories	Nineteen subcategories included within the four HFACS Tiers. The HFACS categories are also referred to as causal categories or causal factors. There are three categories at the organizational influences tier, four at the unsafe supervision tier, seven at the preconditions for unsafe acts tier, and five at the unsafe acts tier
HFACS Tiers	Four overarching levels of the HFACS taxonomy as described by James Reason. Included within these four levels are (1) organizational influences, (2) unsafe supervision, (3) preconditions for unsafe acts (4) unsafe acts
Latent Failure	Background elements which may eventually lead to an adverse event or unsafe act. These failures act as contributory factors that often occur "behind the scenes".
Majority Agreement	The percentage of events in which at least two raters agreed (majority) on the appropriate HFACS code
Pairwise Agreement	Agreement between all possible coder dyads (i.e. rater 1 vs rater 2, rater 1 vs rater 3, and rater 2 vs rater 3).
Reconciled Method	A subsequent analysis of majority agreement that includes those events that were originally disagreed upon, and later reconciled using consensus coding (i.e. discussion between the three coders to determine the appropriate allocation of the event into the HFACS framework)
SHEL Model	Software, Hardware, Environment, Liveware Model
Unanimous Agreement	The percentage of events in which all three coders concurred on the HFACS classification of the event

ABSTRACT

INTRODUCTION: The purpose of the current research was to assess the utility of the Human Factors Analysis and Classification System (HFACS), a tool that has historically been used reactively to look at accidents and incidents, for classifying observational data from various healthcare venues.

METHOD: Three studies are presented to investigate the reliability of HFACS for classifying observational data. In Study I, HFACS was applied to observational human factors data collected from the cardiovascular operating room (CVOR) at an academic medical university. Three trained analysts categorized the data using HFACS and several approaches were used to evaluate its reliability during the categorization task. The same method was repeated for Study II, which utilized CVOR data collected from a non-academic hospital. To investigate the ability of HFACS for differentiating between hospitals, the data from the academic and non-academic hospitals were compared. Finally, to explore the utility of HFACS in another venue, Study III employed the same approach as Study I and II however, observational data from a trauma center was utilized.

RESULTS: Results of the three studies revealed that the framework was substantially reliable ($k=0.635$ (95% CI, .611-.659), $p = 0.000$; $k =0.642$ (95% CI, .633-.652), $p = 0.000$; $k=0.680$ (95% CI, .662 to .698), $p = 0.000$) for classifying observational healthcare data. In all three data sets, preconditions for unsafe acts were the most common area of systemic weakness. However, differences in the distributions of these categories did exist when data-sets were compared.

CONCLUSION: This study is a first step in establishing the reliability of the HFACS framework as a tool for classifying observational human factors data. As HFACS appears

to be a reliable observation tool, findings associated with its use could help to identify where errors and adverse events are likely to occur. Therefore, the proactive identification of human factors issues associated with patient harm represents the next step in the evolution of patient safety. Predictably, hospital administrators could put in place targeted interventions to help mitigate human factors issues before they manifest and become harmful events in the future.

A Human Factors Approach to Identifying Latent Failures in Healthcare Settings

CHAPTER 1: INTRODUCTION

While humans are arguably the most intelligent beings in the world, capable of adapting to our environment, using creative thought, and being cognitively aware, none of us are immune to the several ways in which our bodies can weaken or fail. It is incredibly rare that any of us will live well into old age without experiencing any health detriments. As such, the healthcare industry is important to all of us.

The Center for Disease Control (CDC) has reported that chronic diseases are the leading cause of death and disability in the United States, causing 70% of deaths each year (Center for Disease Control, 2013). To put this in perspective, in 2008, 107 million Americans (nearly 50% of adults age 18 or older) had at least one the following chronic illnesses: cardiovascular disease, arthritis, diabetes, asthma, cancer, and/or chronic obstructive pulmonary disease (Heron et al., 2009). Of these, the leading cause of death in the United States is cardiovascular disease, as over 19,000,000 individuals have succumbed to this disease in 2013 (CDC, 2013).

Although our bodies are built to adapt to most chronic illness, without more invasive treatment, many individuals may still die from these diseases. In 2010, the total number of inpatient surgical procedures performed was 51.4 million, the most common being cardiac procedures (CDC, 2015).

While healthcare in the U.S. is among the best in the world, the delivery of that care is an overwhelming task. Medical professionals are faced with a number of elements while trying to provide optimal care. To name just a few, surgical teams: (1) are constantly faced with the integration of progressively complicated technological devices,

(2) need to communicate and coordinate among various team members within the surgical suite, (3) have to problem solve on the spot for complex cases where there are unforeseen patient challenges, and (4) manage cost and time limitations mandated by various organizations.

While surgery is often the last resort for most patients due to pain, cost and time required for healing, we would like to believe that the surgery itself would not necessarily contribute to complications downstream. Unfortunately, that may not be the case, as the landmark report, *To Err is Human*, conducted by the Institute of Medicine 15 years ago, reported that as many as 98,000 people die in hospitals each year due to preventable medical errors (IOM, 2000). Indeed, experts now say that even that Figure may be too low. More recently, James (2013) has revised the IOM estimate and predicts that the number of deaths resulting from preventable harm in hospitals may be greater than 400,000 per year.

In the healthcare industry, patient safety has been assessed largely by medical outcomes. In other words, many healthcare professionals focus on a single question: did the patient recover or did the patient suffer from complications (perhaps even death) after a given procedure? Unfortunately, this simplistic approach fails to take into account the process itself and does not necessarily shed light on any system-induced failures that may have occurred throughout the procedure.

In his book *Safer Complex Industrial Environments: A Human Factors Approach*, Hollnagel (2010) explains that most safety investigations use either a single- or a multiple- cause philosophy when searching for the genesis of accidents. In the single cause model, individuals focus on root cause analysis and believe that there is a single

cause for any outcome. This philosophy purports that if one acts upon the single cause, the outcome can be prevented from occurring again. While this method is tempting, it is certainly over-simplified. The second approach, multiple-cause philosophy, focuses on the belief that an outcome may be the result of a combination of factors and a “root cause” can exist for each. While this approach is much more comprehensive, it can be unwieldy, given the combination of factors that influence behavior.

While arguments regarding the veracity of the two philosophies continue to be debated, what is common between both approaches is this idea of a “root cause”. Not surprisingly, most of the accident prevention programs in healthcare tend to focus on investigating these “root” causes associated with so-called “sentinel events”, or events that reach a patient and result in either death, permanent harm, or severe temporary harm where intervention is required to sustain life (The Joint Commission, 2016).

Though sentinel events are arguably the largest threat to patient safety and the reputation of the healthcare industry, some would argue that this is not where the focus should be. For example, in Herbert Heinrich’s (1931) book, *Industrial Accident Prevention, A Scientific Approach*, he reported that “it is estimated that in a unit group of 330 accidents of the same kind and involving the same person, 300 result in no injuries, 29 in minor injuries, and 1 in a major lost-time injury” (Heinrich, p.26). In other words, if one simply focuses on a major injury or sentinel event, they are probably missing several minor events and even more “no-injury” events that may have occurred. This is important because many of the cause factors seen in these less serious events are the same cause factors identified in sentinel events (Wiegmann & Shappell, 2003). These Figures and ratios have come to be known as Heinrich’s Triangle or Heinrich’s Law, which became

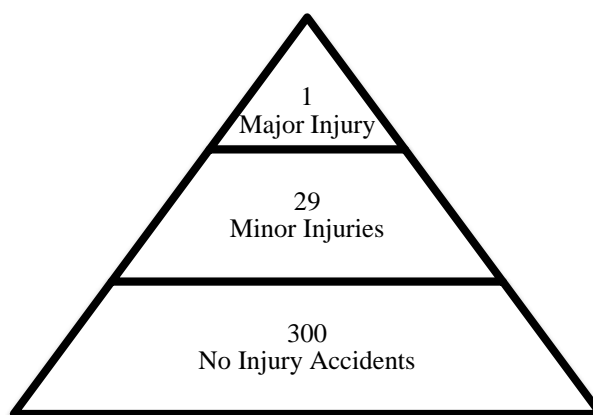


Figure 1. Heinrich's Triangle (adapted from Heinrich, 1931)

the foundation upon which many of today's industrial accident prevention programs were built (see Figure 1).

Because of the high profile nature of sentinel events, most healthcare organizations still prefer to focus on them. However, Heinrich's law would suggest that greater effort should be placed on more common minor/no injury events where the "roots" of many underlying threats to patient safety may exist. Furthermore, by employing more effort to understand non-injury events, organizations would be utilizing a more proactive approach that may broaden our understanding of the genesis of more catastrophic sentinel events and yield novel approaches to their prevention/mitigation.

To illustrate this, consider a patient who was just diagnosed with skin cancer. In terms of Heinrich's triangle, the "Major Injury or Single accident" would be a patient death due to stage 4 skin cancer. Here, treatment would have been difficult, as the cancer cells have spread beyond the skin and regional lymph nodes to distant organs. The "Minor Injury" would be an individual with stage 0 skin cancer where the cancer cells are confined to the epidermis and have not spread, making the cancer more easily treated using surgery or potentially non-invasive techniques. The "No-Injury" cases would be those individuals who have yet to present with skin cancer but may have risk factors that make them predisposed to this disease. For example, these individuals may be older in age, have weakened immune systems, fair skin, a family history of skin cancer, or are

smokers. Unlike the major or minor injury levels, at the “No-Injury” level, skin cancer prevention techniques can be used to decrease the risk for developing melanoma. Here, individuals can decrease their exposure to UV light by avoiding direct sunlight and tanning beds, they can be sure to wear sunscreen, hats and protective clothing when they do go out in the sun, and have regular thorough skin examinations.

If we could choose between the three areas outlined above, most of us would elect to be in the “No-Injury” group, where we can take preventative measures to avoid the risk of developing cancer (Minor-Injury), or having it manifest to be a disease that we die from (Major-Injury). Arguably, patients would prefer to seek preventative treatment for an illness rather than affording the disease the opportunity to spread throughout their bodies to the point where drastic interventions are required.

We can use the example above as a means to understanding the broader problem facing the healthcare industry today. Despite our best efforts in patient safety, medical errors continue to be made in hospitals across the country and in some cases may result in death. In an effort to better understand the genesis of these errors, most investigations have focused on sentinel events. For instance, academic venues like morbidity and mortality (M&M) conferences allow individuals to learn from previous experiences and modify behavior. However, this does little for the people who have already lost their lives or suffered from serious injuries. As Heinrich and others would say, it may make more sense for the healthcare industry to focus its attention on the minor injury/no injury incidents, in the hope to prevent these larger, catastrophic events.

Within healthcare and surgery in particular, one way to investigate the minor injury/no injury events, is to use an observational approach to identify human factors

issues that may lead to patient harm. Many observational studies in healthcare focus on a specific area such as communication (Coiera & Tombs, 1998; Reeve & Lyon, 2002; Gurses, et al., 2009) teamwork (Schraagen, et al., 2010; Steinemann, 2011;) technology and equipment issues (Courdier, et al., 2009; Pennathur, et al., 2013), errors (Bracco, et al., 2001; Catchpole, et al., 2008;), and interruptions (Chisholm, et al., 2000;). Other observational studies have documented the myriad of work system factors that disrupt workflow and team performance overall (Mackenzie & Ziao, 2003; Wiegmann, et al., 2006; Sevdalis, et al., 2008; Henrickson-Parker, et al., 2010; Henrickson-Parker, et al., 2010; Wiegmann, et al., 2010; Palmer et al., 2013).

One area that has received considerable attention for observational research is the cardiovascular operating room (CVOR). This is not surprising given that cardiac surgery is the most common of inpatient surgical procedures performed (CDC, 2015). Further, despite being a fairly structured, planned and organized care setting, cardiac surgery continues to be considered a high-risk procedure (Gurses, et al., 2012). While the CVOR has been researched by many, there is limited research on the differences between CVORs located at different types of facilities (i.e. an academic university medical center vs. a non-academic hospital). Although the differences between these types of venues may be self-evident, the types and frequency of events observed may vary. Understanding these observational differences may prove helpful in designing and implementing targeted interventions to protect against threats to patient safety in the future.

While observations in the CVOR are certainly of interest to the healthcare industry because of its frequent use, several other medical venues have been observed and would

benefit from the investigation of human factors related issues. While certain domains may be similar to that of cardiac surgery because of their high risk and well organized structure (e.g., spinal surgery, neuro surgery) others vary drastically. One area for example, involves care for the traumatically injured patient. Unlike cardiac cases, trauma cases are unexpected, unpredictable and particularly time sensitive. Although the differences between the CVOR and a trauma center may be conspicuous to experts in the field, the differences in observations and the types of interventions that may be put in place are certainly worth investigating.

Unfortunately, while observational studies such as these are helpful in identifying process inefficiencies that threaten the optimal delivery of patient care, they often involve several hours of observations that produce hundreds of data points or instances of potential human factors problems. Without use of a framework or classification system, the analysis of observational data can be arduous, leading to challenges for identification of targeted interventions to reduce specific threats to patient safety.

Perhaps a more systematic, theory-driven approach could be applied to observational healthcare data as a means to classify and analyze the types of human factors issues identified in medical settings. Rather than focusing on errors and adverse events, which is no different than treating a patient who is already sick, the healthcare industry would be better served to address errors by proactively identifying underlying symptoms before they manifest themselves. A method such as this has an advantage in that healthcare professionals are now able to address these threats proactively, rather than continuing to utilize a reactive approach that investigates patient safety by exploring error and preventable death in medicine.

Traditionally, the healthcare industry has focused on developing sophisticated techniques for examining the causes associated with sentinel events (the highest part of Heinrich's triangle). However, little effort has been expended within healthcare to identify potential threats to patient safety (below the peak of Heinrich's triangle) when an adverse outcome did not occur. The purpose of this dissertation is to explore the base of Heinrich's triangle by investigating a tool that has already been established in other domains (e.g., aviation, mining, maritime, defense) and its utility in proactively assessing patient safety.

CHAPTER 2: LITERATURE REVIEW

While there are several different approaches to accident investigation in the literature, there are few tools that specifically apply human factors, and even less that can be modified for use within healthcare or are currently used in this setting. This chapter describes one such tool, the Human Factors Analysis and Classification System (HFACS), as well as the models that have served as its foundation.

Reason's Swiss cheese model of accident causation

While several frameworks have been proposed for identifying human error, perhaps one of the most well-known is James Reason's Swiss cheese model of accident causation (Wiegmann and Shappell, 2003). In fact, in a review of his work, Larouzee and colleagues (2014) identified several areas where this model has been applied since the early 1990s. These areas include aviation (Maurino, 1993; Shappell 2000), maritime (Ren et al., 2008), healthcare (Vincent et al., 1998; Carthey et al., 2001; Lederman & Parkes, 2005), defense (Jennings, 2008), nuclear (Reason et al., 2006), oil and gas (Hudson et al., 1994), and railroad (Reason et al., 2006; Baysari et al., 2008). In developing his model, Reason integrated ideas from several human error perspectives in the literature. Perhaps most influential to the development of Reason's Swiss cheese model were Edward's (1972) SHEL model and Heinrich's (1931) Domino theory.

The SHEL model (a name derived from the initial letters of its factors: Software, Hardware, Environment and Liveware) was developed in 1972 by Elwyn Edwards, a notable ergonomist and aviation psychologist. Edwards maintained that productive processes were systemic in nature, and performed based on a combination of software, hardware, environmental and liveware elements. With respect to this model, *software*

refers to the rules, guidelines, and other written documents that are part of the standard operating procedures of a system. *Hardware* on the other hand, refers to any material tool used within the system (i.e., equipment). The *environment* is the situation in which the other three components must function and it specifically encompasses the social and economic climate, as well as the physical space involved. Finally, *liveware* represents the human beings that operate within the system (i.e., in a hospital, this may be doctors, maintenance workers, administrators or technicians).

Edwards' model places the human component (liveware) as the focus of interest. However, humans are not independent and unrelated factors of a system, rather, they interact with other elements such as hardware, software and the environment. It is within the interactions of these components (i.e., liveware-hardware, liveware-software, liveware-environment, hardware-software, hardware-environment, software-environment) that problems can occur. Captain Frank Hawkins later modified the model into a building block structure and expanded it to include a second "liveware" element (now called SHELL model, see Figure 2) in order to represent group processes between humans, or the liveware-liveware interaction (Hawkins and Orlady, 1993).

In terms of accident investigation, this model supports a systems perspective that maintains that the human is rarely the solitary cause of an accident. Rather, this perspective focuses on a variety of contextual and task-related factors that interact with the human which may in turn affect their performance.

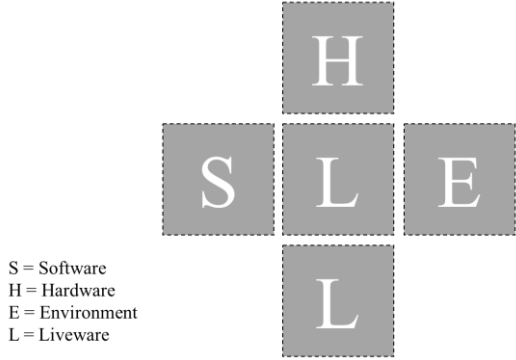


Figure 2. SHELL Model (adapted from Hawkins, 1993)

Another theory that played a key role in the development of Reason’s Swiss cheese model was Heinrich’s domino theory. This theory is based upon a sequential accident model in which accidents transpire as a result of a chain of events that occur in a particular chronological order. In accordance with the domino theory, there are five factors (i.e., dominos) in a given accident sequence: (1) ancestry and social environment, (2) fault of the person, (3) unsafe acts and/or mechanical or physical hazards, (4) the accident and (5) subsequent injuries (see Figure 3).

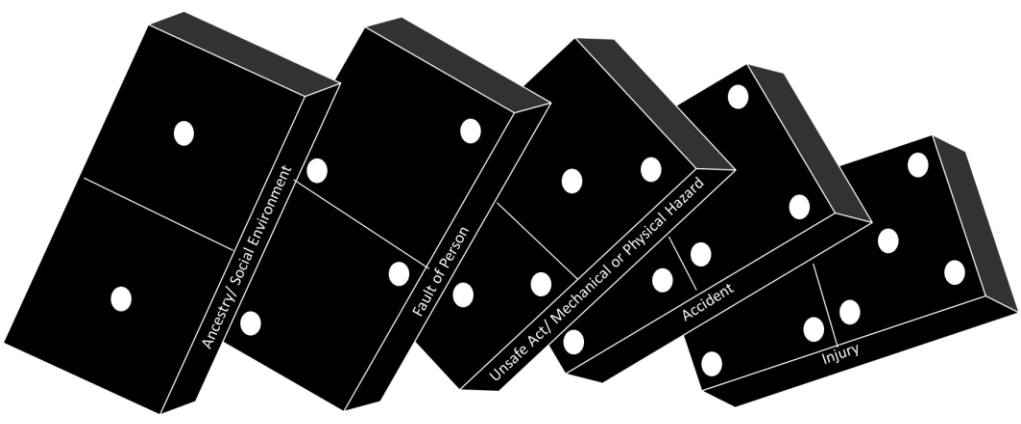


Figure 3. Domino Theory (adapted from Heinrich, 1931)

Heinrich maintains that a resulting injury is always caused by an accident, and the accident in turn is the consequence of the factor that immediately precedes it. Each of the five factors described by Heinrich occur in a chronological order and are explained below.

The first factor, *ancestry and social environment* can be explained by undesirable character traits (e.g., stubbornness, recklessness, or greediness) either passed along through inheritance or influenced by the environment. Both inheritance and environment can cause the second accident factor, fault of a person. *Fault of a person* represents acquired faults of an individual (e.g., violent temper, inconsiderateness, or overconfidence) that may constitute reasons for committing the next accident factor.

In essence, the first two dominos engage what is commonly known as the nature vs. nurture argument. Here, the ancestral/social environmental traits can be thought of as “nature”, representing those qualities that we inherit. While faults of a person can be described as “nurture” and embodies learned and/or acquired behaviors throughout one’s development.

The third accident factor, *Unsafe act and/or mechanical or physical hazard*, is the unsafe performance of individuals (e.g., not wearing personal protective equipment, or failing to wash hands before visiting a patient) and mechanical or physical hazards (e.g., slippery surfaces or wires and tubing tangled across the floor) that result directly in an accident. The fourth domino, the *accident* itself represents specific events like tripping on tangled wires, slipping on a wet floor, or running into another person that can ultimately lead to the final domino, injury. *Injury* occurs as the direct result of an accident and can include anything from minor injuries like bruising, fractures, and lacerations to major injuries like closed head injuries or even death.

The five accident factors are depicted in a domino fashion, such that the collapse of the first domino (ancestry/social environment) will result in the collapse of the remaining dominos. In this theory, an undesirable or unexpected event initiates a sequence of events which leads to an accident. This theory implies that the accident is the result of a sequence of events whose “root” cause resides higher in the system at the ancestry/social environment level. Theoretically, if one were to eliminate that root cause, the accident would not occur.

The theories proposed by Edwards and Heinrich suggest that adverse events rarely occur in isolation. Rather they are the result of a combination of factors that influence the system as a whole. Reason expanded on these theories by describing four layers of interaction: unsafe acts, preconditions for unsafe acts, unsafe supervision and organizational influences.

Much like Heinrich’s domino theory, each layer of Reason’s model is positioned one behind the other, acting as barriers to protect the system as a whole (see Figure 4). Typically, each layer or barrier has been depicted as a slice of Swiss cheese, within which failures (i.e., the holes in the cheese) exist. These holes in the cheese can take the form of either active or latent failures.

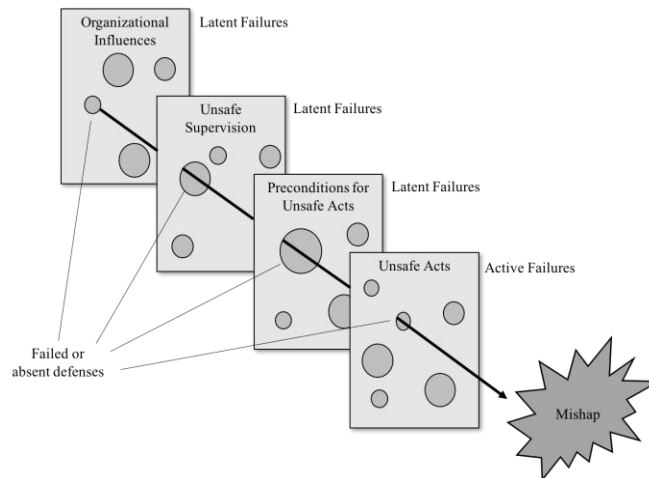


Figure 4. Reason's Swiss cheese model of accident causation (adapted from Reason, 1990)

Active failures, can be thought of as faults that are directly linked to an accident. For example, failing to stop at a stop sign can directly lead to an adverse event. Conversely, latent failures can be thought of as background elements which may eventually lead to an unsafe act. For example, things like inappropriate attitudes, failures in attention and poor communication may not directly cause an accident, but can certainly act as contributory elements in the system. The difference between active and latent failures is that active failures have immediate consequences while latent failures may lie dormant for days, months, or even years, before they contribute to an accident.

Reason expands upon this concept of active and latent failures as he describes his four layers of Swiss cheese, the first of which is the *unsafe acts* of operators. These active failures are easy to identify as they tend to be events that directly result in an accident. For example, consider a floor nurse with 20 years of experience, who one day forgot to administer life sustaining medication to one of her seven patients. In this situation, the inadvertent neglect by the nurse represents an unsafe act.

During the course of the investigation, hospital administrators may ask her how long she has been doing her job, or if she has ever made a mistake like this before. Because she is so experienced and has never made such a devastating error, asking these questions may not necessarily aid the hospital in understanding how or why this event took place.

If instead, the administrator asked the nurse, what was different about this day, then any routine day on the job, he would have learned that she was very fatigued and distracted, which is out of the norm for her. The nurse would have explained that she did not sleep the night before because she had to take her daughter to the emergency room, and although she arrived on time for her shift, she was mentally unfocused, as she was concerned for her daughter.

With this in mind, Reason developed a second layer that could address these underlying factors that may have played a role in how an individual performs on a given day. The second layer, *preconditions for unsafe acts*, involves conditions that can directly affect human performance. Operator factors such as mental or physical fatigue, illness, ineffective communication, and poor coordination practices represent failures at this level. For example, a nurse who is not getting enough sleep or is taking over-the-counter medications may not be able to perform at her best. Likewise, when individuals have difficulty with communicating and coordinating effectively, the likelihood of an unsafe act occurring increases.

While it is certainly important to understand how the preconditions impact humans, Reason took this methodology even further by moving beyond the individual, to investigate the possible latent failures associated with supervisors involved at the third

layer, *Unsafe Supervision*. It is at this level that investigators can begin to understand exactly why failures at the preconditions for unsafe acts level took place. In many cases a breakdown in appropriate communication and coordination practices may be traced up the chain of command to an issue at the supervisory level.

Consider the nurse medication error example. While an investigation at the preconditions for unsafe acts level revealed that she was distracted and fatigued, had we stopped there, we would have missed other failures involved with unsafe supervision. For instance, what we did not know before is that during her shift, the nurse was scheduled to work with a nurse-in-training, who did not have all of the experience necessary to care for her own patients. Traditionally, nurses-in-training are accompanied by other, more experienced nurses with whom they shadow and learn from. However, on this day, the supervisor scheduled the nurse-in-training to work without her mentor who was on vacation. Because of this, the remaining, experienced nurses had to pick up the slack for the nurse-in-training, as well as assist her when she had questions.

However, beyond an issue at the supervisory level, the *organization* itself can contribute to an accident in ways that are often unnoticed by investigators. It is at Reason's fourth level that important decisions are made by the high-level managers of the system. Here, decision-makers use input from the outside world to establish goals for the organization as a whole and determine how these goals should be met. While the overall aim at this level is to maximize productivity and safety, considerable effort is placed on the allocation of finite resources such as money, equipment, people and time.

Why exactly did the supervisor knowingly schedule a nurse-in-training to work a shift on her own, without supervision? Further investigation up the chain of command

may conclude that the hospital was facing a time of fiscal austerity. Because of this, the allocation of resources for hiring was severely cut, and the organization could not afford to hire additional staff. Unfortunately, due to this nursing shortage, the supervisor was left with no choice but to schedule the nurse-in-training to cover a shift on her own.

In sum, by utilizing Reason's Swiss cheese model of accident causation, we have a better understanding of the genesis of the accident. As a result of economic deficits, decision makers at the organizational level of the hospital could not hire new staff members, creating a lack of staff, particularly nurses. Because of this shortage, the supervisor in charge of the weekly scheduling of floor nurses, had no choice but to schedule all of the available nurses during the busiest shift. While the supervisor was aware that nurses-in-training are among the selection of available nurses for the week, he usually only schedules these trainees when they can work alongside an experienced nurse mentor. However, the experienced nurse who traditionally works with the trainees was on vacation, forcing the supervisor to schedule the inexperienced trainee without a mentor.

As a result, the nurse who committed the medication error was overworked, having to manage not only her seven patients but needing to assist the nurse-in-training with her patients as well. On top of her high workload, the nurse was physically and mentally fatigued from spending the night in the Emergency Room with her daughter. As much as she tried to remain focused on the several tasks ahead, she was distracted by concern for her daughter who was still in the Emergency Room, and frustrated that she should not be with her.

Using Reason's Swiss cheese model of accident causation, it is much easier to see how a simple memory lapse could occur during such a complex series of events.

Unfortunately, while this model serves as a good assimilation of different human error perspectives, it is primarily descriptive and does not operationalize the “holes” or failures in the system. In other words, while Reason describes the four layers of accident causation, he fails to explain what the failures in the defenses actually are, instead leaving the identification of the holes to the user. Consequently, the model in its present form, may be impractical for use in healthcare for incident and accident investigation.

The Human Factors Analysis and Classification System (HFACS)

In an effort to further define Reason’s Swiss cheese model of accident causation, Shappell and Wiegmann (1998) utilized accident data from the U.S. Navy and Marine Corps to develop The Human Factors Analysis and Classification System (HFACS). HFACS describes failures (i.e., holes in the cheese) at each of the four levels described by Reason (Wiegmann & Shappell, 2003). Included within these four levels (also called tiers), are 19 causal categories. Each of these causal categories, as they appear at each level are summarized in Table 1, depicted in Figure 5 (white boxes), and described below.

Table 1. Description of the HFACS Categories

Organizational Influences
Organizational climate: Prevailing atmosphere/vision within the organization including such things as policies, command structure, and culture
Operational process: Formal process by which the vision of an organization is carried out including operations, procedures, and oversight among others
Resource management: Management of necessary human, monetary, and equipment resources
Unsafe Supervision
Inadequate supervision: Oversight and management of personnel and resources including training, professional guidance, and operational leadership among other aspects.
Planned inappropriate operations: Management and assignment of work including aspects of risk management, crew pairing, operational tempo, etc.
Failed to correct known problems: Those instances when deficiencies among individuals, equipment, training, or other related safety areas are “known” to the supervisor, yet are allowed to continue uncorrected
Supervisory violations: The willful disregard for existing rules, regulations, instructions, or standard operating procedures by management during the course of their duties
Preconditions for Unsafe Acts

Environmental Factors

Technological Environment: This category encompasses a variety of issues including the design of equipment and controls, display/interface characteristics, checklist layouts, task factors and automation

Physical Environment: Included are both the operational setting (e.g., weather, altitude, terrain) and the ambient environment, such as heat, vibration, lighting and toxins

Conditions of the Operator

Adverse mental states: Acute psychological and/or mental conditions that negatively affect performance such as mental fatigue, pernicious attitudes, and misplaced motivation

Adverse physiological states: Acute medical and/or physiological conditions that preclude safe operations such as illness, intoxication, and the myriad of pharmacological and medical abnormalities known to affect performance

Physical/mental limitations: Permanent physical/mental disabilities that may adversely impact performance such as poor vision, lack of physical strength, mental aptitude, general knowledge, and a variety of other chronic mental illnesses

Personnel Factors

Communication, coordination and planning: Includes a variety of communication, coordination, and teamwork issues that impact performance

Fitness for duty: Off-duty activities required to perform optimally on the job such as adhering to crew rest requirements, alcohol restrictions, and other off-duty mandates

Unsafe Acts**Errors**

Decision errors: These “thinking” errors represent conscious, goal-intended behavior that proceeds as designed, yet the plan proves inadequate or inappropriate for the situation. These errors typically manifest as poorly executed procedures, improper choices, or the misinterpretation and/or misuse of relevant information

Skill-based errors: Highly practiced behavior that occurs with little or no conscious thought. These “doing” errors frequently appear as breakdown in visual scan patterns, inadvertent activation/deactivation of switches, forgotten intentions, and omitted items in checklists often appear. Even the manner or technique with which one performs a task is included

Perceptual errors: These errors arise when sensory input is degraded, as is often the case when flying at night, in poor weather, or in otherwise visually impoverished environments. Faced with acting on imperfect or incomplete information, aircrew run the risk of misjudging distances, altitude, and decent rates, as well as responding incorrectly to a variety of visual/vestibular illusions

Violations

Routine violations: Often referred to as “bending the rules” this type of violation tends to be habitual by nature and is often enabled by supervision/management that tolerates such departures from the rules.

Exceptional violations: Isolated departures from authority, neither typical of the individual nor condoned by management

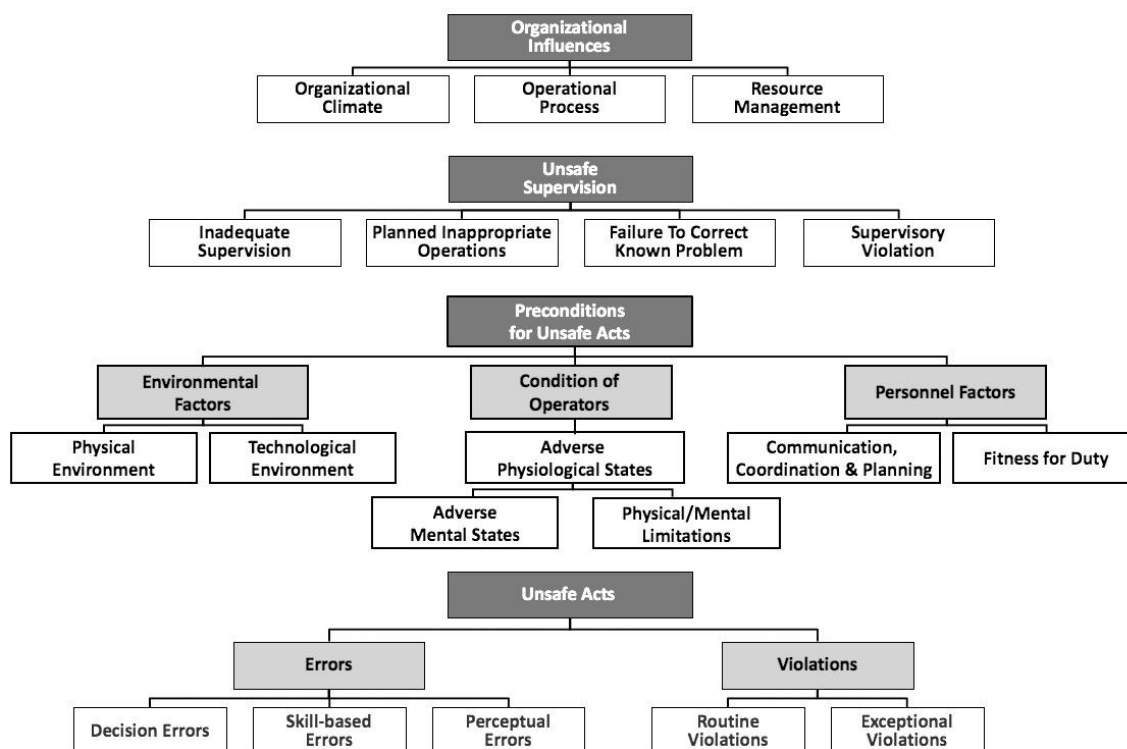


Figure 5. The Human Factors Analysis and Classification System (HFACS)

Description of the Categories

Unsafe Acts: The first tier of HFACS, “unsafe acts”, describes the active failures of operators that may ultimately lead to an unintended outcome, and in some cases an accident. There are two major types of unsafe acts: errors and violations.

Errors. Errors can be best described as mental or physical actions that proceed as intended yet do not achieve their desired outcome. Wiegmann and Shappell describe three error types in their taxonomy: (1) decision errors, (2) skill-based errors and (3) perceptual errors.

Decision errors. This subdivision of unsafe acts can best be described as products of conscious, volitional behavior that proceeds according to plan but proves inadequate for the situation. These errors can be thought of as “honest mistakes” that typically result due to poor information, a lack of knowledge, or issues associated with experience. The

latter of which could be due to too much or not enough experience with respect to a given task.

Cases of poor information refer to instances in which an individual may lack important material or data related to a situation that could help them make an appropriate decision. Consider for example an unconscious patient who arrives at a trauma center after a motor vehicle collision. Upon arrival, the resident notes that the patient has several lacerations and fractures and calls for the administration of morphine to minimize pain. Because the trauma team could not find the patient's wallet, and the patient was unconscious, no one could inform the resident that the patient was in fact allergic to morphine. As a result, upon administration, the patient's throat began to close and his respiratory rate dropped drastically. Assuming that the attending physician sees this in time, it may turn out to be benign; however, if this is missed, the patient may not survive.

Decision errors can also occur because of a lack of knowledge. As an example, consider the case of a patient who died because his healthcare provider did not recognize the symptoms of his illness. Dr. Wachter, who at the time was a second-year medical student, started rounds on his 71-year-old patient who had just had a hip operation. Upon seeing that the man was sweating profusely and panting, Dr. Wachter checked his chart and noted that his respiration and heart rate had been climbing, but his temperature was steady. After listening to his lungs (which sounded clear), Dr. Wachter concluded that the problem did not seem like heart failure or pneumonia, and he attributed his patient's condition to the stuffy room. He told his patient to rest until the other members of the medical team could return later to check on him.

Later that morning, a *Code Blue* was called to the patient's room, and when Dr. Wachter arrived, he found that his patient had died (i.e., the patient was in cardiac arrest). The autopsy revealed that the patient suffered a massive pulmonary embolism. Despite having good information from the symptoms expressed by his patient, which indicated a blood clot, Dr. Wachter did not know this, as he had yet to read that chapter in his medical textbook (Kita, 2012).

Poor decisions may also be a result either too much, or too little experience on a given task. As an example of someone lacking experience, consider the following case. A 28-year-old woman who recently had a bilateral lung transplant, was admitted to the hospital when she presented with sudden onset of severe shortness of breath. Diagnostic studies revealed that she was producing donor-specific antibodies, and as a result she was treated for humeral rejection. This treatment included placing a large bore central line (similar to a hemodialysis catheter) on the right side of her chest and daily bedside plasmapheresis therapy. A registered nurse received orders to draw the patient's morning labs. Although she had worked with many other types of catheters, the nurse asked the charge nurse for specific instructions for this type of catheter, as she had never used one before. The charge nurse provided her with the following basic verbal instructions: "waste 3 cc, draw labs, flush with saline, HEP-LOCK." The nurse felt confident that the verbal instructions were sufficient and she went to the patient's room to complete the task.

When the nurse entered the room to draw the labs, the patient was awake and in no apparent distress. However, after all the tubes had been filled, the patient sat upright and said, "something isn't right." The nurse reached around the bedside Table to grab the saline flush, and the patient began to convulse. She called for help as the patient lost

consciousness and fell to the floor and bled from her catheter. The patient spent the next 3 days in the intensive care unit (ICU) and testing revealed a cerebral air embolism. The nurse manager conducted an immediate and thorough incident review, which revealed that the nurse had failed to clamp the catheter prior to removing the syringe, allowing air to enter the catheter and obstruct the patient's circulatory system. The devastated nurse requested a temporary leave of absence, but never returned to work. The hospital enacted a policy requiring all nurses receive education on catheter placement and allowed only trained nurses to access the catheters (Swayze & James, 2013)

Skill-based errors. This category of errors arises during automatized behavior that requires little conscious thought, and are often attributable to memory, attention or technique failures. Failures in memory can manifest as omitted items on a checklist, place losing or forgotten intentions. For example, forgetting to roll a patient with a penetrating wound to examine his back during a trauma resuscitation effort may be the result of a memory failure. As a result, the patient bled out due to a puncture wound to the back side of the left lung.

As another example, consider 17-year-old Jesica Santillan who received the heart and lungs of a patient whose blood type did not match hers. Unfortunately, doctors at the medical center had forgotten to check the compatibility of the transplant organs before the surgery began. As a result, Ms. Santillan died two weeks later (Kopp, 2013).

In a similar situation, medical resident Dr. Danielle Ofri was already juggling patients with acute heart failure and rampant infections when she was assigned to a nursing home patient with dementia. This patient was a perfect candidate for the intermediate unit, and by transferring her there, the workload would be reduced for Dr.

Ofri. However, before she could transition the patient, she had to rule out any treatable medical conditions by getting the labs, reviewing the head CT scan and chest X-ray.

Because it was 4:45pm and the doctors at the intermediate unit left by 5:00pm, Dr. Ofri scanned through the labs, called the ward's doctor and ran through the case. The doctor on the phone verified that "the labs and everything [were] normal" and Dr. Ofri agreed. Within a few minutes the patient was sent to the intermediate ward.

The next afternoon the intermediate ward doctor found Dr. Ofri and told her that she was called overnight by the radiologist because the patient's head CT showed an intracranial bleed. The patient had been rushed into neurosurgery to get the blood drained from inside her skull. In an effort to decrease her patient load, Dr. Ofri completely forgot to check the head CT. Luckily, someone else caught this mistake, and the patient survived (Ofri, 2013).

Skill-based errors that appear as attentional failures typically involve incorrect or omitted actions, distractions, or task overload. As an example, consider an overloaded ED nurse working the night shift. During her shift she is managing several patients, one of which is a baby boy who needs a heparin injection. While she is pulling the vial of heparin from the automated dispensing cabinet, she hears an overhead page that a trauma alert patient is about to arrive. Because she is on the trauma team, the nurse quickly checks the label on the medication, prepares it, and administers it intravenously to the boy. She runs out of the room towards the trauma bay not realizing that instead of administering the heparin in a concentration of 10 units/mL, the infant receives heparin in a concentration of 10,000 units/mL and dies.

Finally, regardless of training, experience, and educational background, specific differences in techniques can also set up individuals for specific failure modes. Two surgeons with identical training and skill may differ significantly in the techniques they use to operate on a patient and manage their operating room. One may be slow and methodical, requiring that no one speaks during the procedure, while the other may be fast and efficient, allowing music and background chatter.

In and of itself, one technique may not be better than the other; however, each demonstrates potential opportunities for failure. For example, a surgeon who demands a quiet OR, may create a fearful atmosphere where team members are frightened to speak up or even communicate aloud to one another. This environment could lead to issues in miscommunication and poor coordination.

While the setting may be vastly different in another room where a different surgeon tolerates chatter and music, team members may be equally susceptible to failures. For instance, individuals in the “noisy” operating room may be more easily distracted and less likely to pay attention to events in the room.

Perceptual errors. Finally, *perceptual errors* involve improper actions based on the misinterpretation of sensory input. These errors tend to occur when sensory input is degraded or “unusual” as is the case with visual illusions and spatial disorientation. Because of context cues and our past experiences, our brains often fill in “missing” information without us knowing. For example, if a word is missing on a page of a book, most individuals won’t even realize this because our brains fill in this information based on the context provided. Perceptual errors occur when individuals inadequately respond to information that they may or may not have actually perceived.

For example, take a few seconds to look at Figure 6, an X-ray computed tomography (CT) scan of a human lung.



Figure 6. X-ray CT (from Drew, Vo & Wolfe, 2013)

Do you notice anything strange about this image? If not, keep looking. Did you spot the gorilla? The image comes from a study published in *Psychological Science* where Trafton Drew and colleagues (2013) working at the Brigham and Women’s Hospital found that when people focus on searching these images for bright white cancer nodules, they never notice the gorilla. Perhaps more shocking is that radiologists – individuals who are specifically trained to read CT scans, usually miss the gorilla as well.

As an example of an auditory perceptual error, consider a cardiovascular procedure that requires the patient to go on cardiopulmonary bypass. This procedure involves a surgeon, perfusionist and anesthesiologist. During the case, the surgeon may ask the perfusionist, who manages the heart-lung machine, to “go up” or “go down” on pump. This changes the rate at which blood flows in and out of the patient. On the other hand, the surgeon may also ask the anesthesiologist, who is in charge of keeping the patient asleep, as well as managing the patients position on the operating Table, to “go up” or “go down” in terms of the patient’s positioning.

During one of these cases, the surgeon declared “up on pump” just after he had been speaking to the anesthesiologist. Because the anesthesiologist assumed the surgeon was still communicating with him, as soon as he heard the word “up” the anesthesiologist started to reposition the patient on the operating Table. Immediately, the surgeon declared “up on pump! I wasn’t talking to you.” The surgeon was requesting that the perfusionist increased the flow of cardiopulmonary bypass, and the anesthesiologist misheard the request, because of the context and the routine nature of his brain to fill in the missing information.

Violations. While errors are often referred to as “honest mistakes” that occur within the safety norms of an organization, *violations* involve an operator’s willful departure from the rules or regulations of safety. In other words, the individual knew what the rule or regulation was but elected not to adhere to it. While there are many ways to differentiate between types of violations, the HFACS methodology makes the distinction between those that are *routine* in nature and those that are *exceptional* to an individual’s normal behavior.

Routine violations. Routine violations, often referred to as “bending of the rules”, tend to be habitual in nature and are often tolerated by supervisory authority. Individuals who commit this kind of violation have the ability to work within the rules when they want to, however, in these situations they chose not to. For example, consider an interstate with a speed limit of 65 mph. Many drivers will consistently drive five mph above this limit, often without reprimand. However, in a school zone, individuals rarely speed because they choose not to violate the rules in a situation where someone (e.g., a child) could get hurt. Similarly, while the rules and regulations of a given hospital may

dictate that employees must wash their hands before and after they see each patient, an Emergency Department (ED) physician who elects not to wash his hands is not likely to get reprimanded. Further, the supervisor at the hospital may not only allow for this rule to be broken, but they themselves may violate said rule, which makes matters more difficult.

Exceptional violations. Unlike routine violations, exceptional violations appear as isolated withdrawals from authority. These violations are not condoned by management and are not usually indicative of an individual's typical behavioral patterns. These types of violations are not considered "exceptional" because of their extreme nature. Rather, they are considered exceptional because they are outside of the individual's typical behavior. An isolated event of driving 105 mph in a 65 mph zone would be representative of an exceptional violation, as it is highly unlikely that the individual behaves in this way on a regular basis.

Consider Dr. Sulieman Al Hourani who was dismissed from his position after removing the entire right testicle of a patient who was only supposed to have a cyst removed. As soon as the assisting nurse turned away to get a transfusion stitch, the incident occurred and the testicle was "mistakenly" removed. Upon further investigation, examiners found that Dr. Al Hourani not only stole supplies from the hospital (two boxes of dihydrocodeine) but also injected himself with medication that was meant for a patient. Specifically, Dr. Al Hourani had consulted a colleague and was advised to inject a patient with 10 milligrams of midazolam. He gave the patient 8mg and injected himself with the remaining 2mg (Surgeon cut, 2010).

Preconditions for Unsafe Acts: Very few unsafe acts are isolated events; rather they are often the end result of latent failures intrinsic to the system. The second tier,

preconditions for unsafe acts, captures those latent failures associated with the individual and the general working environment. This level includes three over-arching categories (i.e., environmental factors, conditions of the operator, and personnel factors), which can further be broken down into seven distinct causal categories.

Environmental factors. Within HFACS environmental factors have been separated into two causal categories: physical environment and technological environment. The first, *physical environment*, refers to both the operational environment (e.g., weather, terrain and altitude), as well as the ambient environment (e.g., heat, vibration, lighting and toxins). Certain aspects of the physical environment can make it difficult for the individual to complete their tasks. For example, the high temperature of a room can cause dehydration, reducing the operator's concentration level. This is particularly true in a trauma resuscitation bay, where the room temperature must be kept high to reduce the risk of hypothermia. Trauma team members are often dripping with sweat while trying to resuscitate patients as they must wear personal protective equipment and lead vests for protection during x-ray imaging.

The second environmental factors category, *technological environment*, includes traditional usability issues associated with equipment, software, and several forms of documentation including checklists and procedures. As an example, consider the case of 79-year-old Richard Smith who was receiving dialysis for kidney disease. During treatment, he started to experience shortness of breath and was admitted to the ICU. The next day, Mr. Smith complained of a stomach ache and was prescribed an antacid, or so he thought. Rather than metoclopramide (an antacid), Smith was given pancuronium, a paralytic and muscle relaxant that is used for intubation in small doses, and for lethal

injection in larger doses. Unfortunately, the pancuronium had put Mr. Smith into respiratory arrest (Gora, 2016).

How were two such different medications so severely mixed up? Take a look at Figure 7 below from a Medication Safety Alert (1999). On the left is pancuronium, (the incorrect medication issued to Mr. Smith), while metoclopramide (the medication he should have received) is on the right.

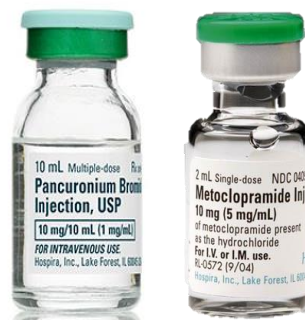


Figure 7. Left to right - Pancuronium (paralytic), Metoclopramide (antacid)

Condition of the operator. The conditions of the individuals within a system can, and usually do, influence their performance on the job. Wiegmann and Shappell use three categories to address the issues involving individuals: (1) adverse mental states, (2) adverse physiological states and (3) physical/mental limitations.

Adverse mental states refer to the mental conditions of operators that may affect performance. Mental conditions include cognitive states such as distraction, inattention and mental fatigue, as well as personality traits and attitudes such as anger, overconfidence, and frustration. Predictably, if an individual is mentally tired, frustrated, has a loss of situational awareness or task fixation, there is an increased chance of an error occurring. Similarly, overconfidence and other hazardous attitudes such as egotism may increase the likelihood that a violation will be committed.

As an example, consider a patient who came to the hospital for glaucoma surgery but lost his vision as a result. The nurse on the case was confident that she knew what preoperative medications were required for the surgery, so when the physician did not provide preoperative eye drop orders, she was convinced he had made a mistake. Rather than discussing this with the physician, the nurse created an order sheet listing the medications she was certain the doctor would have ordered, and then administered them to the patient. Unfortunately, the prescription eye drops were contraindicated for the specific type of glaucoma this patient had. Once the surgeon became aware of the nurse's mistake, he tried to reverse the effects of the medication. Unfortunately, the patient had a poor outcome and lost vision as a result of the nurse's overconfidence (Overconfidence, 2012).

Consider another case where a doctor administered a lethal dose of diamorphine, killing his patient within hours. Dr. Ubani, who had been recruited by an agency to provide out-of-hours cover in Cambridgeshire, arrived in the United Kingdom one day before the incident. The doctor explained that he was extremely tired and couldn't concentrate, leading him to use the wrong drug, resulting in the death of his patient (Tired German doctor, 2009).

The second type of operator condition involves *adverse physiological states*. This category refers to those medical or physiological conditions that may preclude safe operations. Issues involving operator illness, physical fatigue, and several pharmacological and medical abnormalities that affect performance are considered adverse physiological states. While many individuals continue to go to work when they have a head cold, working in this state typically results in degraded performance,

increasing the risk of an error. Consider for example an anesthesiologist who is suffering from a bad cold. Determined to come to work he stocks up on over-the-counter antihistamines, acetaminophen, and other non-prescription pharmaceuticals. However, taking this myriad of medications makes him extremely drowsy and less alert to the patient monitors, increasing the chance of missing the decline of the patient's status.

The third and final category involves an individual's *physical/mental limitations*. While adverse mental and physiological states tend to be acute, physical/mental limitations tend to be longer lasting and include such things as hearing loss, visual acuity changes and traditional anthropometric issues such as height and weight. Specifically, this category includes those instances when operational requirements exceed the capabilities of the individual. For example, some tasks require that individuals be a certain height (e.g., adjusting a headlamp in an operating room) or have a given amount of strength (e.g., lifting a patient's legs for sterile scrubbing). Unfortunately, when the requirements are outside of the scope of an individual's abilities, errors may occur.

Consider a floor nurse who needs to reposition her patient to check for bed sores. She is quite small, but because she does not want to ask the other nurses for help, the nurse tries to roll the patient herself. Unfortunately, she loses her balance and falls onto the patient, pulling on his stitches from his earlier procedure, requiring that the doctor come back and re-stitch the patient.

Personnel factors. Some preconditions for unsafe acts can may also be considered personnel factors. While there are a number of ways that an individual's condition can lead to the commission of unsafe acts, there are also several things that an individual may do to themselves to create these preconditions for unsafe acts. Wiegmann and Shappell

describe personnel factors based on two categories: (1) communication, coordination and planning and (2) fitness for duty.

The first, *communication, coordination and planning*, accounts for occurrences of poor communication or coordination among personnel. Many industries and organizations have noted the importance of good communication and coordination skills in maintaining a well-functioning system. There have been several instances in which the lack of team coordination or improper communication has led to confusion and poor decision making. One of the most famous examples was the crash of a civilian airliner in the Florida Everglades. Here, the entire crew was vigorously trying to troubleshoot what turned out to be nothing more than a burnt out indicator light, while the autopilot was inadvertently disconnected. Because no one in the cockpit was monitoring the aircraft's altitude, or "flying" the aircraft, the plane entered a slow, unrecognized descent that ended in numerous fatalities.

Issues of poor communication, coordination and planning can also take place in a medical setting. Consider the case of Sarah Fudacz, a 24-year-old patient with end-stage renal failure. Ms. Fudacz was under anesthesia and ready to receive a perfect-match kidney donated by her brother when the operation was suddenly stopped. A part-time nurse accidentally threw away the donor kidney when she discarded the contents of the slush machine before the kidney was relocated to Ms. Fudacz's operating room. The nurse who was cleaning up, had just returned from a lunch break and thought the kidney was already in Ms. Fudacz's room when she discarded the machine's contents. Unfortunately, despite best efforts by the doctors to resuscitate the kidney, it was rendered unusable and Ms. Fudacz was forced to find another donor (James, 2013).

The second category of personnel factors, *fitness for duty*, sometimes referred to as fitness for duty, typically includes activities performed off the job that influence a person's ability to perform when they come to work. Breakdowns in personnel readiness can occur when individuals fail to adequately prepare physically or mentally for their obligations. Failing to get proper rest, self-medicating, and drinking alcohol before work, can all lead to adverse mental states, which may ultimately lead to errors and accidents down the road.

As an example consider a Northern Kentucky family who was awarded 2.55 million dollars in compensatory and punitive damages against Dr. Gregory Duma, who was intoxicated when he assisted in the delivery of their son in 2005. Nurses testified that Dr. Duma had fallen asleep during the procedure and the baby boy had a damaged right arm and broken humerus after delivery (Family wins, 2009).

Unsafe Supervision: Unsafe supervision, the third tier of the HFACS methodology, focuses on actions and decisions at the supervisory level of an organization that can adversely affect operator performance and/or the overall safety and efficiency of a system. There are four causal categories of unsafe supervision: (1) inadequate supervision, (2) planned inappropriate operations, (3) failure to correct a known problem, and (4) supervisory violations.

Inadequate supervision. The role of any supervisor, regardless of the industry, is to provide their personnel with whatever it takes to ensure the job is done safely and efficiently. Because supervisors must provide guidance, training, leadership and oversight amongst other resources, adequate supervision is not easy, and it is not always done. Inadequate supervision describes factors related to flawed oversight and management of

an organization's personnel and resources. Examples of issues that fall into this category include the lack of professional guidance, poor leadership and insufficient training. When a supervisory system lacks guidance and oversight, the likelihood that violations are committed increases. As such, it is important that accident investigators consider the role that supervision plays (i.e., the supervision was inappropriate vs. the supervision did not occur at all).

Consider the case of Emily Jerry, a two-year-old girl who lost her life due to a preventable medical error. Emily was diagnosed with a yolk sac tumor at one-and-half-years-old. After many successful surgeries and treatments, the tumor had completely vanished, and her doctors said it was as if she had never had cancer in the first place. Regardless, Emily was still scheduled to receive her last chemotherapy session on her second birthday. On the third day of this treatment, Emily was administered a fatal dose of sodium chloride solution that left her brain dead. A pharmacy technician that had been working for the hospital very several years decided not to use a standard prepared bag of sodium chloride solution (with less than 1% of sodium chloride). Instead, she filled a plastic bag with a concentrated sodium chloride solution of 23.4%, of which she had compounded herself. Eric Cropp, the pharmacist in charge, failed to detect the mixing error. Cropp was called to dispense the chemotherapy and after the technician mixed the solution, he felt rushed to check the chemotherapy, which was among many other solutions, vials and syringes. He reported seeing an empty 250 mL bag of 0.9% sodium chloride near the bag of mixed chemotherapy and assumed the technician had used it to prepare the base. Because of this inadequate assumption, Cropp served a 6-month jail

sentence, 6 months of home confinement, 3 years of probation, 400 hours of community service, and a \$5,000 fine (Eric Cropp weighs in, 2009).

Planned inappropriate operations. The second category, planned inappropriate operations, involve situations where supervisors engage in actions that have a direct negative impact on the operator's performance. Examples include improper staffing, failure to evaluate risk associated with a task, or having goals that are in opposition of the organization's rules. Occasionally, the operational tempo and/or scheduling of the operators is such that the individuals are put at unacceptable risk, or they cannot get appropriate rest, performance is adversely affected

During one winter near the end of her training as a third-year medical student, Dr. Pauline Chen came down with a terrible cold. The constant coughing and runny nose made her miserable and tired during her shifts, she even had to wear two masks every time she scrubbed in for a surgical case. Other doctors and nurses on her team were becoming ill, and after weeks of coming into work, she finally asked her senior doctor-in-training if she could go home because of an upset stomach. Rather than telling Pauline "yes", the senior doctor made her feel bad for wanting to go home, and told her "just remember that I've never missed a day at the hospital in my life. They'll have to put me in the hospital to keep me from my patients" (Chen, 2013).

Failure to correct a known problem. A third category of unsafe supervision involves cases where a supervisor failed to correct a known problem. This category refers to situations in which deficits in some aspect of the organization are recognized by the supervisor, yet continue unabated. These deficiencies can include issues among individuals, equipment, training, or other related safety areas. The failure to consistently

correct issues or discipline inappropriate behavior can foster an unsafe environment that promotes violations to the rules.

Dr. Christopher Dunstch, conducted a delicate spinal surgery on Barry Morguloff, a patient suffering from back pain. However, after his surgery, Mr. Morguloff continued to feel pain, and in fact felt worse than he did before his operation. Another doctor examined Mr. Morguloff and found that not only had bone fragments had been left on the nerves, but the hardware in his spine was installed incorrectly. After further investigation, Dr. Dunstsch was not only found to be “completely incompetent”, but he was using drugs while working. A bottle of vodka was in his desk, he was using painkillers, and a bag of white powder was found in his private bathroom. Dr. Dunstsch left one patient in the operating room so he could go to Las Vegas, and skipped five drug tests during his time at the hospital. He was allowed to operate despite these problems because the hospital had advanced him \$600,000 to move from Tennessee to Dallas, and they wanted to earn their investment regardless of the consequences (Gora, 2016).

Supervisory violations. The last causal category of unsafe supervision, supervisory violations, is similar to violations at the unsafe acts tier, and describe a supervisor’s willful disregard for the rules and regulations of personnel safety. While these violations don’t occur often, some supervisors have been known to violate the rules mandated by their organization when managing their resources. For example, allowing an unqualified individual to engage in a certain task that is outside of their allowed privileges would qualify as a supervisory violation. Consider a trauma team manager who is short staffed and assigns one of the un-trained nurses to retrieve a patient from the helipad. Here, trauma team members are required to go through a helicopter safety

training class to be qualified for patient retrieval from the helipad. Because the nurse was un-trained, she approached the helicopter before she received permission from the pilot, causing chaos on the helipad.

Organizational Influences. Supervisory practices and the conditions and actions of operators are directly impacted by the decisions made by upper-level management. As such, the fourth and final tier of the HFACS taxonomy examines the impact of organizational influences on failures in a system. Organizational influences are further broken down into three causal categories: (1) resource management, (2) organizational climate, and (3) operational process.

Resource Management. How an organization manages the allocation and maintenance of organizational assets such as human resources, monetary resources, equipment and facilities is critical to maintaining the organization's goals. A breakdown in any of these areas is captured within the category of resource management. Traditionally, decisions based on the management of these resources is based upon two, sometimes conflicting, objectives: (1) safety and (2) on-time, cost-effective operations. While both goals are usually satisfied during times of prosperity, times of fiscal austerity often lead to a give-and-take between both goals. In these situations, excessive cost-cutting can lead to a decreased focus on safety and training, and reduced funding for equipment. As a result, low-cost, less effective alternatives are often utilized. Unfortunately, these alternatives may fail or cause problems for the users, resulting in poor performance.

As an example of the negative effects of poor resource management, consider Wesley Medical Center in Kansas, a facility that has faced a number of medical

malpractice lawsuits. In one case, Ms. Holt checked in to the hospital for an elective induction of labor. However, her baby girl who was born limp, pale, and without spontaneous respirations now suffers from permanent brain damage. Ms. Holt's condition mandated a one-on-one nursing ratio throughout her entire hospital stay, however because of staffing shortages she did not receive this care.

In another case at the same hospital, Becky Hartman's mother died during heart failure because the hospital was too understaffed to properly assist her. Hartman brought her 61-year-old mother, Shirley Keck to the hospital when she had difficulty breathing. For seven hours, Hartman watched as her mother's condition deteriorated and tried to get help from the nurses. However, because the primary nurse was overburdened with 20 patients (several more than the hospital's own guidelines) she did not have time to observe Keck until she was being resuscitated after a heart attack. Unfortunately, as a result, Keck suffered brain damage and was paralyzed (Stampalia, 2006).

Organizational climate. The category of organizational climate refers to the working atmosphere within the organization and allows investigators to consider how the vision of the organization may have adverse effects on operators. The structure, culture and policies of an organization are all important variables related to its climate. The structure of an organization is reflected in the chain-of-command, designation of authority, communication networks, and accountability of actions. An organization's culture refers to the unofficial or unspoken values, beliefs, attitudes, rules, and customs within an organization. Policies can be described as official standards that guide management's decisions for issues such as hiring and firing, promotion, sick leave, and other issues important to the day-to-day business of the organization.

One of the most dangerous hospital cultures is one that encourages the hiding or ignoring of errors. For example, a physician may make an error, that results in an accident, perhaps even the loss of a patient life. However, if the hospital they work within involves the discounting of errors or worse, a blame culture, the physician is likely to feel ashamed when he makes a mistake and may even try to cover it up. Instead, the ideal environment involves a strong safety culture that promotes the idea that errors are more often the result of poor systems rather than bad caregivers. Here, physicians and nurses may be more likely to report errors, allowing others to learn from, and hopefully reduce the likelihood of committing another in the future.

Operational process. The last category of organizational influences involves the operational process. This category is reserved for issues surrounding the formal practice by which the vision of an organization is carried out. Issues associated with the operational process involve the formation and use of standard operating procedures, formal methods for maintaining oversight between the workforce and management, organizational tempo, time pressures, and work schedules. Any of these factors have the ability to adversely impact safety in an organization. While most organizations have formal procedures in place to address these factors, some do not. Further, not all organizations utilize anonymous reporting systems and safety audits as a means to actively monitor issues. Because of this supervisors and managers are often unaware that any problem exists before the accident occurs.

Consider for example, a new CEO who is hired at a large medical center. He recently heard of a new technological system that will help to keep track of lost equipment. Because the hospital has historically been known to lose track of expensive

equipment, he finds it absolutely imperative that the new system is integrated hospital-wide. Because of his rules to use the new system, employees immediately start interacting with it. However, within a short period of time, the nurses, doctors and technicians are confused with how the system works, they are making errors because they did not receive proper training, nor do they have experience using the system. Further, managing these issues distracts them from their time with their patients. While the intentions of the CEO were not malicious, he did not consider the results of his actions, which in turn may have caused patient harm.

Framework background. It is important to note that the HFACS framework was not originally created as a method for identifying latent failures. The framework was specifically developed to define both the latent and active failures implicated in Reason's "Swiss cheese" model so it could be used as an accident investigation and analysis tool. Since its development in 2003, HFACS has received considerable attention and has been employed in a variety of industrial settings such as aviation (Li & Harris, 2006), mining (Patterson & Shappell, 2010), maritime (Chen et al., 2013), rail (Reinach & Viale, 2006), and medicine (ElBardissi et al., 2007). A handful of studies involving HFACS have included reliability measures during its use (Li & Harris, 2006; Wiegmann & Shappell, 2001; Olsen, 2011; Olsen & Shorrock, 2010) however there has been disagreement upon its reliability.

One of the fundamental applications of HFACS involves the classification of incident/accident causal factors into the HFACS causal categories. This classification or coding process is often performed on pre-existing causal factors associated with events that were not originally investigated using HFACS. Rather, the HFACS framework is

applied post hoc in attempt to identify meaningful trends in the human causal factors that were not apparent in the original data structure. The reliability of this HFACS coding process impacts the subsequent validity and utility of the HFACS output. If more than one person codes the same causal factors differently, or if the coding results vary for the same person over time, the final results become suspect. Decisions based on such analyses may therefore lead to onerous comparisons across industries and produce ineffective mitigation/prevention plans that have little meaningful impact on reducing risk or improving system safety.

Two recent studies (Cohen et al., 2015; Ergai et al., 2015) specifically investigated HFACS reliability and found that overall it is a reliable system. These studies and their findings will be discussed in more detail below.

Examining HFACS Reliability. (*adapted from Cohen et al., 2015 and Ergai et al., 2016 see Appendix B and C*). The use of HFACS involves the identification and subsequent classification of causal factors into categories based on their presumed underlying etiology. Although the methodology has been widely implemented, the results of research investigating the reliability of the taxonomy are mixed, with some studies showing very high levels of reliability (Li & Harris, 2006; Wiegmann & Shappell, 2001), while others have shown moderate or lower levels (Olsen, 2011; Olsen & Shorrock, 2010).

Reliability, in the present context, involves the degree to which results from an instrument, such as a framework for classifying accident causal factors like HFACS, are consistent or replicable (Carmines & Zeller, 1989). Reliability is crucial for ensuring consensus and consistency and is a vital foundation for establishing the validity of such

analyses (Wallace & Ross, 2006). The purpose of this section therefore, is to review the literature regarding HFACS to summarize previous findings and identify factors that either enhance or detract from the system's reliability. Implications for ensuring the reliability of HFACS as an accident analysis tool will also be discussed.

The description above (see Table 1) represents the HFACS methodology in its most common and fundamental form as described by Wiegmann & Shappell (2003). Over the last decade, individuals have modified the existing approach to make it more applicable to particular industries and/or domains (Chen, all, Davies, Yang, Wang & Chou, 2013; O'Connor, 2008; Olsen & Shorrock, 2010). Some derivatives of the HFACS framework include small changes (e.g. changing a category name) whereas others involve more significant changes (e.g. the addition of a new tier to the framework). Examples of derivatives include Department of Defense HFACS (DoD-HFACS), HFACS for maritime accidents (HFACS-MA), HFACS for mining (HFACS-MI), and HFACS for the Australian Defense Force (HFACS-ADF).

Some derivatives of the approach, such as DoD-HFACS, contain additional sub-levels of classification that further define the causal factor (O'Connor, 2008). This deeper level of analysis represents classifications at the exemplar or "nanocode" level. These nanocodes represent specific forms of the overarching causal category observed within a particular industry. For example, within the causal category of skill-based errors, DOD-HFACS includes the following nanocodes: inadvertent operation, checklist error, procedural error, over control/under control, breakdown in visual scan, and inadequate anti-G training maneuver. Note however, that nanocodes are specific to an industry and

typically created by individual organizations. They are not part of the original HFACS methodology.

Inter-rater and Intra-rater reliability: Practitioners use the HFACS framework to systematically examine the underlying causal factors of an adverse event; the classification of these causal factors is referred to as “coding.” Consistency of the classifications is measured to determine reliability, and this is typically done in two ways: assessing consistency “between” raters or “within” raters. Inter-rater reliability is examined when consistency is measured between different raters during the same time period. In contrast, intra-rater reliability is assessed when consistency is measured within the same rater during different time periods.

There are a variety of methods for measuring these two types of reliability (see Table 2). Arguably the most common measurement is percent agreement (PA) amongst raters. Other, more stringent, measures of reliability that account for chance agreement probabilities are Krippendorff’s Alpha (α), Cohen’s Kappa (K), Fleiss’ Kappa (K_F) and Free-marginal Multirater Kappa (K_{free}). Krippendorff’s Alpha is used to assess both inter-rater and intra-rater reliability, as it can be used for multiple coders; nominal, ordinal, interval and ratio data; and small sample sizes (Hayes & Krippendorff, 2007). Values of (α) range between 0.0 and 1.0 where values of 0.80 and above indicate excellent reliability, and values greater than 0.60 indicate substantial agreement (Landis & Koch, 1977). Some measurements are reserved for inter-rater analysis only. For example, Cohen’s Kappa (K) is used to examine inter-rater reliability in nominal data and ranges from -1.0 to 1.0, Fleiss’ Kappa (K_F), like K tests for inter-rater reliability and measures nominal data on a scale of 0.0-1.0, and free-marginal Multi-Rater Kappa $K_{[Free]}$ is an

alternative to K_F and is used in situations where the rater is unaware of the number of cases to be distributed into each category.

Table 2. Measures of reliability and associated values

Percent Agreement (PA)		Cohen's Kappa (K)/ Fliess' Kappa (K_F)		Krippendorff's Alpha (α)		Free-marginal Multirater Kappa ($K_{[free]}$)	
Value	Conclusion	Value	Conclusion	Value	Conclusion	Value	Conclusion
70% - 100%	Reliable	> 0.80	Reliable	0.80-1.0	Reliable	> 0.80	Reliable
60%-70%	Moderately Reliable	0.60-0.80	Substantially Reliable	0.667-0.80	Tentatively Reliable	0.60-0.80	Substantially Reliability
0-60%	Unreliable	0.40-0.60	Moderately Reliable	0-0.667	Unreliable	0.40-0.60	Moderately Reliable

Factors that Impact Reliability: There are a number of factors that could impact the reliability of classification and coding tasks. For example, the conceptual distinctiveness, definition clarity, and number of categories are all variables that can impact reliability. In general, reliability becomes degraded as the conceptual similarity and number of categories increases and/or the clarity of category definition decreases (Wallce & Ross, 2006). The number of items to be classified can also influence reliability, with a tendency towards improved reliability with a larger number of items due to a reduction in the impact of random error (Button & Ioannidis, Mokrysz, Nosek, Flint et al., 2013). Finally, intra-rater and inter-rater reliability are also contingent upon adequate coder training as well as task standardization (Rousson, Gasser & Siefert, 2002).

Task standardization is particularly of concern with the classification of accident/incident data. For instance, in some cases, raters are provided with causal factors relating to an incident or accident and are asked to classify them using HFACS. However, in other cases the casual factors are not provided, rather participants are given a number of accident/incident reports and are asked to first derive the causal factors from narratives.

The derived causal factors are then classified using the categories provided by HFACS. This latter, two-stage approach may also yield a higher variability and reduces the reliability among raters.

Cohen and colleagues (2015) did a major review on the topic, identifying 111 studies conducting involving the use of HFACS since it was first published in the open literature in 2001. In this study, inclusion criteria for further analysis required that all manuscripts be published in peer-reviewed published journals. This criterion resulted in the exclusion of 73 papers, which in most cases were similar or earlier versions of the peer-reviewed manuscripts that were in the form of laboratory technical reports or conference proceedings. The search was further restricted to include only articles that reported a reliability measure associated with the data coding process; thereby, eliminating an additional 58 papers. Consequently, 14 studies were included in the current analysis.

The 14 published HFACS reliability studies are presented in Table 3. The Table presents those studies that reported reliability measures for the prototypical HFACS framework and those of its derivatives, such as the Department of Defense HFACS (DOD-HFACS), HFACS for the Australian Defense Force (HFACS-ADF) and HFACS maintenance extension (HFACS-ME).

Table 3. Articles that reported reliability indices of HFACS

Authors/Year	Domain	HFACS version	# Raters	Training Involved	Data set utilized for coding	# Events Coded	Inter- Rater Reliability
Wiegmann & Shappell, 2001	Aviation-Commercial	HFACS, 2001	2	Not reported	Retrospective accident reports	319	Reliable
Gaur, 2005	Aviation-Civil	HFACS, 2003	2	Not reported	Retrospective accident reports	153	Reliable
Li & Harris, 2006	Aviation-Military	HFACS, 2003	2	10 didactic hours	Retrospective accident descriptions	1762	Reliable/ Moderately reliable
Shappell et al., 2006	Aviation - General	HFACS, 2003	6	16 didactic hours	Retrospective accident reports	2210	Reliable

Baysari, McIntosh & Wilson, 2008	Other Transport - Railroad	HFACS, 2003	3	“brief overview” by author & self-paced training	Retrospective incident and accident reports	360	Moderately Reliable
Lenne, Ashby & Fitzharris, 2008	Aviation-General	HFACS, 2003	3	training package ¹	Retrospective GA insurance claims	Not reported	Reliable
O'Connor, 2008	Aviation-Military	DoD-HFACS	123	2 didactic hours	Retrospective US Navy mishaps	47	Reliable
Li, Harris & Yu, 2008	Aviation-Civil	HFACS, 2003	2	3 didactic half-day modules	Retrospective civil aviation accidents	330	Substantially Reliable/Reliable
Baysari, et al., 2009	Other Transport - Railroad	HFACS, 2003	6	“brief overview” by author & self-paced training	Retrospective incident and accident reports	144	Reliable
Rashid, Place & Braithwaite, 2010	Aviation-Civil	HFACS-ME	2	Not reported	Retrospective accident and incident reports	197	Reliable/Substantially Reliable
O'Connor, Walliser & Philips, 2010	Aviation-Military	DoD-HFACS	22	4 didactic hours	Retrospective aviation incident report	147	Reliable
Olsen & Shorrock, 2010*	Aviation-ATC	HFACS-ADF	#1:11, #2: 1, #3: 4	Not reported	Retrospective incident report	Not reported	Unreliable **Unreliable
O'Connor & Walker, 2011	Aviation - Military	DoD-HFACS	204	2 didactic hours	Retrospective mishap reports	Not reported	Substantially Reliable
Olsen, 2011	Aviation-ATC	HFACS, 2003	7	Self-paced training workbook	Retrospective incident reports	Not reported	Moderately reliable / Unreliable

*These studies also reported intra-rater reliability, **Intra-rater reliability reported

As seen in Table 3, a majority of the studies reported acceptable values of reliability per the standards outlined above. The studies vary considerably based on the HFACS version used (nine utilized the original HFACS framework, while five utilized HFACS derivatives), domain the framework was applied to, the training involved (most had two hours or more of didactic training on HFACS), and causal factors coded (range: 47-2,210 causal factors). Although each of the studies reported the overall observed reliability, examining the reliability was often secondary to the main objective. In other words, many of the studies reported reliability for quality purposes prior to subsequent analyses, whereas others specifically assessed reliability as the primary purpose of the research.

Of the 14 studies reviewed, only six were dedicated to specifically examining the reliability of the framework (see Table 4). The studies varied considerably in terms of the specific reliability measure used to analyze the data. The most commonly used measure was percent agreement (PA) however each of the more stringent measures represented in Table II were used in at least one of the above studies. In general, the majority of the studies reported adequate levels of reliability ranging from substantially reliable to reliable reports (Li & Harris, 2008; O'Connor, 2008; O'Connor & Walker, 2011).

Table 4. Comparison of HFACS reliability-specific studies reported in the literature

Authors/Year	HFACS version	# Raters	Data set utilized for coding	# Causal Factors	Reliability Measure Used	Inter-rater Reliability	Intra-Rater Reliability
Li & Harris, 2006	HFACS, 2003	2	Retrospective accident descriptions	1762	Percent Agreement (PA), Cohen's Kappa (K)	Avg PA: 88.8 % Avg K: 0.67	-
O'Connor, 2008	DoD-HFACS	123	Retrospective US Navy mishaps	47	Percent Agreement (PA)	Fixed Wing Avg PA: 77.8%, Rotary Wing Avg PA: 78.8%	-
O'Connor, Walliser & Philips, 2010 *	DoD-HFACS	22	Retrospective aviation incident report	147	Multi-rater Kappa Free (Kfree)	Avg Kfree (nanocodes): 0.76	-
Olsen & Shorrock, 2010	HFACS-ADF	#1: 11, #2: 1, #3: 4	Retrospective incident report	Not reported	Percent Agreement (PA)	#1 PA: 40% #2 PA: 40.1%	#3 PA: 44.6%
O'Connor & Walker, 2011	DoD-HFACS	204	Retrospective mishap reports	Not reported	Multi-rater Kappa Free (Kfree)	Helicopter Avg. Kfree: 0.58 Tacair avg. Kfree: 0.69	-
Olsen, 2011	HFACS, 2003	7	Retrospective incident reports	Not reported	Percent Agreement (PA)	ATCO group PA: 36.1% HF specialist group PA: 34.5%	-

**This study reported reliability at the "nanocode" level*

Two of the six studies (Olsen & Shorrock, 2010; Olsen, 2011) found the HFACS framework to be unreliable. Olsen & Shorrock (2010) examined both inter-rater and

intra-rater reliability at the category level and reported unreliable levels of reliability in terms of percent agreement (inter-rater reliability PA: 40%; intra-rater reliability PA: 44.6%). Olsen (Olsen, 2011) examined inter-rater reliability at the category level using two groups of participants (ATCO group and HF Specialist group) and also reported unreliable levels of reliability with respect to percent agreement (ATCO PA: 36.1%, HF Specialist PA: 34.5%).

The studies were diverse on several other factors listed (see Table 6). Three different versions of HFACS were used between the six papers: two used the original HFACS framework from Shappell & Wiegmann (27), three used DoD-HFACS and one used HFACS-ADF. The amount of training involved ranged from no training to two days of hands on instruction. However, none of the studies reported the experience level of those providing the training. There was also substantial variation regarding the number of causal factors coded (range: 47-1,762 causal factors).

Overall, a majority of the peer reviewed articles reported acceptable levels of reliability for HFACS, even when the most stringent metrics were used. Nevertheless, there was considerable variability across studies and two studies reported less than favorable reliability. An examination of those studies whose primary purpose was to assess reliability, revealed several factors known to affect the reliability of classification tasks including training, sample size, and category variability; each are discussed in turn below.

The amount of time dedicated to HFACS training varied considerably across the studies identified in this review. Participants in some studies received no formal training (e.g. see Olson & Shorrock, 2010); whereas, others received elaborate training over

multiple days (e.g. see Li & Harris, 2006). Moreover, only a few of the studies sufficiently explained the specifics of the training provided to participants, raising questions surrounding the quality of training as well as the proficiency of the instructors.

Not surprising, the studies that reported lower values of HFACS reliability (Olsen & Shorrock, 2010; Olsen, 2011) also reported the least amount of training. These findings are similar to those of studies that investigate rater variability and tool consistency (Barrett, 2001; Congdon & McQueen, 2000; Bayzari McIntosh & Wilson; 2008). For example, Weigle (1998) found that training was a substantial element affecting reliability, where levels were higher for trained, experienced coders than for untrained, inexperienced coders. Further, Baysari et al., (2001) explained that hands-on training with repeated practice problems is more likely to result in higher levels of reliability than a study that does not use the same methodology.

In general, studies reporting lower levels of reliability provided fewer causal factors for raters to categorize using the HFACS framework. Unfortunately, only three of the six reliability-specific studies reported the number of causal factors used for classification. Notably however, the study with the largest number of reported categorized causal factors found the methodology to be reliable. The raters in the Li and Harris' (2006) study, coded 1,762 causal factors and reported an average percent agreement of 88.8% and a Cohen's K of 0.67. The larger number of coded causal factors provides for practice over time and can reduce the impact that random coding error has on reliability. This is not to say that larger numbers will always ensure high levels of reliability, since systematic error can degrade reliability as numbers increase. Likewise, small numbers may not always compromise reliability. O'Connor (2008) reported

generally high levels of HFACS reliability despite using a small sample. It should be noted, however, that O'Connor (2008) did provide participants with two hours of hands on HFACS training.

There are several adaptations of the HFACS methodology used across industries, including the Department of Defense HFACS (DOD-HFACS), HFACS for the Australian Defense Force (HFACS-ADF), HFACS for air traffic control (HFACS-ATC), HFACS for aircraft maintenance (HFACS-ME), and HFACS for mining (HFACS-MI). Classifying cases of human error using a derivative of HFACS rather than the original framework, is not only more challenging for the coder but can result in lower levels of reliability. Research conducted by O'Connor specifically examined the reliability of DOD-HFACS for classifying incident data using nanocodes (O'Connor, 2008; O'Connor, Walliser & Philips, 2010; O'Connor & Walker, 2011). He explains, "The main difference between HFACS and DOD-HFACS is the inclusion of an additional level of fine-grain classification. Each DOD-HFACS category has between 1 and 16 associated nanocodes" (O'Connor, 2008, p. 599).

It is important to recognize that the O'Connor studies examined the reliability of a framework that has a total of 147 potential causal factor classifications when the nanocodes are included, rather than the 19 causal categories associated with the original HFACS framework. The additional number of possible categories potentially increases the level of ambiguity among subclass or nanocodes and places higher cognitive demands on coders during the classification task. Consequently, modest reliability values for the HFACS-DOD generally occur at the nanocode level, particularly when combined with limited participant training (O'Connor, Walliser, Philips; 2010).

Overall, the authors found The Human Factors Analysis and Classification System to be a reliable approach for classifying accident causal factors. Based on the data presented here, the majority of the 14 peer-reviewed journal articles identified in this paper reported acceptable levels of inter-rater and intra-rater reliability even when the most stringent metrics were used. Reliability levels were generally highest when training was provided, a large number of causal factors were coded, and deviations from the core components of the original framework were minimized.

Unfortunately, many of these conclusions are speculative given that many of the articles reviewed by the authors lacked sufficient detail to discern the exact nature of the conditions under which the coding was conducted. To further address these issues, Ergai and colleagues (2016) assessed the reliability of the HFACS framework as a general accident analysis tool using a large number of trained coders and multiple real-world accident causal factors from a variety of industries.

In this study, one hundred and twenty-five safety professionals from a variety of industries (e.g. aviation, mining, medicine and manufacturing) were recruited from a series of two-day training workshops designed and provided by the developers of HFACS (S.S. & D.W.). Trainers with expertise in both human factors and the HFACS methodology provided participants two days of intensive HFACS training (see Ergai, et al., 2016 for details).

Five causal factors associated with each of the 19 HFACS causal categories ($n = 95$) were extracted from actual accident reports from the National Transportation Safety Board (NTSB), Occupational Safety and Health Administration (OSHA), and other HFACS accident databases such as aviation, maintenance, food service, healthcare,

mining and lodging. Each causal factor was selected for clarity and was presented to participants in the format in which it was reported in the accident database. Some causal factors were modified for grammatical purposes only, and in a manner that did not affect the content of the statements. This approach was adopted to enhance content and external validity of the causal factor statements. Given the original accident reports were not coded using HFACS, a scoring key of correct answers was created for each of the 95 causal statements. Correct answers were determined a priori by consensus among the HFACS expert research team.

The study required participants to fill out two assessments: the first given at the conclusion of the second day of training, the second given 14 days later. During the first session, the 95 causal statements were presented via the Internet and participants completed the coding on their own personal laptop computer or Tablet. Two weeks later, participants were emailed a link to the online form and the same 95 casual factors were randomly presented again for coding. There was no time limit for completing this coding process and participants were allowed to use notes and reference materials to help complete the task. This process was used to ensure that the assessment was not simply a memory test but rather mimicked the coding process commonly utilized in the real-world.

Data from 125 participants who completed the first coding activity, were collected and used for the inter-rater reliability analysis. Of these 125 participants, 59 (47%) completed the activity two weeks following the training. Therefore, data from these 59 participants were included in the intra-rater reliability analysis. Krippendorff's Alpha was used to analyze the data.

The overall inter-rater reliability for both the HFACS tier and the causal category levels were computed and are presented in Table 5. When calculated across the four tier levels (*unsafe acts, preconditions for unsafe acts, unsafe supervision and organizational influences*), a value of $\alpha = 0.79$ was obtained for Krippendorff's Alpha. When calculated across the 19 categories, the Krippendorff's Alpha value was somewhat lower ($\alpha = 0.67$) indicating that coders were less likely to agree on the particular category a specific causal factor belonged within each level.

Table 5. Overall α for each HFACS Tier and Category; Inter-rater Reliability

HFACS Tier	HFACS Category	Average α
Unsafe Acts	Skill Based Error	0.56
	Decision Error	0.46
	Perceptual Error	0.72
	Routine Violation	0.76
	Exceptional Violation	0.63
Preconditions for Unsafe Acts	Physical Environment	0.82
	Technological Environment	0.65
	Adverse Mental State	0.68
	Adverse Physiological State	0.63
	Physical / Mental Limitations	0.73
	Communication Coordination & Planning	0.78
	Fitness for Duty	0.73
Unsafe Supervision	Inadequate Supervision	0.51
	Planned Inappropriate Operations	0.49
	Failed To Correct a Known Problem	0.82
	Supervisory Violation	0.53
Organizational Influences	Resource Management	0.62
	Organizational Climate	0.80
	Organizational Process	0.69

Krippendorff's Alpha values at the causal category level were lower than at the tier level (tier: $\alpha = 0.73$ to 0.82 ; category: $\alpha = 0.46$ to 0.82). The highest reliabilities were *physical environment* ($\alpha = 0.82$) and *organizational climate* ($\alpha = 0.80$). *Skill based error*

($\alpha = 0.56$), *decision error* ($\alpha = 0.46$), *inadequate supervision* ($\alpha = 0.51$), *planned inappropriate operations* ($\alpha = 0.49$) and *supervisory violation* ($\alpha = 0.53$) resulted in the lowest reliability levels.

Upon closer inspection of the 95 causal factors the analysis identified six in which less than 50% of the coders properly identified the causal category. Many of these cases included compound causal factors, which could explain why inter-rater reliability was particularly low for certain categories. Causal factor 92, for example was: “The electrical operator got distracted by an external noise and forgot to take readings on the main transformers”. Most participants classified the code as *skill based error*; however, many coders selected *physical environment* because of the term “external noise” or *adverse mental state* because of the word “distracted”. Because factors like these were included, a number of people classified them differently, which could explain why there were such low levels of reliability for some of the categories.

Of the original 125 participants who took the survey on the first day, 59 (47%) participants returned the survey two weeks later. Intra-rater reliability was determined by calculating Krippendorff’s Alpha individually for each of the 59 participants who participated in both sessions. The intra-rater agreement results of these measures were tabulated for each participant. Krippendorff’s Alpha calculated across all 59 participants at the HFACS tier level and at the HFACS category level and are presented in Table 6.

Table 6. Average α for each HFACS Tier and Category; Intra-rater reliability

HFACS Tier	HFACS Category	Average α
Unsafe Acts	Skill Based Error	0.66
	Decision Error	0.57
	Perceptual Error	0.82
	Routine Violation	0.82
	Exceptional Violation	0.75

Preconditions for Unsafe Acts	0.87
Physical Environment	0.87
Technological Environment	0.75
Adverse Mental State	0.78
Adverse Physiological State	0.68
Physical / Mental Limitations	0.82
Communication Coordination and Planning	0.83
Fitness for Duty	0.73
Unsafe Supervision	0.83
Inadequate Supervision	0.66
Planned Inappropriate Operations	0.64
Failed to Correct a Known Problem	0.85
Supervisory Violation	0.62
Organizational Influences	0.87
Resource Management	0.75
Organizational Climate	0.89
Organizational Process	0.80

The overall average intra-rater agreement values were slightly yet consistently higher than those observed for inter-rater reliability. However, a similar pattern emerged; Krippendorff's Alpha values were generally higher at the tier level in comparison with the HFACS category levels (tiers range $\alpha = 0.83 - 0.88$; category range $\alpha = 0.57 - 0.89$). With the exception of *decision error* ($\alpha = 0.57$) intra-rater reliability was found to be above $\alpha = 0.62$. Note also that because the same causal factors were used in both sessions, those six causal factors that had low reliability in the first session subsequently had low reliability in the second session, and may account for the lower intra-rater reliability seen among *decision errors* and *supervisory violations*.

In this study, HFACS was found to be generally reliable at both the tier and category levels. That is, the coders generally agreed upon which level of the HFACS hierarchy a causal factor belonged and were consistent upon retesting. Albeit, they had a slightly more difficult time agreeing upon the particular category a specific causal factor belonged within each level. The general reduction of reliability levels within HFACS

from the tier level to the category level is not uncommon (e.g. Olsen, 2011). Research in other domains has shown that as the number of categories increases, reliability decreases. For example, Gwent (2010) demonstrated through a Monte-Carlo experiment, reliability decreases as the number of categories increases, while the number of causal factors is kept constant.

Although the data presented here suggest that HFACS is a reliable framework for classifying human factors associated with accidents and incidents, inter- and intra-rater reliability was not maximized. Clearly there are opportunities for improving the reliability of the HFACS coding process including: improving training quality, further clarifying causal factors, and coder selection to name a few. Nonetheless, HFACS remains one of the most actively used frameworks for accident and incident investigation.

Research Questions

Despite the established reliability of HFACS for investigating retrospective accident reports, explorations of HFACS reliability for coding data other than accident causal factors have been limited. More specifically, HFACS has never been used to identify threats to safety from observational healthcare data. Unlike accident investigations that are reactive, other approaches to safety can be proactive, attempting to identify threats to safety before accidents happen. Sources of such proactive data often include the direct observation of work-related activities and the identification of conditions that potentially threaten safe performance. As such, the first research question seeks to investigate the following:

Q1: Can HFACS, a tool developed originally for investigation of accident and incident data, be reliably applied to observational data within

healthcare? Here, observational data is used to describe human factors issues that are observed in various healthcare domains.

Perhaps one of the most highly observed areas in clinical medicine is the cardiovascular operating room (CVOR). This environment is arguably one of the most demanding, challenging, and complex environments in the healthcare domain. In this setting, the medical team must work together and interact with a number of technologically advanced tools while coordinating care for the patient in the room.

While medical teams are composed of highly trained and uniquely qualified healthcare professionals, threats to patient safety can and still do occur. It has been noted that cardiac surgery is particularly predisposed to pitfalls because it features multiple specialties, close coupling of concurrent tasks, changing plans, and high workload (Carthey, et al., 2001). It is not surprising then, that direct observation analysis in cardiac surgery has provided evidence that small errors have the potential to influence the outcome of a given patient (de Leval, et al., 2000).

Observational studies in the CVOR have identified several factors that may represent threats to patient safety. Palmer and colleagues (2013) found that observations associated with the physical layout were most prominent, followed by general interruptions, usability concerns and communication issues. Similarly, Wiegmann and colleagues (2006) identified 341 events during 31 cardiac operations. Of these disruptions, teamwork/communication problems accounted for the most events followed by interruptions. A more recent investigation of the CVOR found most issues to involve interruptions, layout, communication and coordination (Cohen et al., 2016).

With respect to HFACS, the categories that most closely align to the issues identified above involve adverse mental states (e.g., distractions, interruptions), physical environment (e.g., physical layout, layout) and communication, coordination and planning (e.g., communication, coordination, teamwork), which all exist at the preconditions for unsafe act level.

While several studies have investigated HFACS in medicine, none have applied it to certain medical specialties at different facilities (i.e., a CVOR in an academic hospital vs. a CVOR in a non-academic hospital) to investigate differences in populated categories. This information would be valuable to medical practitioners and administrators as it could help them to understand if specific investigations are needed for certain types of facilities (i.e. teaching vs non-teaching) or if the findings are generalizable regardless of facility. Understanding the differences in populated categories can help the healthcare industry to put targeted barriers in place that may protect against specific threats to patient safety.

Although the types of preconditions for unsafe acts populated in any given CVOR are likely to involve adverse mental states, communication, coordination and planning issues, and problems with the physical environment, the distribution of workflow events would be expected to vary based on the facility investigated. For example, in a private academic university hospital, there are often medical students, interns and residents not only observing in the room but also participating. This situation lends itself to overpopulated operating rooms, which can lead to issues involving the physical layout. With this in mind, this dissertation seeks to also answer a second research question:

Q2: Given the answer to Q1 is yes, Q2 will investigate if the utilization of HFACS in two different CVOR environments will result in the population of primarily adverse mental states, physical environment issues and communication, coordination and planning problems. More specifically, will there be a greater number of physical environment issues identified in a private academic hospital compared to a non-academic hospital setting?

Given that the answer to the second research question is “yes”, this research will represent a starting point for establishing the utility of HFACS as a tool for identifying human factors issues in observational healthcare data. However, it is important to note that the first two research questions are examined through similar observation of individual team roles in the CVOR. Several other diverse areas of observation exist within the healthcare realm, and the data collection processes expectedly vary between the types of healthcare domains observed.

While cardiovascular surgery is predisposed to pitfalls due to its complexity (Carthey et al., 2001) it is typically predictable, well planned and conducted in a controlled environment. Although the individuals who work in the CVOR must communicate effectively and work together to accomplish their tasks, they are often separated based on their role-type and responsibilities. Because of its predictability and structured nature, observation in the CVOR allows for data collection based on certain team member areas or specific role-groups. In the following studies, the CVOR team was observed based on role: anesthesiologist, circulating nurse, perfusionist, and surgeon.

Unlike cardiovascular surgery, providers in other treatment domains such as trauma care may not have the luxury of a foreseeable procedure. Specifically, Sarcavic (2009) explains that trauma resuscitation is unpredictable in nature and occurs in a fast-paced, dynamic environment where healthcare professionals must quickly and accurately evaluate and diagnose potentially life-threatening injuries to the patient. This process is information laden and team dependent, relying heavily not only clinical skill but also on the efficiency of the system.

A trauma resuscitation team is generally made up of physicians, nurses, and allied health personnel. The American College of Surgeons Committee on Trauma (2014) explain that a high-level response to a patient who is severely injured would typically include the following individuals or groups of individuals: (1) a general surgeon; (2) an emergency physician; (3) surgical and emergency residents; (4) emergency department nurses; (5) a laboratory technician; (6) a radiology technologist; (7) a critical care nurse; (8) an anesthesiologist or certified registered nurse anesthetist; (9) an operating room nurse; (10) security officers; (11) a chaplain or social worker; and (12) a scribe.

When caring for the traumatically injured trauma patient, time is of the essence and cases are almost always unpredictable, making teamwork that much more important. Harkins (2009) explains that although each of the team members in trauma have individual purpose and responsibility, they must work in concert to pursue the goal of saving the patient. She explains that the whole team is greater than the sum of its parts and if one individual member “drops the ball, it may be a matter of life or death despite the extraordinary efforts of all the other players” (p. 61). As a result, data collection in

trauma care facilities generally involves the observation of the entire team rather than individual roles.

Teams are fundamentally different than individuals or groups. Salas and colleagues (2008) explain that because teamwork involves multiple individuals' taskwork as well as their coordinated efforts, teamwork is more complex than individual performance. This can play into the notion that the observation of teams is different than the observation of individuals. Unlike individual research, the complexity of team's research involves different levels of attention and skill in order to evaluate or diagnose performance (Salas, et al., 2008).

Boquet and colleagues (2016) investigated observations at a level II trauma center and identified communication issues to occur most frequently (28%) followed by interruptions and coordination issues (24% each). Similarly, in another study on disruptions in trauma, Blocker and colleagues (2012) found that unfavorable coordination issues made up 28% of the events and was followed by communication issues at 24%. Likewise, Bergs and colleagues (2005) noted that knowledge transferal during trauma resuscitations was sub optimal. In terms of HFACS categories, the findings noted above would likely translate as communication, coordination and planning issues.

If HFACS can function as a tool for identifying human factors issues in observational healthcare data, it should theoretically be able to classify observations from domains outside of the the CVOR regardless of the collection method or technique used. Therefore, research question three addresses this:

Q3: This research question investigates if the utilization of HFACS to classify data collected in a different setting, using a different data

collection technique would yield expected results. Specifically, this question investigates if HFACS can be used reliably to classify data collected in a trauma care setting, where events were recorded based on how they impacted the entire trauma team rather than by individual team members and if a majority of the issues would be classified as communication, coordination and planning.

CHAPTER 3: STUDY I (CVOR - ACADEMIC HOSPITAL)

Introduction

Whether HFACS can be reliably applied to observational healthcare data is entirely unknown. Therefore, the purpose of chapters three and four was to explore this issue. Data was collected from two hospitals; a academic hospital and a non-academic hospital, producing two different data sets. Three trained analysts then independently classified the data using HFACS and agreements between the analysts were computed. It is important to note that while the data comes from similar operating theatres, involving similar team members, the methodology for collecting each data set was different.

The Cardiovascular Operating Room (CVOR). While the architectural layout of each CVOR can vary between hospitals, most have similar layout features. In Study I, data was collected from a large, metropolitan medical school with over 700 hospital beds and two state of the art surgical suites (see Palmer et al., 2015 for a full description). Each cardiac suite is connected to a sterile CORE. This area is outside of the physical cardiac operating room but houses several important pieces of equipment and different supplies (i.e. ice, surgical wire, replacement aortas, saline, etc.) that the team members (usually circulating nurses) must retrieve during the case. While there are certainly pieces of equipment and supplies within the room, contained in locked 'pixus', these machines cannot house all of the equipment and supplies needed for every case. This may be because of a thermal requirement for saline, because the different rooms share equipment that must be housed in a shared space, or equipment is too large to fit into the individual room.

The layout of the CVOR is usually quite standardized and will be described from a “birds-eye-view” perspective (see Figure 8). The patient is placed on the examination Table in the center of the room. The anesthesia team (see number 1) is found at the top of the bed (near the patient’s head) next to the transthoracic echocardiography instrument (TEE). The circulating nurses typically reside in the half of the room near the patient’s feet. When they are not gathering equipment and supplies for the team, they are typically sitting/standing at a computer/phone completing charting and other room-managing tasks (see number 2). The Perfusion team (see number 3) sits to the right of the patient with their extracorporeal equipment (namely the heart/lung machine). The surgeon (depending on the procedure) stands to the left of the patient (see number 4). Being a academic hospital, observers often saw students in the room either participating or watching.

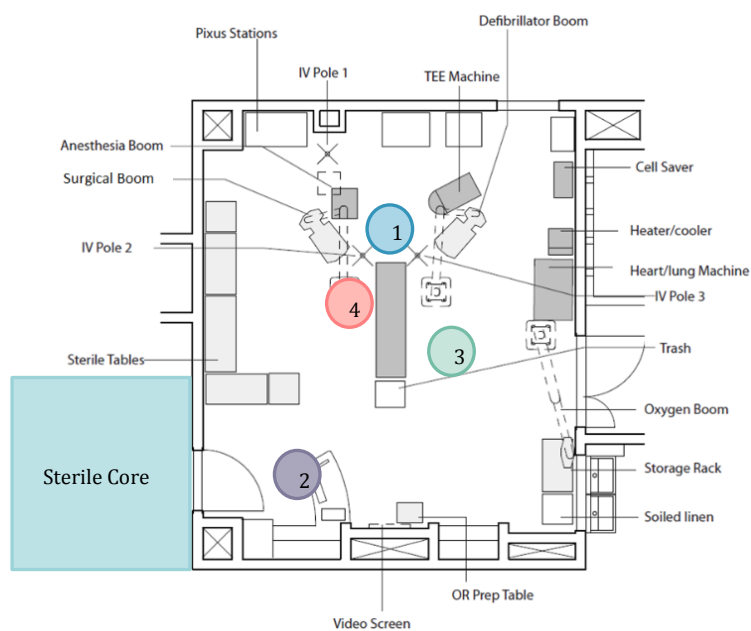


Figure 8. Layout of the Cardiovascular Operating Room (CVOR)

Method

Data Collection

The data-set used for this study was originally collected to identify and analyze workflow disruptions that occurred during cardiac surgery at a medical university (see Cohen et al., 2016). The data was collected from a academic hospital over 4 months (11/21/13 – 2/27/14), across 15 surgical cases, totaling 73.08 hours of observations. Observers were trained but did not use any structured format for documenting workflow disruptions. Each observer was embedded within one of three cardiac team areas (anesthesiology, circulating nursing, and perfusion). Results produced 878 events or workflow disruptions to be coded and classified using HFACS in this study. See Cohen and colleagues (2016) for a detailed description of the data collection procedures.

Coder Training

An expert in human factors and the HFACS methodology provided three coders with two days of general HFACS training. This training included an overview of human error and human factors, an overall description of the HFACS framework, extensive discussion of each causal category including examples of each. The raters participated in hands-on exercises that allowed them to practice classifying generic causal factors within the framework. Following HFACS training, the author provided raters with more detailed instruction on specific cardiovascular OR topics including general CVOR terminology, surgical team member titles, roles and responsibilities, and common procedures and equipment/supplies in the OR (one-hour session).

Data Coding / Classification

In an effort to calibrate the coding process, all three raters classified 100 randomly selected events from the observational dataset into the HFACS framework as a group (one-hour session). After this period, each rater independently classified 50 additional events. The author and the three raters then discussed any disagreements (one-hour session). Following, the raters individually coded the remainder of the data, and later returned to independently code the practice data. Coders individually completed the coding between 10/7/15 and 10/27/15 (20 days) for a data set consisting of 878 workflow disruptions.

Reliability

While percent agreement is arguably the most common method for examining reliability, it does not correct for chance agreement among raters. Another more stringent measure of reliability is Fleiss' Kappa. Fleiss' Kappa tests for inter-rater reliability, and measures nominal data on a scale of 0.0-1.0. Here, values between 0.40 and 0.60 are considered moderately reliable, values between 0.60 and 0.80 are considered substantially reliable and values above 0.80 are reliable (Cohen et al., 2015).

Results

Data Inclusion

The analysis was primarily focused on evaluating coder agreement across the various aspects of the coding process. This was done using three methods. The first method is based on unanimous agreement in which results were considered only from those events in which all three raters had complete agreement on the appropriate code. The second method focuses on majority agreement and considers any event in which at

least two raters agreed (majority) on the appropriate code. Reliability was also evaluated based on pairwise comparison (i.e. rater 1 vs. rater 2, rater 1 vs. rater 3, and rater 2 vs. rater 3) as a quality control measure. This methodology was used as a means to investigate whether any single rater was influencing the reliability.

In all three methodologies, percent agreement was used to analyze the data. Percent agreement can range from 0 to 100%, with agreement of 70% or higher being considered reliable, 60% to 69% being moderately reliable and below 60% being unreliable (Cohen et al., 2015).

In the unanimous method, first, agreement amongst coders as to whether any given event could be coded using HFACS (i.e., inclusion vs. exclusion) was examined. In this case, of the 878 original observations, all three coders agreed that 867 events could be coded using HFACS. Of the 867 “codable” events, all three coders agreed on the allocation of 847 events at the tier level of HFACS. Based on this, there was unanimous agreement that one event was considered an unsafe act, while the remaining 846 events were preconditions for unsafe acts. Of the 847 total agreed upon tier level events, 567 (67%) were unanimously agreed upon at the category level (see Figure 9).

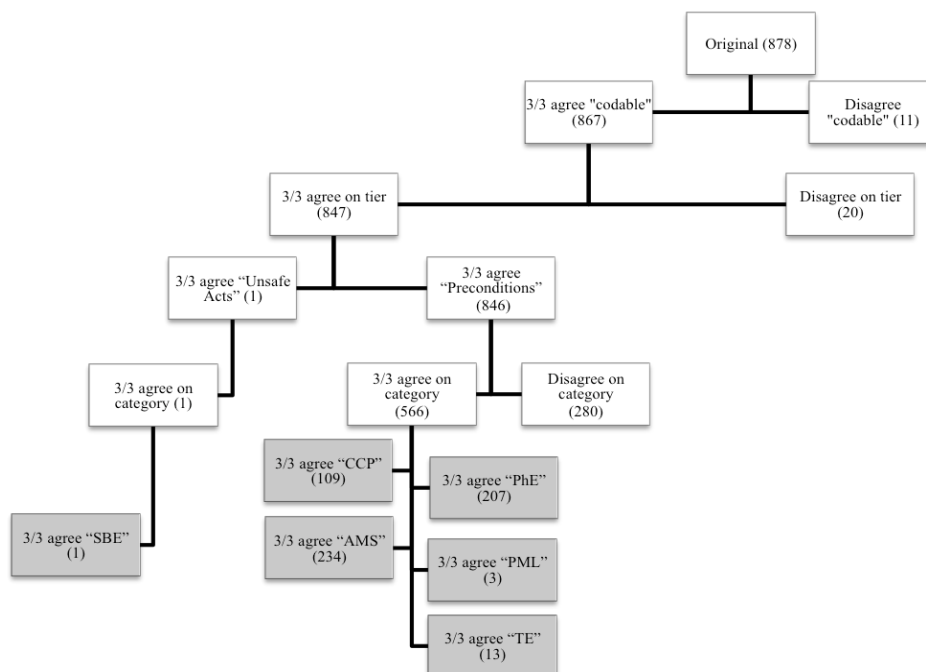


Figure 9. Data-set I line diagram (unanimous method)

Similar to the unanimous method, the first step of the majority method involved investigating how many events were considered “codable” by at least 2 of the 3 coders. Here, coders came to majority that all 878 of the events could be coded using the HFACS framework. Of the “codable” events, coders came to majority regarding their appropriate tier allocation for all 878 events. At this point, coders determined that the majority of the events (875) could be classified as preconditions for unsafe acts, followed by unsafe acts (3). Of the 3 unsafe acts identified, 2/3 coders agreed on the appropriate category for two of the events and disagreed on one event. Of the 875 preconditions for unsafe acts, coders came to majority on 820 of the events and disagreed on 55. Overall majority agreement was 93.6 % (see Figure 10). Following the majority method, coders and researchers met to reconcile those issues in which there was no agreement in the reconciled method (see Figure 11).

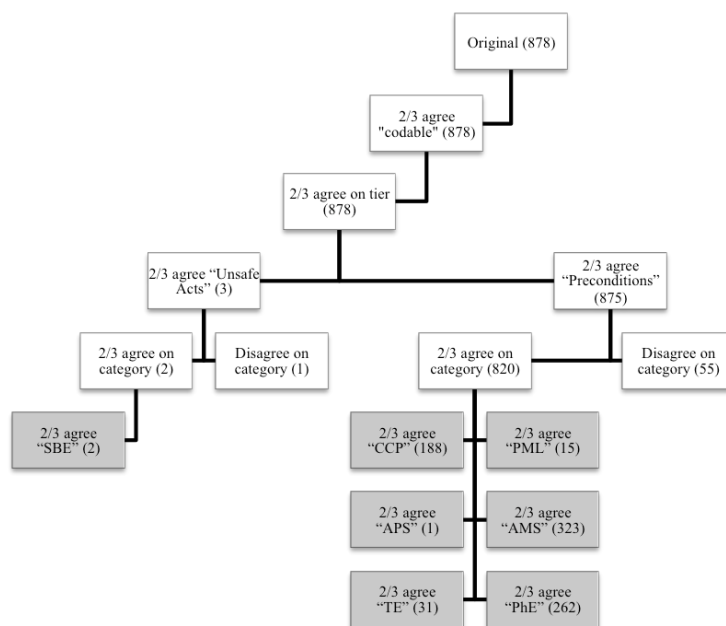


Figure 10. Data-set I line diagram (majority method)

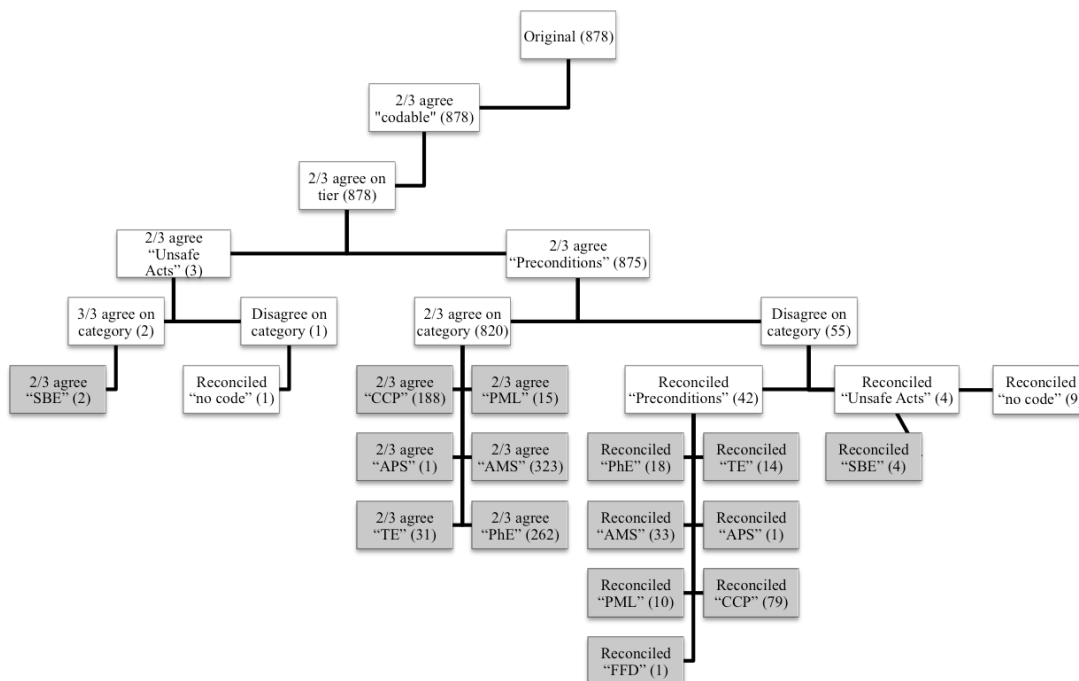


Figure 11. Data-set I line diagram (reconciled method)

Percent agreement was also explored between each possible dyad of the three coders (i.e. 1 vs. 2, 1 vs. 3, and 2 vs. 3). Reliability at the pairwise level was calculated

based on the total number of agreements between raters divided by 878 possible events (see Table 7).

Table 7. Data-set I methodology comparisons at the tier and category level

	Unanimous (3/3 agree)	Majority (2/3 agree)	Reconciled Majority (2/3 + consensus)	Pairwise Comparisons		
				1vs2	1vs3	2vs3
Tier	98%	100%	100%	97%	98%	97%
Category	67%	94%	94% +	77%	73%	73%
Inclusion	567 total 566 preconditions	824 total 822 preconditions	864 total 862 preconditions			

Reliability

Fleiss' Kappa was also used to investigate inter-rater where two Kappa values were calculated depending on the amount of data included. First, an overall Kappa was calculated to investigate inter-rater reliability for all potential events ($n = 878$). This first method investigates how well the three raters agreed on the allocation of a particular event into any of the 19 causal categories represented in HFACS. Here, Fleiss' Kappa showed substantial agreement ($k=0.635$ (95% CI, .611-.659), $p = 0.000$).

Because an overwhelming majority of the data was considered preconditions for unsafe acts in both studies, Kappa was also calculated based on those events that all raters unanimously agreed were "codable" at the preconditions for unsafe acts tier ($n = 846$). This second method investigates how well the three raters agreed on the allocation of a particular event into the seven preconditions for unsafe acts categories represented in HFACS. In this case, Fleiss' Kappa also showed substantial reliability again ($k = 0.660$ (95% CI, .635-.685), $p = 0.000$).

Findings:

As nearly all of the data (99.8% and 99.7% for unanimous and majority methods respectively) was coded as a precondition for unsafe acts, a more fine-grained analysis to

indicate the types of preconditions was conducted. For comparison, Figure 13 depicts results from the unanimous method, the majority method and a reconciled version of the majority method which includes an additional 42 reconciled preconditions events. There were originally 55 events that a majority agreed were preconditions for unsafe acts, however there was no agreement on the particular category it belonged. Of these 55, 42 were reconciled as preconditions, four were reconciled as unsafe acts and nine were determined “uncodable” (See Figure 12).

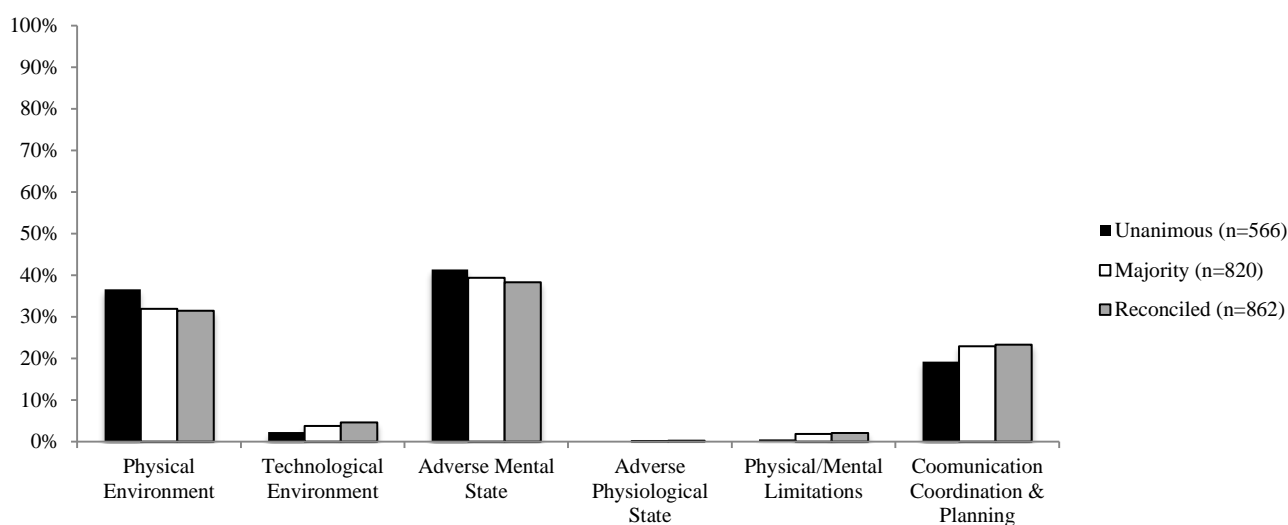


Figure 12. Data-set I unanimous vs. majority vs. reconciled preconditions for unsafe acts

Because all three methods produced such similar results, only the reconciled data (n = 862) are discussed below. Most failures resulted from adverse mental states that involved being distracted or interrupted (38.28%) followed by issues in the physical environment (31.44%) and problems with communication, coordination and planning (23.32%). Other preconditions involved problems with the technological environment (4.64%) and physical/mental limitations (2.09%)

CHAPTER 4: STUDY II (CVOR-NON-ACADEMIC HOSPITAL)

The findings of the prior study indicate that HFACS can be used to classify retrospective observational data in a cardiac operating room. The data set utilized for the aforementioned study was collected from a medical teaching university. While a considerable amount of data was collected and nearly all of it could be analyzed using HFACS, the findings from that study may not be generalizable to all CVORs. In an effort to investigate this, Study II utilized HFACS to classify retrospective observational data from a different data set from another hospital. In this case, a much larger sample of data was used, and was collected from a community hospital located in Orlando, Florida.

The Cardiovascular Operating Room: The description of a general CVOR was discussed during chapter three (see Figure 8). The operating room layout was very similar for Study II; however, it is important to note some differences about the hospital overall. This community hospital houses over 2000 beds and over 16,000 employees serving seven county areas throughout Orlando and central Florida and stands as the second largest hospital in the state (Florida Hospital, 2012). As such, there are a greater number of CVOR suites and more CORE areas than the hospital in Study I. Further, being a non-academic hospital, this environment saw little to no students, unlike that of Study I.

Method

Data Collection: The data-set used for the second study was also collected to identify and analyze disruptions that occurred in a CVOR. However, this data was collected from a non-academic hospital over four months (1/26/15 – 4/30/15), across 25 surgeries totaling 145.04 hours of observation. Observers were experienced with collecting data in the cardiac operating theatre, but again utilized no specific tool or

method for data collection. While similar to the first study in which only two observers could observe a procedure at a given time, here each observer collected workflow disruptions that impacted *two* cardiac team areas. In other words, in each surgery one observer collected data involving the anesthesiologist and surgeon, and the other collected data involving the Perfusionist and circulating nurse. Results produced 4233 observations that were to be coded using HFACS.

Coder Training: The same individuals who coded the data for Study I also coded the data for Study II. (Refer to coder training under “Study I” for a complete description)

Data Coding / Classification: Because the data in data-set II was very similar to data-set I, a brief training period took place but a calibration process did not. Each rater individually coded all 4233 observations. Raters completed the coding within three months (11/5/15 – 1/10/16).

Reliability: Fleiss’ Kappa was used again to investigate inter-rater reliability overall, as well as inter-rater reliability in terms of agreement only at the preconditions for unsafe acts level.

Results

Data Inclusion

The same methods from Study I were used to analyze the reliability of the data. With the unanimous method, of the 4233 original observations, all three coders agreed that 3416 events could be coded using HFACS. Of the 3416 “codable” events, all three coders agreed on the allocation of 3308 events at the tier level of HFACS. Based on this, there was unanimous agreement that 10 events were considered unsafe acts, while the remaining 3298 events were preconditions for unsafe acts. Of the 3288 preconditions,

2411 (73.1%) were unanimously agreed upon at the category level (see Figure 13).

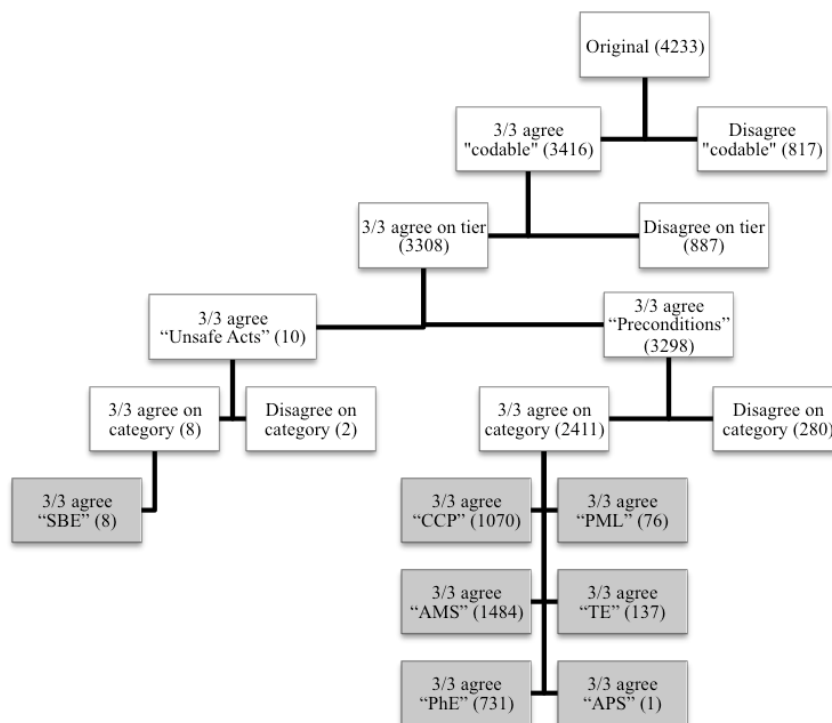


Figure 13. Data-set II line diagram (unanimous method)

In the majority method, 2/3 coders came to agreement that 3789 of the 4233 events could be coded using HFACS. Of the “codable” events, coders came to majority regarding their appropriate tier allocation for 3775 events. At this point, coders determined that the majority (3729) could be classified as preconditions for unsafe acts, followed by unsafe acts (41) and organizational influences (5). Of the 3729 preconditions, coders came to majority as to which category 3499 events belonged (93.8%) (see Figure 14); the remaining 230 events were agreed upon during the reconciliation process (see Figure 15).

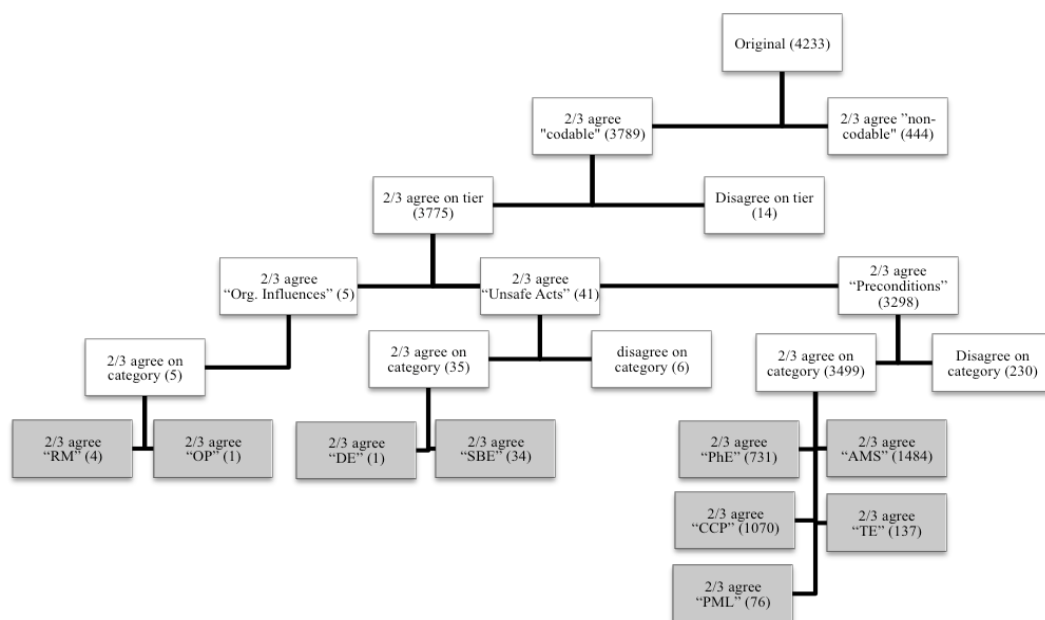


Figure 14. Data-set II line diagram (majority method)

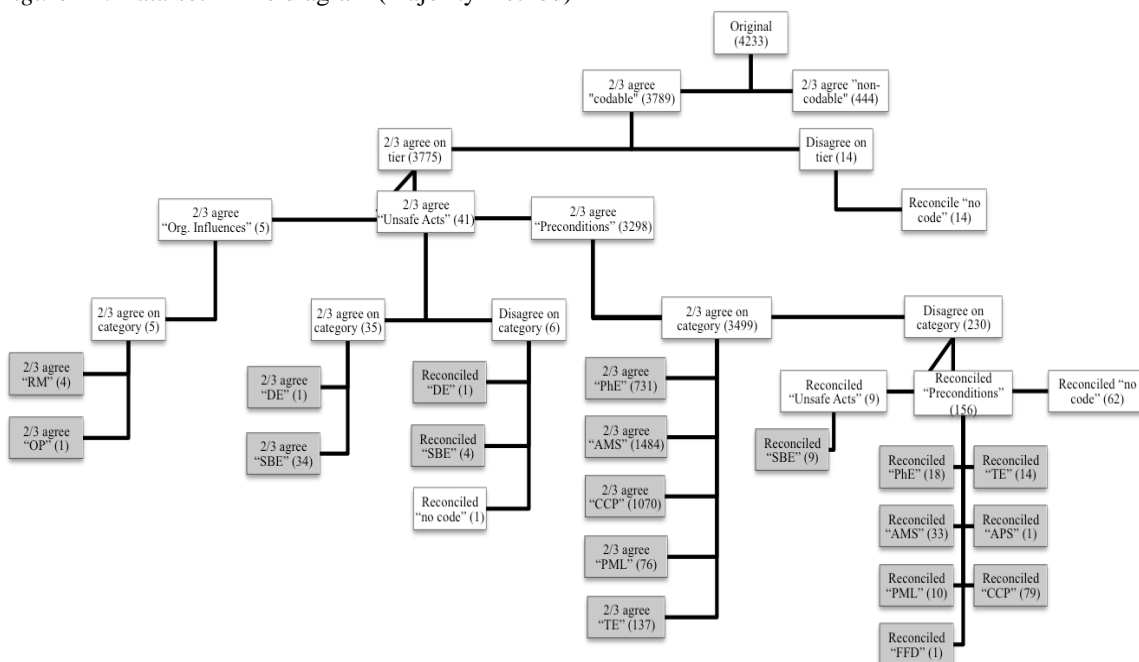


Figure 15. Data-set II line diagram (reconciled method)

Similar to data-set I, we also calculated pairwise comparisons along with the other two methodologies for investigating reliability (see Table 8).

Table 8. Data-set II methodology comparisons at the tier and category levels

	Unanimous (3/3 agree)	Majority (2/3 agree)	Reconciled Majority (2/3 + consensus)	Pairwise Comparisons		
				1vs2	1vs3	2vs3
Tier	97%	99.6%	99.6% +	89%	88%	89%
Category	73%	96%	96% +	76%	71%	73%
Inclusion	2419 total 2411 preconditions	3539 total 3499 preconditions	3709 total 3655 preconditions			

Reliability

Again, Fleiss' Kappa was calculated in two ways to represent a more stringent measure of inter-rater reliability. First an overall Kappa was calculated to investigate inter-rater reliability for all potential events ($n = 4233$). Here, Fleiss' Kappa showed substantial reliability ($k = 0.642$ (95% CI, .633-.652), $p = 0.000$). Because an overwhelming majority of the data was again considered preconditions for unsafe acts, Kappa was also calculated based on those events that all raters unanimously agreed were "codable" at the preconditions for unsafe acts categories represented in HFACS. Here, Fleiss' Kappa showed an increase in reliability ($k = 0.726$ (95% CI, .713-.738), $p = 0.000$).

Findings

Again, nearly all the data fell into the preconditions for unsafe acts category (99.7% and 98.7% for the unanimous and majority methods respectively). Because of similarity in results, data is discussed based on reconciled findings. Here, most failures resulted from adverse mental states (41.50%) followed by issues with communication, coordination and planning (31.44%) and problems in the physical environment (20.49%) Other preconditions involved problems with the technological environment (4.13%) and physical/mental limitations (2.35%) (See Figure 17).

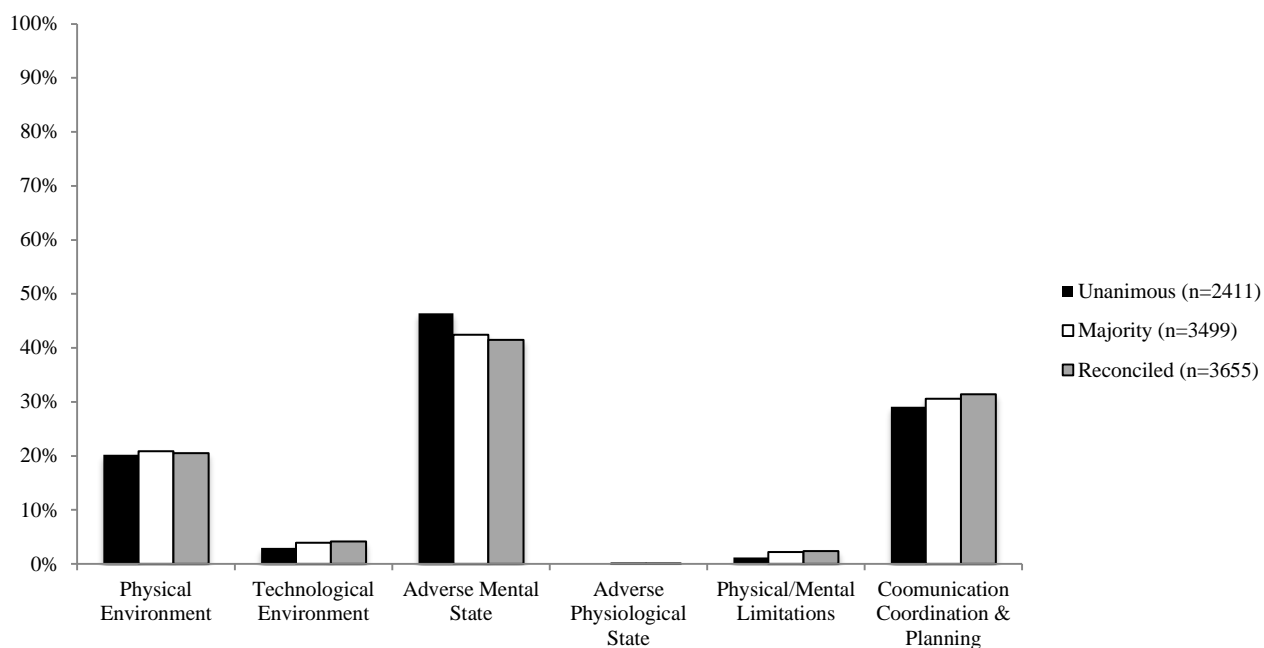


Figure 16. Data-set II unanimous vs. majority vs. reconciled preconditions for unsafe acts

Comparison of Data-set I and Data-set II

Data Comparison

The retrospective data-sets utilized for addressing the second research question were not only different in terms of the hospitals in which they were collected but also in the methodology used for collecting the data. Study I utilized data from an academic hospital where each researcher (two in the room) observed workflow disruptions while embedded within one of three specialties (circulating nurse, perfusionist and anesthesiologist). Study II utilized data from both the academic hospital in which each researcher (two in the room) observed work flow disruptions while observing two of four specialties (circulating nurse and perfusionist or anesthesiologist and surgeon).

Because both sets of data, while inherently different, came from observations of workflow disruptions in the cardiac operating room, it was also of interest to compare the two data sets with respect to the preconditions for unsafe acts findings (see Figure 17).

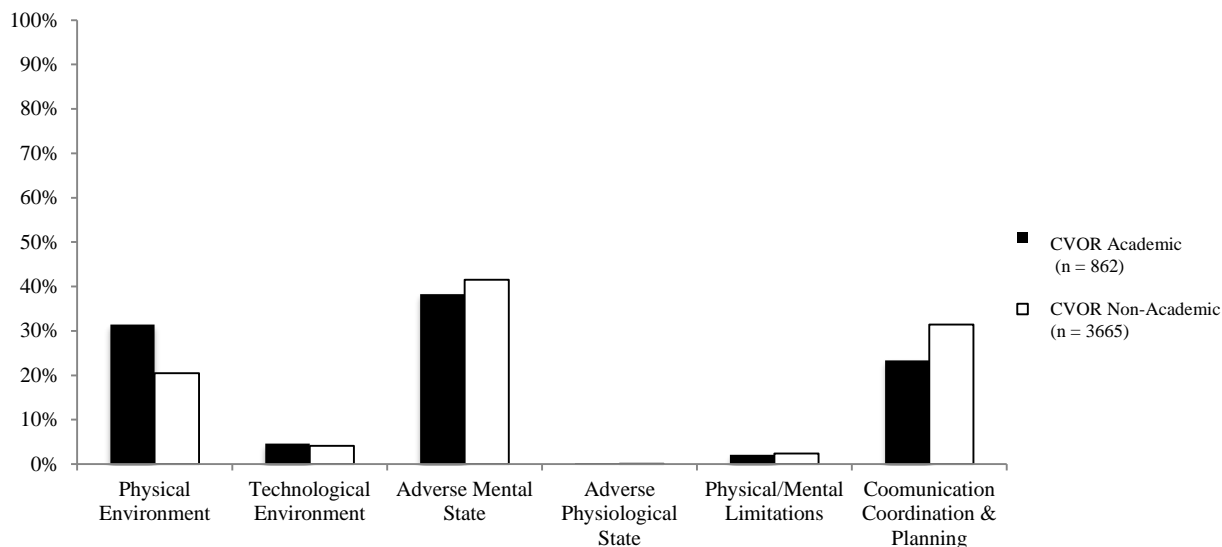


Figure 17. CVOR academic and non-academic comparison of reconciled findings for preconditions for unsafe acts

Chi-square tests (X^2) were used to evaluate differences in the frequency of preconditions for unsafe acts events relative to the two hospital types (academic hospital (Study I) and non-academic hospital (Study II)). Although the overall distribution of events appears similar, individual differences emerged by hospital. For example, the academic hospital experienced an average of 31.4% of physical environment issues as compared with 20.5% for the non-academic hospital; $X^2 (1, N = 4515) = 47.357, p = 0.000$. In contrast, with respect to communication, coordination and planning issues, the academic hospital experienced less, an average of 23.3% while the non-academic hospital experienced more, an average of 31.4%; $X^2 (1, N = 4515) = 22.238, p = 0.000$.

Overall, the findings from both studies indicate that HFACS can be used to identify latent failures in retrospective observational data collected from a CVOR. Despite that the two data sets were collected without HFACS, under different methodologies and included very different sample sizes, results were very similar. In other words, when the HFACS framework was utilized to classify data in two different cardiac operating rooms, similar precondition categories were populated. However, differences were identified between the populated preconditions for unsafe acts categories at the academic hospital vs. the non-academic hospital.

CHAPTER 5: STUDY III (TRAUMA CENTER)

Introduction

Chapters three and four (Cohen et al., 2016) explored the utilization and reliability of HFACS for the classification of observational data in the cardiovascular operating room (CVOR). HFACS was applied to two different data-sets collected from CVORs and found that HFACS was not only reliable for both data sets (especially when using majority coding) but it could be used to detect subtle differences in types of latent failures identified within one domain. In other words, raters were successful in using HFACS to classify observations from both hospitals, and results were able to show differentiation between the two venues.

In an effort to better understand the types of threats facing providers in a different medical setting, researchers from Embry-Riddle Aeronautical University investigated flow disruptions at a trauma center located at an East Central Florida community hospital (see Boquet et al., 2016 for more detail).

The Trauma Domain: The trauma patient has a very different experience than the cardiac patient. In most cases trauma patients arrive to the hospital either via helicopter or ambulance (on rare occasions they arrive by personal vehicle). Upon arrival, the emergency medical technicians (EMTs) bring them to a resuscitation room or area, where the trauma team attempts to stabilize the patient. The trauma resuscitation team is made up of physicians, nurses and other allied health professionals. The American College of Surgeons, in their resources for optimal care of the injured patient (2014), explain that the general size and composition of the team may vary based upon the hospital size, severity of injury, and the corresponding level of trauma team activation. Traditionally, a high

level response to a severely injured patient includes a general surgeon, an emergency physician, surgical and emergency residents, emergency department nurses, a laboratory technician, a radiology technician, a critical care nurse, an anesthesiologist or nurse anesthetist, an operating room nurse, security officers, a chaplain or social worker and a scribe. However, a low level response to a less critically ill patient usually consists of only an emergency physician and the emergency department nurses until the general surgeon arrives (American college of Surgeons, 2014).

While most trauma cases occur in a series of steps as outlined above, there are also specific differences inherent to different hospital structures. As such, Figure 18 can be used as a diagram for the flow of a typical trauma patient at the hospital where data was collected for Study III. The dashed lines show the pathway for the patient to enter the trauma resuscitation area, and the solid line shows how the patient is transferred from this area to imaging. In most cases, patients arrive to the hospital either via helicopter (see number 1) or ambulance (see number 2). After they arrive the EMTs transfer the patient to the resuscitation area, which consists of three resuscitation bays (see number 3). Once here, the trauma resuscitation team treats the patient (as quickly as possible) to be sure that they are hemodynamically stable and ready for transfer. If needed, the patient is then transferred to the imaging suite. In this hospital, the patients are wheeled directly into a CT scan room (see number 4) which is out the doors of the trauma area, down a hallway and on the left side (see solid arrow). Here, some members of the trauma team go into the CT Scan Room and position the patient onto the scanner, and supply with contrast if needed, while other members of the team (namely the trauma surgeon, emergency department physician, nurses and scribe) go to the imaging observation area (see number

5) to look at the patient scans and wait for further assessment. This room is very small, and also serves as the observation area for two other CT Scan rooms. Once images are taken, a decision can be made as to whether the patient should be sent to surgery, sent to the ICU for recovery, or back to the trauma resuscitation room for further care (see Figure 18).

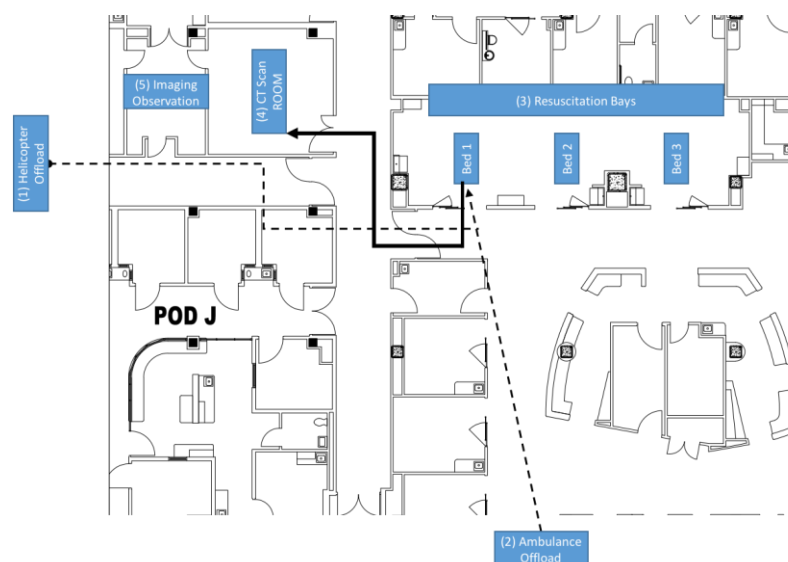


Figure 18. Level II trauma center (Study III)

Method

Data Collection: The data-set for this study includes workflow disruption events collected during trauma cases observed at a Level II trauma center (see Boquet et al., 2016 for a complete description). While Level I trauma centers (where a vast majority of research has taken place) have a dedicated team of specially trained staff on site 24 hours a day, 7 days a week, a level II trauma center requires the mobilization of a non-dedicated, multidisciplinary team of individuals who must respond quickly to the arrival of a trauma (American Trauma Society, 2016).

Sixty-five cases (yielding 1137 events) were collected from an East Central Florida community hospital over 23 months (4/29/14 - 3/12/16) and approved by the hospital's Institutional Review Board. During each case, researchers observed the team for workflow disruptions from the time the patient arrived in the resuscitation bay and continued through imaging (if needed), until disposition to surgery, the medical floor unit, or the emergency department for further assessment.

Specifically, case observation occurred in two parts: (1) observation of the team in the resuscitation bay (the area used to stabilize patients) and (2) observation of the team in imaging (CT-scan room where patients are brought following resuscitation, for in depth images to help providers gain a better understanding of the patient's medical status). Researchers observed and recorded workflow disruptions during both phases (if applicable), as well as the amount of time the patient and team spent in each observation area.

Coder Training: The same three raters who coded the CVOR data (from chapters three and four) will code the trauma data. Coder training will be the same as that discussed in studies one and two with the addition of more detailed instruction on specific trauma care topics including general emergency medicine terminology, trauma team member titles, roles and responsibilities, and common procedures and equipment/supplies in both the resuscitation bay and imaging. During this session, coders will be shown images of both the trauma resuscitation bays, and the imaging room.

Data Coding / Classification: All three raters classified a group of randomly selected events (n =50) into the HFACS framework with the author during a subsequent one-hour training session. After this period, each rater independently classified 100

additional events to practice coding on their own. The author and the three raters then discussed any disagreements (one hour). Following, the raters individually coded the remainder of the data (n = 987), and returned later to recode the original 150 data points.

Overall, the training period consisted of three hours (initial trauma-specific detailed lesson (1hr) followed by a group coding session (1hr) and finally a period of discussion for the first 100 items (1hr)).

Results

Reliability

The same methods from Study I and II were used to analyze the reliability of the data in Study III. With the unanimous method, of the 1137 original observations, all three coders agreed that 993 events could be coded using HFACS. Of the 993 “codable” events, all three coders agreed on the allocation of 929 events at the tier level of HFACS. Based on this, there was unanimous agreement that one event was related to organizational influences, 52 were unsafe acts, and the remaining 876 events were preconditions for unsafe acts. Of the 929 total agreed upon tier level events, 743 (74.8%) were unanimously agreed upon at the category level (see Figure 19).

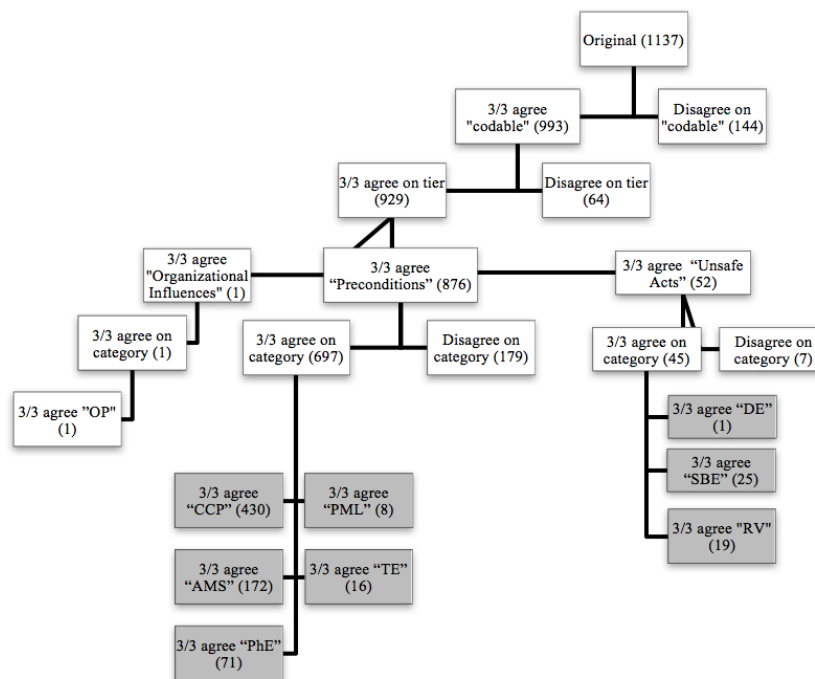


Figure 19. Data-set III line diagram (unanimous method)

In the majority method, 2/3 coders came to agreement that 1068 of the 1137 events could be coded using HFACS. Of the “codable” events, coders came to majority regarding their appropriate tier allocation for 1057 events. At this point, coders determined that the majority (976) could be classified as preconditions for unsafe acts, followed by unsafe acts (78) and organizational influences (2) (see Figure 20); Following the majority agreement method, coders and researchers met to reconcile those issues in which there was no agreement. Subsequently, 45 events that were originally disagreed upon were included in the reconciled data set. Of these, 33 were reconciled as preconditions for unsafe acts (see Figure 21). See Table 9 for a comparison of methods used and the inclusion data.

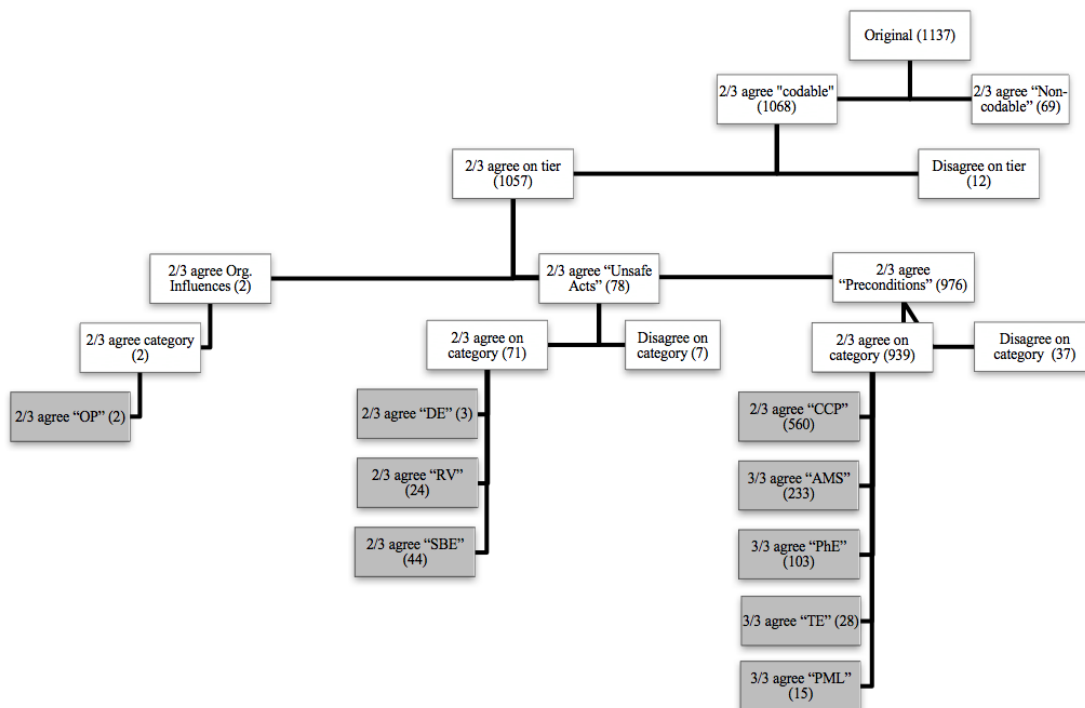


Figure 21. Data-set III line diagram (majority method)

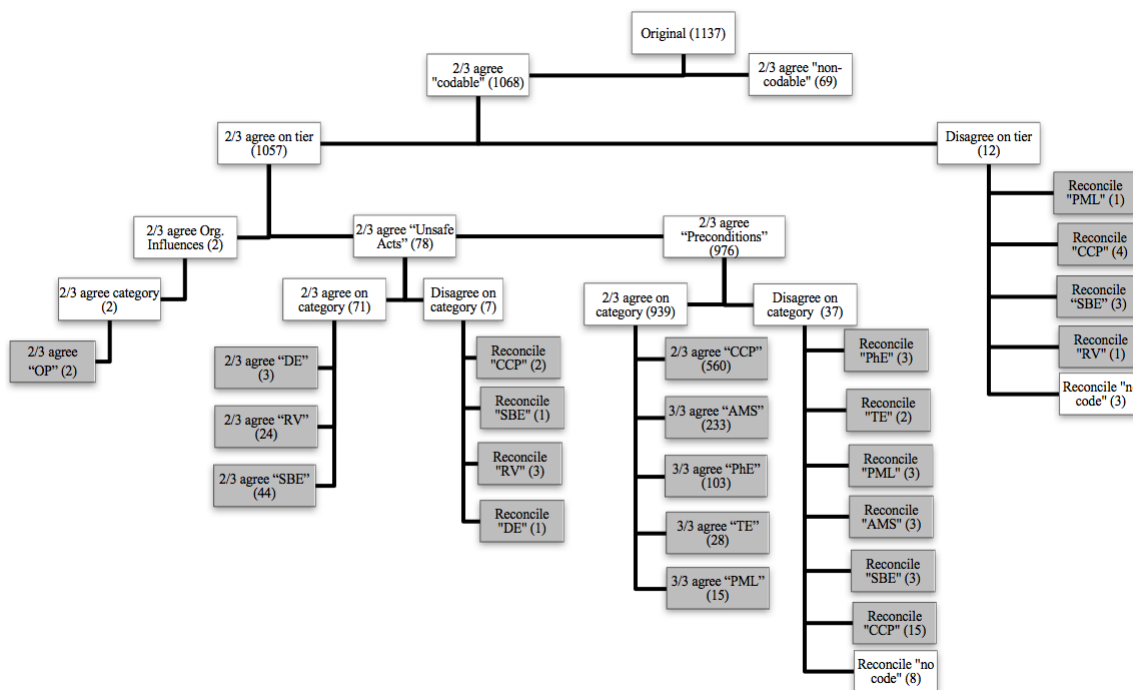


Figure 20. Data-set III line diagram (reconciled method)

Table 9. Data-set III methodology comparisons at the tier and category level

	Unanimous (3/3 agree)	Majority (2/3 agree)	Reconciled Majority (2/3 + consensus)	Pairwise Comparisons		
				1vs2	1vs3	2vs3
Tier	90%	99%	99% +	90%	89%	90%
Category	73%	94%	94% +	79%	77%	76%
Inclusion	742 total 697 preconditions	1010 total 939 preconditions	1054 total 972 preconditions			

Again, Fleiss' Kappa was calculated in two ways to represent a more stringent measure of inter-rater reliability for Study III. The overall Kappa was calculated to investigate inter-rater reliability for all potential events ($n = 1137$). Here, Fleiss' Kappa showed substantial reliability ($k=0.680$ (95% CI, .662 to .698), $p = 0.000$). Because an overwhelming majority of the data was again considered preconditions for unsafe acts, Kappa was also calculated based on those events that all raters unanimously agreed were "codable" at the preconditions for unsafe acts categories represented in HFACS. Here, Fleiss' Kappa showed an increase in reliability ($k = 0.757$ (95% CI, .731 - .784), $p = 0.000$).

Findings

Similar to data-set I and II a great majority of the data fell into the preconditions for unsafe acts category (93.8% and 92.7% for the unanimous and majority methods respectively). Because of similarity in results, data is discussed based on reconciled findings as an effort to include as much data as possible. Unlike that of the CVOR, in trauma most failures resulted from communication, coordination and planning issues (59.8%), followed by adverse mental states (24.3%) and issues in the physical environment (10.9%) (see Figure 22).

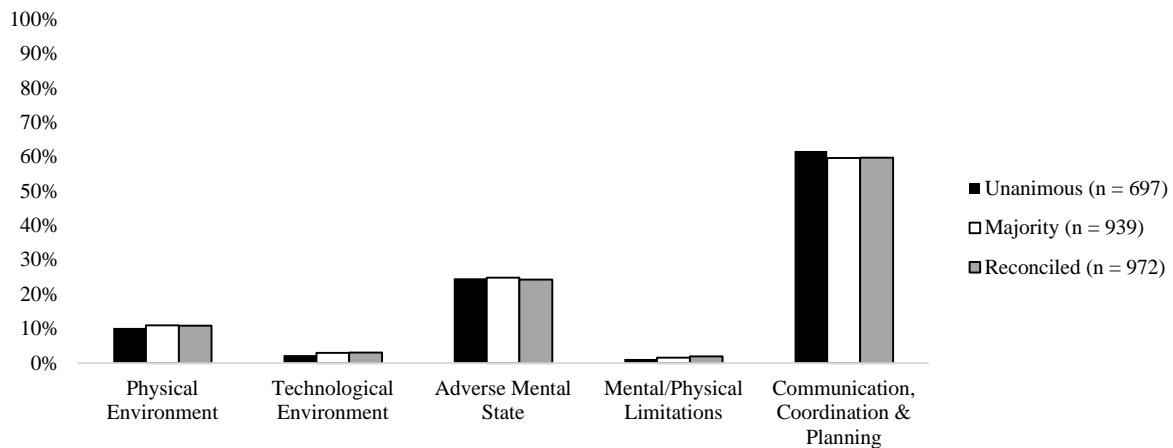


Figure 22. Data-set III unanimous vs. majority vs. reconciled preconditions for unsafe acts

CHAPTER 6: EXPLORATORY ANALYSES

While it is possible that the data collected in the CVOR studies (I and II) is inherently different from the data collected in the trauma center (study III) because of the differences in data collection, it can certainly be value added to explore the differences of each of these data sets with respect to the preconditions for unsafe acts findings. Although different methods were applied to collect the data in the trauma center than what was used in the CVOR, that does not necessarily limit this study from investigating differences in the populated HFACS preconditions between these hospitals.

Chi-square tests (X^2) were used to evaluate differences in the frequency of preconditions for unsafe acts events relative to the three hospitals utilized in the three studies (academic hospital CVOR, non-academic hospital CVOR, and trauma care facility). An assumption for Chi-square tests is that no more than 20% of the expected counts are less than five (Yates, Moore & McCabe, 1999, p. 734). As a result, two preconditions for unsafe act categories were not included in the analysis (i.e., adverse physiological state and fitness for duty).

Three of the precondition for unsafe acts categories were found to be different between the hospital types. These included physical environment; X^2 (2, N = 5487) = 117.532, $p = 0.000$), adverse mental state; X^2 (2, N = 5487) = 96.790, $p = 0.000$), and communication, coordination and planning; X^2 (2, N = 5487) = 333.926, $p = 0.000$).

Sharpe (2015) explains that one approach available to further investigate a statistically significant omnibus chi-square test result is to use a Bonferroni adjustment to control the family wise error rate. Therefore, post hoc analyses were conducted using

pairwise Chi-square tests with Bonferroni adjustments to investigate the statistically significant differences discussed above.

With respect to issues involving physical environment, differences existed between the academic CVOR and the non-academic CVOR; $X^2(1, N = 4515) = 47.357, p = 0.000$), the academic CVOR and the trauma center; $X^2(1, N = 1832) = 117.442, p = 0.000$), and the non-academic CVOR and the trauma center; $X^2(1, N = 4627) = 48.848, p = 0.000$). Specifically, physical environment issues were most prominent for the academic CVOR, making up 31.4% of the overall preconditions for unsafe acts followed by the non-academic CVOR at 20.5% and the trauma center at 10.9% (see Figure 23).

In terms of adverse mental state, differences existed between those studies conducted in the CVOR, and that which was conducted in the trauma care facility. Specifically, differences were highlighted between the academic CVOR and the trauma center; $X^2(1, N = 1832) = 117.442, p = 0.000$), and non-academic CVOR and the trauma center; $X^2(1, N = 4627) = 96.806, p = 0.000$). While the academic CVOR and non-academic CVOR did not differ from each other (38.3% and 41.5% respectively), both experienced significantly more adverse mental state issues than the trauma center (24.3%) (see Figure 23).

Finally, with respect to issues involving communication, coordination and planning, differences existed between the academic CVOR and non-academic CVOR; $X^2(1, N = 4515) = 22.238, p = 0.000$), the academic CVOR and the trauma center; $X^2(1, N = 1832) = 148.798, p = 0.000$), and non-academic CVOR and trauma center; $X^2(1, N = 4627) = 263.375, p = 0.000$). Communication, coordination and planning issues were most prominent in the trauma center, making up 59.8% of the overall preconditions for

unsafe acts followed by the non-academic CVOR at 31.4% and the academic CVOR at 23.3% (see Figure 23).

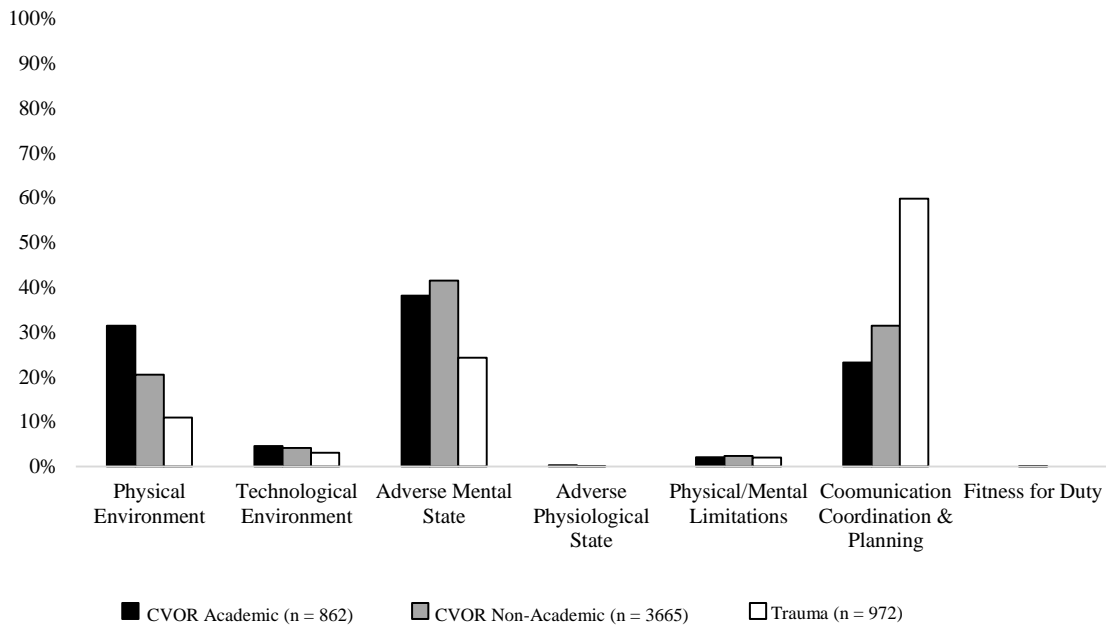


Figure 23. Comparison of preconditions for unsafe acts categories by hospital

CHAPTER 7: DISCUSSION

This dissertation investigated the utility of HFACS, a system originally designed for use in accident and incident investigation, for proactively classifying observational human factors data collected in various healthcare domains. Specifically, this project investigated three research questions:

Q1: Can HFACS, a tool developed originally for investigation of accident and incident data be reliably applied to observational data within healthcare? Here, observational data is used to describe human factors issues, that are observed in various healthcare domains.

Q2: Given the answer to Q1 is yes, Q2 will investigate if the utilization of HFACS in two different CVOR environments will result in the population of primarily adverse mental states, physical environment issues and communication, coordination and planning problems. More specifically, will there be a greater number of physical environment issues identified in a private academic hospital compared to a non-academic hospital setting?

Q3: Given that HFACS can be used on retrospective observational healthcare data collected in CVORs, can it be used in an entirely different domain (i.e., trauma care), where a different method was used for collecting the observations (i.e., observation of the entire team vs. observation of individuals).

Each research question was addressed and answered throughout the manuscript and will be described in more detail below.

Study I

Study I (chapter three) addressed RQ1, by investigating the reliability of HFACS as a tool for classifying retrospective observational healthcare data. In Study I, three human factors graduate students classified data from a CVOR at an academic university medical center using HFACS. Reliability was evaluated using percent agreement and Fleiss' Kappa. When using percent agreement as a method to investigate inter-rater reliability, values were calculated in terms of unanimous agreement (all three raters agree) and majority agreement (at least two raters agree). Based on unanimous agreement, reliability was low (67%). However, when majority agreement was considered, reliability was very high (94%). When reliability was calculated using a more stringent measure, Fleiss' Kappa, substantial reliability was reported (0.635).

Overall, HFACS was found to be reliable at the tier level and less reliable at the category level. This means that more often than not, the three raters agreed upon which level of the HFACS hierarchy each observation belonged. It is not uncommon for inter-rater reliability to decrease when raters are asked to code observations at the category level (Olsen, 2011; Ergai et al., 2015). It has been noted in the literature that as the number of possible categories for coding increases, the reliability generally decreases (Gwent et al., 2010).

It is important to note here that the data used in this study was not originally collected using the HFACS framework. The data set consisted of observations that were collected with the intent of analyzing flow-disruptions. While it is apparent that nearly all of the events could be classified using HFACS, this statement is not absolute. The data was not collected using HFACS, nor were the raters involved with its collection. While

the raters were trained in HFACS and in general terminology and procedures involved in the CVOR, they were not fully exposed to the environment in which the data was collected, therefore it is not surprising that there may have been some disagreement among raters.

Study II

In Study II (chapter four), the second research question was addressed by applying HFACS to observational healthcare data from another CVOR at a non-academic hospital. As expected, while the identified preconditions for unsafe acts were very similar between hospitals, there were differences in the distribution of these categories. For example, the academic hospital experienced more physical environment issues than the non-academic hospital. In contrast, the academic hospital experienced less communication, coordination and planning issues than the non-academic hospital. The findings of the second study highlight the ability of HFACS to differentiate between observational data collected in the CVOR in two very different hospitals.

While further research is needed to describe these findings perfectly, a number of factors may contribute to the differences identified. First, an increased number of individuals in the operating room may cause issues relating to the physical environment. Being that an academic hospital was utilized for the first data-set there were often several individuals observing, leading to overpopulated operating rooms. When more individuals were in an operating room, many issues involving the physical layout of the room were observed. Specifically, there were numerous instances where individuals did not have enough space to move past one another and several occasions where equipment and furniture had to be repositioned in a way not to negatively impact personnel.

With respect to the increase in communication, coordination and planning issues at the non-academic hospital, one factor to consider is the level of expertise of those in the room. At the non-academic hospital, rather than having anesthesiologists in each CVOR, a certified registered nurse anesthetist (CRNA) will often fulfil this position. According to the American Association of Nurse Anesthetists (AANA, 2016), A CRNA is an advanced practice registered nurse who has acquired graduate-level education and board certification in anesthesia. The American Society of Anesthesiologists (2016) explains that in the state of Florida (as well as 45 other states) physician supervision, collaboration, direction consultation, agreement, accountability, or discretion over nurse anesthetists providing anesthesia services is required. These rules and regulations may explain some of the disparity in coordination issues involving personnel being unavailable.

At the non-academic hospital, there were many events involving the CRNA waiting on or looking for assistance from the anesthesiologist who was oftentimes out of the room or assisting in another surgery. This may be related to the number of cardiac procedures occurring simultaneously on a given day. Essentially, the more cases occurring, the more difficult it may be to obtain the immediate assistance of an anesthesiologist. The academic hospital conducts an average of 600 cardiac procedures each year, while the non-academic hospital conducts an average of 2000 cardiac procedures each year (Blalock, 2016; Florida Hospital – Cardiovascular Institute, 2012). With the elimination of weekends and public holidays there are about 250 work days in a given year. This means that there was an average of 2.4 cardiac procedures occurring each day at the academic hospital, while there was an average of 8 procedures each day at

the non-academic hospital. While the increase in communication, coordination and planning issues observed at the non-academic hospital, was not originally hypothesized, the disparities in the number of procedures performed per day may partially explain these differences. Conceivably, a hospital with several ongoing procedures, with limited supervisory staff (i.e., anesthesiologist overseeing CRNAs), would likely have more issues involving coordination and planning (i.e., CRNAs waiting on anesthesiologist for guidance) than a hospital with one or two procedures per day.

The increase in procedures occurring at the non-academic hospital may also speak to an increased level of workload required for anesthesiologists who must shift their attention from case to case. Kleinman and Serfaty (1989) found that workload could be related to communication. Specifically, the authors imposed different workload levels on two person teams and found that teams exposed to high workload communicated significantly less than did teams with low or moderate workload. While this was not studied directly, it would be interesting to explore if the anesthesiologist and CRNA teams did communicate less often when more cases were occurring simultaneously. Perhaps this could be a focus of future research.

Another element that may impact the communication, coordination and planning category may involve differences in personality types of the individuals at the different hospitals. Attri and colleagues (2015) explain that personality traits, differences in beliefs and values, and personal factors can all affect the working environment in an operating room. Traits such as perfectionism, compulsiveness, and aggressiveness have been identified in both surgeons and anesthesiologists, which can make it more difficult to acknowledge one another's expertise, leading to challenges in terms of cohesive

teamwork. In the book “The Cultural Study of Work”, Wilson explains that every operating room is like no other operating room, and the personality of the surgeon can set the tone for the entire surgery. During interviews with nurses, Wilson reported that the operating room “is not a joking place if the surgeon does not make jokes, and not a talking place if the surgeon does not like to talk while operating” (p. 18). Further, he found that some surgeons are friendly and their rooms are filled with witty exchanges, while others are strict, allowing no talking whatsoever in their room (Harper & Lawson, 2003).

Currently, there is little to no literature exploring the differences between surgical personality types and styles and communication issues at academic vs. non-academic medical centers. However, several have noted the benefits of each hospital type which may be related to this concept. For example, David Shahian, a professor of surgery at Harvard Medical School explains that academic hospitals are at the forefront of medical research as they encourage surgical attendings to be up to date on literature so that they may better interact with and engage residents (Webster, 2014). While research is certainly needed to investigate this, it may be the case that surgeons at academic medical centers foster an environment with better communication because of the styles and personality types of individuals who are comfortable answering questions and working with residents and students. This is not to suggest that surgeons at non-academic universities promote operating rooms with poor communication; however, individuals in these settings may not need to take into consideration the learning style and personalities of students.

Study III

Finally, in an effort to investigate the utility of HFACS in a vastly different domain, where data was collected in a very different manner, Study III (chapter five) addressed RQ3. Here, HFACS was applied to observational healthcare data collected in a level II trauma care facility. Overall, in terms of unanimous agreement, inter-rater reliability was acceptable at 73%. However, this number increased substantially when using majority agreement (94%). Regardless of the observational data type, as well as the method in which the data was collected, HFACS appears to be sufficient for investigating failures in healthcare settings.

The findings from Study III were very much as expected, as the preconditions for unsafe acts made up a majority of the events (92.0% for the reconciled agreement). Further, a majority of the precondition events involved communication, coordination and planning issues. While these findings certainly appear different than those of the CVOR studies, they are not surprising given the nature of trauma care. Perhaps of initial interest is the notion that while a majority of the events in trauma were considered preconditions for unsafe acts, there were certainly more unsafe acts identified than in either of the CVOR studies. The unsafe acts category captures active failures of operators that may, ultimately, lead to an unintended outcome. The slight increase in these failures at the trauma center (7.9% compared to 0.69% and 1.4% in data-set I and II respectively) is not surprising considering the need to work at a fast pace in order to resuscitate a given patient. This finding speaks to the concept of the speed-accuracy trade off that occurs during response execution, the notion that when an individual must speed up their processes, their accuracy declines potentially leading to unforgiving errors (Wickens, 1998).

Another explanation for this increase in active failures involves the inherent differences in trauma care as opposed to cardiac care. Unlike cardiac cases which are scheduled and planned in advance, traumatic injury can happen at any point, resulting in the need for trauma care at all hours of the day and night. Barach and Weinger (2007) explain that sleep deprivation and fatigue are common among trauma team members who work regularly on recurring call or night shifts. The authors continue to note that “a sleep deprived or fatigued trauma team will make more errors, be less likely to recover from these errors, and provide lower quality care than a well-rested team” (p. 104).

With respect to the preconditions for unsafe acts findings, as expected, the area with the most issues involved communication, coordination and planning, making up nearly 2/3 of the preconditions for unsafe acts data. The treatment and care for the traumatically injured patient is oftentimes unpredictable and must occur in a fast-paced environment. In a level II trauma center, such as the one utilized in this particular study, the trauma team is not an in-tact team that always works together under the same conditions. Manser (2009) explains that this is not unfamiliar as most trauma teams must work under conditions that change frequently, may be assembled ad hoc, have a dynamically changing team membership and often work together for short periods of time.

Further, Roberts and colleagues (2014) explain that emergency medical teams have little time for deliberate planning and communication while providing care for the patient. However, effective performance is only ensured if the roles and tasks of the team members are clearly defined and communication and leadership aspects are regulated.

Because of the constraints placed on the trauma team because of the nature of trauma care, unfortunately proper communication and coordination is very difficult.

Overall Reliability

To reiterate the reliability findings, all three raters had substantial agreement across all three studies. For simplicity and ease of comparison, data will be discussed in terms of the Fleiss' Kappa findings. In these studies, values ranged from 0.635 to 0.680 when investigating overall Fleiss' Kappa and were slightly increased when reliability was explored based on only the preconditions for unsafe acts events (range from 0.660 to 0.757)

These findings are quite similar to others who have investigated the reliability of the original HFACS framework. For example, Ergai and colleagues (2016) evaluated inter-rater reliability for 125 participants and reported Krippendorff's alpha values of 0.67 overall. Similarly, Li and Harris (2006) utilized two raters to investigate 1762 events and reported a Cohen's Kappa value of 0.67 overall. These findings indicate that HFACS can be applied to retrospective observational healthcare data and yield similar reliability results as other studies investigating the reliability of HFACS.

Exploratory Discussion

While there were undoubtedly differences in the methods used to collect each data set, the comparison of the preconditions for unsafe acts findings from each data set is certainly of value. One of the underlying messages of this dissertation is the idea that hospital administrators can not apply a "one size fits all" generic approach when attempting to mitigate threats to patient safety.

The notion of comparing different types of data sets to gain a better understanding of the differences in domains is not dissimilar from applications of HFACS in aviation. For example, Wiegmann and Shappell (2001) analyzed human error data associated with aircrew-related commercial aviation accidents. The authors compared HFACS findings from both FAR Part 121 Schedule Carriers (major commercial airlines whose operations are governed by the Federal Aviation Regulations (FAR), Part 121) and FAR Part 135 (smaller commuter airlines or air services whose operations are governed by the Federal Aviation Regulations, Part 135). The authors found that the overall number of accidents associated with most error types was generally higher for FAR Part 135 scheduled carriers between the two groups were identified with respect to accident causal factor classification. Overall, the investigation provided valuable information about the differences between FAR Part 121 and 135 schedule carriers, highlighted the critical areas of human factors in need of further safety research and subsequently provided valuable insight aimed at the reduction of aviation accidents through data-driven strategies.

Similarly, Shappell and colleagues (2007) compared findings of HFACS analyses from general aviation accidents in Alaska versus the rest of the United States. Alaskan aviation is very different from general aviation in that Alaska is known for its varied and unique landscape, temperamental weather, and seasonal lighting conditions, making it one of the most difficult flying environments in the world. The authors identified a number of differences between general aviation accidents in Alaska and the rest of the US, concluding that efforts to generate intervention solutions must be needs-based and data-driven.

The aforementioned studies both utilized HFACS to investigate two different areas in the same domain as well as compare the findings between the two. Because general aviation in Alaska is conceivably different than general aviation in the rest of the United States, it is not surprising that there were also differences identified in the accident causal factors identified using HFACS. By comparing the two types of general aviation accidents, investigators and safety professionals are better able to understand the types of accidents occurring with respect to the specific area in which they are occurring. Subsequently, data-driven interventions can be put in place to reduce the risks to pilot safety, creating a more resilient aviation system overall.

While these studies were conducted in a very different domain than healthcare, the logic is still applicable here. By comparing the findings between latent failures identified across all three hospitals, healthcare administrators can gain a better understanding of the types of threats that face patient safety and be better equipped to design targeted interventions directed at reducing risk to specific areas in healthcare.

Overall, differences in preconditions for unsafe acts were identified when the three studies were compared. Perhaps most striking are the three areas with the highest frequency of events; physical environment, adverse mental state and communication, coordination and planning. There were significantly more physical environment issues at the academic hospital CVOR than the non-academic hospital CVOR as discussed earlier (refer to chapter 4) and at either CVOR compared to the trauma care facility. While the issues in physical environment were attributed to a larger number of personnel in Study I, this was not true in Study III. However, this may be related to the fact that the trauma bay is not heavily equipment laden like the CVOR. In other words, while there was still a

great deal of individuals in the room, because there was less equipment to interact with and navigate around, there were less issues identified with respect to the physical environment (see Figure 24 a/b).



(a)



(b)

Figure 24. (a) image of a typical cardiovascular operating room (b) image of a typical trauma resuscitation room

When looking at the differences in adverse mental state, it is clear that both CVORs, regardless of facility (academic vs. non-academic) experienced more of these issues than did trauma. This finding is not surprising given the life-threatening condition of trauma care patients and the pace at which they must be cared for. When a patient arrives at a trauma center, they are in critical condition and personnel must quickly assess the severity of the patient's injuries, while developing a plan for treating the patient. As such, there is little to no "down time" where team members can engage in non-essential communication, and they are rarely interrupted by any outside distractions such as non-essential personnel, text messages, pages or phone calls.

Finally, communication, coordination and planning issues differed between both CVOR types as previously explained (see Chapter 4), as well as between each CVOR and

the trauma facility. A few factors that could contribute to this finding are team stability (i.e., intact vs. ad hoc teams), time sensitivity involved in the case, and the ability to plan based on available information prior to a case. With respect to team stability, most CVOR teams are intact, in that they are made up of similar individuals who have worked together on a regular basis while the trauma teams (especially at level II trauma centers) are comprised of team members who may not have even met each other before. This could certainly cause issues for teams as they may not know who they are working with or how to communicate and coordinate with those individuals.

The time sensitive nature of trauma care may also lend itself to more communication and coordination issues, as individuals must act as quickly as possible to resuscitate the patient. While time is still of value in cardiovascular surgeries as increased time on cardiovascular bypass is of concern, teams in these settings have much more time to communicate with one another during a case. One factor that may influence how much communication and coordination is needed during a case, is the ability to plan ahead of time. In the case of cardiovascular surgeries (unless there is an emergency case) the surgeon knows exactly what procedure he/she will be doing, the full history of the patient, the risks associated with the type of procedure, and much more. In contrast, the trauma team is given little to no information regarding the case they are about to work on. This only becomes even more complex when multiple trauma cases arise at the same time.

General Discussion

Other important findings can be drawn from the categories that were *not* largely populated across the three data-sets (i.e., technological environment, adverse physiological state, physical/mental limitations, and fitness for duty). Perhaps most

interesting is the absence of many technological environment issues in either of the three studies. Several hospitals and healthcare systems allocate a great deal of their resources to new and improved technology. While practitioners have certainly reported usability problems associated with technology (new or old), this research suggests improving technology, at least in the CVOR and in trauma care, may not necessarily be the most important place to focus efforts when it comes to patient safety. Future research is needed to investigate if this holds true, not only in more CVORs and trauma care facilities but also in other healthcare domains.

Next, the lack of events for adverse physiological states, mental/physical limitations, and fitness for duty, speaks to the abilities of the team members at each of the facilities to perform their jobs as required. There were little to no events describing illness, physical fatigue, or other physiological states that would impair performance of team members. Similarly, observed practitioners were both mentally and physically capable of completing their required or necessary tasks. Finally, the individuals were always fit to perform their duties at work. There was never an indication that individuals did not adhere to rest requirements, alcohol restrictions or other off-duty mandates.

While the findings from these studies may not be true for all CVORs and level II trauma care facilities, they certainly provide insight into the types of issues that may be or *may not be* occurring within each domain, and how those issues may differ. By understanding the types of failures that occur in different settings, we can better provide hospital administrators with information to proactively address patient safety.

Limitations

While this dissertation was able to answer all three research questions proposed, each of the studies were certainly not without limitations. Perhaps the most recognizable being the differences between the samples of each of the studies. The sample size varied considerably between all three data-sets: (Data-set I: 15 cases yielding 878 events, Data-set II: 25 cases yielding 4233 events, Data-set III: 65 cases yielding 1137 events). Likewise, the method used for the data collection of each data-set was different. For data-set I, three positions were observed overall (anesthesiologist, circulating nurse, perfusionist) and only two positions were observed for disruptions during each case. In the second data-set, four positions were observed overall (anesthesiologist, circulating nurse, perfusionist and surgeon) and data from all four positions was collected during each case (each researcher observed two positions at a time). Finally, in data-set III, the trauma team was observed (rather than individual positions or specialty areas).

Another possible limitation involves the lower levels of reliability reported for unanimous agreement and Fleiss' Kappa. While reliability was generally high for the classifications at the tier level, agreement levels were typically lower at the specific causal factor level. However, this finding is consistent with other studies that have found that as the number of subcategories or coders increases, acceptable levels of reliability are generally more difficult to obtain.

One factor that may have influenced reliability is that the coders of the observational data were not involved in the data collection. While the coders were trained in HFACS and general terminology/procedures involved in the CVOR and trauma care, they were not fully exposed to the environment in which this data was collected.

Another issue involves the use of observational data in general. Because these studies were conducted using retrospective observational healthcare data, only the failures that were observable to researchers during the time of data collection were recorded and thus classified into HFACS. For example, failures at the organizational and supervisory tiers of HFACS (i.e., organizational influences and supervisory factors) were rarely observed, as these failures would typically occur outside of the observation area. Similarly, while a few unsafe acts were observed, it is likely that many more occurred and were missed by observers. Although issues like slips, trips and falls (skill-based errors) are easily observable, because the observers were not clinicians themselves, it was difficult for them to pick up on any error or violation made by practitioners unless directly told beforehand (i.e., improper scrub techniques used; a break in sterile protocol; improper vocabulary used for tools and equipment; inappropriate procedures, etc.).

While studying and exploring the observable failures certainly provides a good start on the path for proactive patient safety, this method alone may not be adequate. If hospital administrators are to truly get a full picture of the latent failures occurring within the hospital or care setting, other methods outside of observation may be necessary. For example, surveys could be used to investigate issues at the organizational and supervisory tiers. Along similar lines, perhaps a practitioner or an observer with innate knowledge of procedures is needed to observe cases to truly capture the unsafe acts that occur.

Contributions and Future Directions

Historically, the medical community has focused most of its efforts on identifying and mitigating human error via the analysis of sentinel events. While this approach has been successful, the healthcare industry continues to face issues involving morbidity and mortality. As Woods and colleagues (2010) explain, focusing on a single root cause or sentinel event in the case of medicine, “retards” our ability to understand the role of multiple contributors or latent failures. A vital component to the proactive approach to patient safety involves understanding that these multiple contributors can create conditions that reduce the resilience of a system. If we are able to capture and study these contributors, we can develop interventions that may be able to mitigate them before they reach the patient.

The proactive identification of human factors issues associated with patient harm therefore represents the next step in the evolution of patient safety. This study is a first step in establishing the reliability of the HFACS framework as a tool for classifying observational human factors data in two different medical venues. As HFACS appears to be a reliable observation tool, findings associated with its use could help to identify where errors and adverse events are likely to occur.

The studies in this dissertation further demonstrate that when applied to retrospective healthcare data, HFACS can be used to identify different areas of threat depending on the particular system. Predictably, hospital administrators could put in place targeted interventions to help mitigate human factors issues before they manifest and become errors in the future.

One way to develop targeted interventions as described above, would be to utilize the Human Factors Intervention Matrix (HFIX), a system developed based on human factors engineering principles that allows for the implementation of targeted, data-driven interventions with the ultimate goal of reducing human error (Shappell & Wiegmann, 2006). The matrix is designed so that identified threats are positioned against five intervention approaches that capture the underlying casual mechanisms of human error. The five dimensions include: (1) human/crew, (2) technology/engineering, (3) technical/physical, (4) task/procedure and (5) organizational/supervisory. To employ the HFIX methodology, the organization must utilize subject matter experts (SMEs) to brainstorm intervention strategies that are aimed at addressing the specific threats.

While several intervention prototypes may be generated while using HFIX, it is unlikely that all interventions can be implemented by the organization. Shappell and Wiegmann (2006) developed a method by which to evaluate each particular intervention that focuses on certain factors that should be considered before it is employed. These factors include: Feasibility, Acceptability, Cost, Effectiveness, and Sustainability (FACES). Each of these factors is rated on a 5-point Likert scale (1= “worst”, 5 = “best”) which can then be used to determine which interventions should be selected for implementation. The final product represents a cubed structure that can be used to visualize the threats identified against the intervention approaches and the subsequent evaluation criteria (see Figure 25).

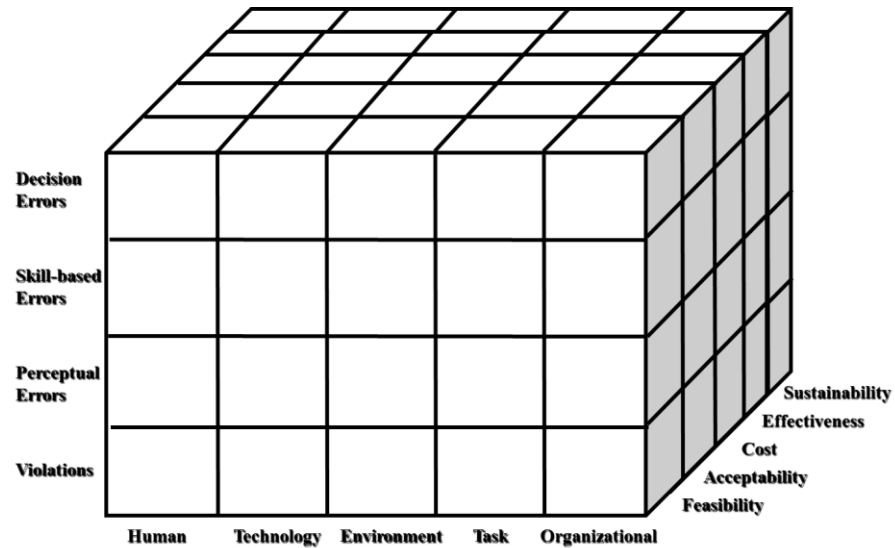


Figure 25. The HFIX Cube (Adapted from Shappell & Wiegmann, 2006)

While HFIX has certainly been employed successfully in several industries, it is important to note that it is not the only intervention implementation technique available. This is simply one example of how a hospital system could proactively use observational process inefficiency data to create targeted interventions that would be designed specifically for their needs.

This concept of using proactive methods to reduce harm can be compared to physicians who inoculate patients. Inoculations are used as preventative measures that help to reduce the risk of death or suffering. While inoculations are widely used, physicians do not simply administer “omnibus” vaccinations that prevent the likelihood of any and all disease. Rather, they give a specific vaccination to prevent specific illness and disease. Similarly, this research allows us to investigate and identify the types of human factors issues in medicine that may lead to catastrophe in the future. Rather than waiting for catastrophe to happen and fixing the problem after the fact (a reactive approach), we may be able to use HFACS in a proactive way, much like preventative

medicine, to design targeted interventions that reduce adverse events and patient harm in the future.

While this dissertation discusses the implications of HFACS as a proactive tool, that theory must be expanded into practice. In other words, further efforts should be placed on conducting observational studies using HFACS rather than applying HFACS to retrospective observational data sets. In order to do this effectively, efforts are also needed to design and evaluate a prospective HFACS-specific observational tool to help identify human factors failures, which would allow for targeted interventions to be deployed before harm occurs.

BIBLIOGRAPHY

- Abernathy III, J. H. (2015). Patient Safety in the Cardiac Operating Room: What Can, Will, and Might Make Patients Safer and You Happier? *ASA Refresher Courses in Anesthesiology*, 43(1), 1-6.
- American Association of Nurse Anesthetists (2016). *Certified Registered Nurse Anesthetists Fact Sheet*. Retrieved August 20, 2016, from <http://www.aana.com/ceandeducation/becomeacrna/Pages/Nurse-Anesthetists-at-a-Glance.aspx>
- American College of Surgeons: Committee on Trauma (2014). *Resources for the Optimal Care of the Injured Patient* (6th ed.). Chicago, IL.
- American Society of Anesthesiologists (2016). *CMS Supervision Rules for Nurse Anesthetists*. Retrieved August 20, 2016, from <http://www.asahq.org/advocacy/federal-activities/items-of-interest/cms-supervision-rules-for-nurse-anesthetists>
- American Trauma Society (2016). *Trauma Center Levels Explained*. Retrieved March 06, 2016, from <http://www.amtrauma.org/?page=TraumaLevels>
- Attri, J. P., Sandhu, G. K., Mohan, B., Bala, N., Sandhu, K. S., & Bansal, L. (2015). Conflicts in operating room: Focus on causes and resolution. *Saudi journal of anaesthesia*, 9(4), 457.
- Barach, P., & Weinger, M. B. (2007). Chapter 6: Trauma Team Performance. In *Trauma: Resuscitation, Perioperative Management, and Critical Care* (Vol. 1, pp. 101-111). Boca Raton, FL: Taylor & Francis Group.

- Barrett, S. (2001). The impact of training on rater variability. *International Education Journal*, 2(1), 49-58.
- Baysari M.T., Caponecchia, C., McIntosh, A.S., & Wilson, J.R. (2009). Classification of errors contributing to rail incidents and accidents: A comparison of two human error identification techniques. *Safety Science*, 47, 848-857.
- Baysari, M.T., Caponecchia C., & McIntosh A.S. (2001). A reliability and usability study of TRACER-RAV: The technique for the retrospective analysis of cognitive errors – For rail, Australian Version. *Applied Ergonomics*, 42(6), 852-859.
- Baysari, M.T., McIntosh, A.S., & Wilson, J.R. (2008). Understanding the human factors contribution to railway accidents and incidents in Australia. *Accident, Analysis and Prevention* 40,1750-1757.
- Bergs, E. A., Rutten, F. L., Tadros, T., Krijnen, P., & Schipper, I. B. (2005). Communication during trauma resuscitation: do we know what is happening? *Injury*, 36(8), 905-911.
- Birkimer, J. C., & Brown, J. H. (1979). Back to basics: Percentage agreement measures are adequate, but there are easier ways. *Journal of Applied Behavior Analysis*, 12(4), 535-543.
- Blalock, J. (2016). Division of Cardiothoracic Surgery Adult Cardiac Section. Retrieved September 28, 2016, from <http://academicdepartments.musc.edu/surgery/divisions/cardiothoracic/cardiac.htm>
- Blocker, R. C., Duff, S., Wiegmann, D., Catchpole, K., Blaha, J., Shouhed, D., ... & Gewertz, B. (2012, September). Flow disruptions in trauma surgery: Type, impact,

- and affect. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 56, No. 1, pp. 811-815). SAGE Publications.
- Blocker, R. C., Shouhed, D., Gangi, A., Ley, E., Blaha, J., Gewertz, B. L., ... & Catchpole, K. R. (2013). Barriers to trauma patient care associated with CT scanning. *Journal of the American College of Surgeons*, 217(1), 135-141.
- Bracco, D., Favre, J. B., Bissonnette, B., Wasserfallen, J. B., Revelly, J. P., Ravussin, P., & Chioléro, R. (2001). Human errors in a multidisciplinary intensive care unit: a 1-year prospective study. *Intensive care medicine*, 27(1), 137-145.
- Button, K.S., Ioannidis, J.P.A., Mokrysz, C., Nosek, B.A., Flint, J., Robinson, E.S.J., & Munafò, M.R. (2013). Power failure: why small sample size undermines the reliability of neuroscience. *Nature Reviews*, 14(5), 365-376.
- Carayon, P., Alvarado, C. J., Hundt, A.S., Springman, S., Borgsdorf, A., & Hoonakker, P. (2005). An employee questionnaire for assessing patient safety in outpatient surgery. In Henrikson, Battles, Marks, & Lewin (Eds.). *Advances in Patient Safety: From Research to Implementation*, AHRQ Publication. [PMID: 21250041](#)
- Carayon, P., Hundt, A.S., Hoonakker, P.L.T., Kianfar, S., Alyousef, B., Salek, D., Cartmill, R., Walker, J.W., & Tomcavage, J. (2015). Perceived impact of care managers' work on patient and clinician outcomes. *European Journal for Person Centered Healthcare*, 3(2), 158-167.
- Carayon, P., Hundt, A.S., Karsh, B.T., Gurses, A.P., Alvarado, C.J., Smith, M., & Brennan, P.F. (2006). Work system design for patient safety: The SEIPS model. *Quality and Safety in Health Care*, i50-i58. PMID: PMC2464868

- Carayon, P., Li, L., Kelly, M.M., DuBenske, L.L., Xie, A., McCabe, B., Orne, J., & Cox, E.D. (2014). Stimulated recall methodology for assessing work system barriers and facilitators in family-centered rounds in a pediatric hospital. *Applied Ergonomics*, 45 (6), 1540-1546. [PMCID: PMC4119819](#)
- Carmines, E.G., & Zeller, R.A. (1989). Reliability and Validity Assessment. Sage University Paper series on Quantitative Applications in the Social Sciences, series no. 07-017. Beverly Hills and London: Sage Publications.
- Carmines, E.G., & Zeller, R.A. (1979). *Reliability and Validity Assessment*. Thousands Oaks, CA: Sage Publications.
- Carthey J, de Leval, M.R., Reason, J.T. (2001) The human factor in cardiac surgery: errors and near misses in a high technology medical domain. *Annals of Thoracic Surgery*, 72, 300–305.
- Catchpole, K.R., Giddings, A.E., Wilkinson, M., Hirst, G., Dale T., & de Leval, M.R. (2007). Improving patient safety by identifying latent failures in successful operation. *Surgery*, 142, 102-110.
- Catchpole, K.R., Gangi, A., Blocker, R.C., Ley, E.J., Blaha, J., Gewertz, B.L., & Wiegmann, D. A. (2013). Flow disruptions in trauma care handoffs. *journal of surgical research*, 184(1), 586-591.
- Catchpole, K., Mishra, A., Handa, A., & McCulloch, P. (2008). Teamwork and error in the operating room: analysis of skills and roles. *Annals of surgery*, 247(4), 699-706.

- Catchpole, K., Ley, E., Wiegmann, D., Blaha, J., Shouhed, D., Gangi, A., ... & Starnes, B. (2014). A human factors subsystems approach to trauma care. *JAMA surgery, 149*(9), 962-968.
- Centers for Disease Control and Prevention. (2013). Morality Multiple Cause Micro-data files. Retrieved from http://www.cdc.gov/nchs/data/nvsr/nvsr64/nvsr64_02.pdf
- Chen, P. W. (2010, October 13). The Doctor is in (but Shouldn't Be). Retrieved May 15, 2016, from <http://www.nytimes.com/2010/10/14/health/views/14chen.html>
- Chen, S., Wall, A., Davies, P., Yang, Z., Wang, J., & Chou, Y. (2013). A human and organizational factors (HOFs) analysis method for marine causalities using HFACS-Maritime Accidents (HFACS-MA). *Safety Science, 60*, 105-114.
- Chisholm, Carey D., et al. "Emergency Department Workplace Interruptions Are Emergency Physicians "Interrupt-driven" and "Multitasking"?" *Academic Emergency Medicine* 7.11 (2000): 1239-1243.
- Cohen, T., Wiegmann, D., & Shappell, S. (2015). Evaluating the reliability of the Human Factors Analysis and Classification System. *Aerospace Medicine and Human Performance, 86* (8), 728-735.
- Coiera, E., & Tombs, V. (1998). Communication behaviours in a hospital setting: an observational study. *Bmj, 316*(7132), 673-676.
- Collins, R. L. (2011). Heinrich and beyond. *Process Safety Progress, 30*(1), 2-5. doi:10.1002/prs.10422
- Congdon P.J., & McQueen, J. (2000). The stability of rater severity in large-scale assessment programs. *Journal of Educational Measurement, 37*(2), 163-178.

- Courdier, S., Garbin, O., Hummel, M., Thoma, V., Ball, E., Favre, R., & Wattiez, A. (2009). Equipment failure: causes and consequences in endoscopic gynecologic surgery. *Journal of minimally invasive gynecology*, 16(1), 28-33.
- de Leval, M.R., Carthey, J., Wright, D.J., and Reason, J.T. (2000). Human factors and cardiac surgery: a multicenter study. *Journal of Thoracic and Cardiovascular Surgery*, 119, 661-672.
- Donchin, Y., Gopher, D., Olin, M., Badihi, Y., Biesky, M., Sprung, C. L., ... & Cotev, S. (2003). A look into the nature and causes of human errors in the intensive care unit. *Quality and Safety in Health Care*, 12(2), 143-147.
- Drew, T., Võ, M. L. H., & Wolfe, J. M. (2013). The invisible gorilla strikes again: sustained inattention blindness in expert observers. *Psychological science*, 24(9), 1848-1853.
- Edwards, E., (1972). Man and machine: systems for safety. In: Proceedings of British Airline Pilots Associations Technical Symposium. British Airline Pilots Associations, London, pp. 21–36.
- ElBardissi, A.W., Wiegmann, D.A., Dearani, J.A., Daley, R.C., & Sundt, T.M. (2007). Application of the human factors analysis and classification system methodology to the cardiovascular surgery operating room. *The Society of Thoracic Surgeons*, 83, 1412-1419.
- Ergai, A.O., Cohen, T.N., Wiegmann, D.A., Sharp, J.L., Gramopahdy, A., & Shappell, S.A. (2016). Assessment of the Human Factors Analysis and Classification System (HFACS): Intra-rater and Inter-rater Reliability. *Safety Science*, 82, 393-398.

- Eric Cropp weighs in on the error that sent him to prison. (2009, December 3). Retrieved May 15, 2016, from <http://ismp.org/Newsletters/acutecare/articles/20091203.asp>
- Family wins \$2.5M in drunk doctor case. (2010, March 29). Retrieved May 15, 2016, from http://www.bizjournals.com/cincinnati/morning_call/2010/03/family_wins_25m_in_drunk_doctor_case.html
- Florida Hospital- Cardiovascular Institute (2012). *2012 annual outcome report*. Retrieved from https://www.flheart.org/sites/www.flheart.org/files/2012%20FHCVI%20Annual%20Report_1.pdf
- Gelman, L. (2012). Doctors Confess Their Fatal Mistakes | Page 6 of 7 | Reader's Digest. Retrieved May 15, 2016, from <http://www.rd.com/health/conditions/doctors-confess-their-fatal-mistakes/6/>
- Gora, G. (2016, January 30). 10 Nightmarish Stories About Terrifying Medical Errors - Listverse. Retrieved May 15, 2016, from <http://listverse.com/2016/01/30/10-nightmarish-stories-about-terrifying-medical-errors/>
- Gruen, R. L., Jurkovich, G. J., McIntyre, L. K., Foy, H. M., & Maier, R. V. (2006). Patterns of errors contributing to trauma mortality: lessons learned from 2594 deaths. *Annals of surgery, 244*(3), 371-380.
- Guar D. (2005). Human factors analysis and classification system applied to civil aircraft accidents in India. *Aviation, Space, and Environmental Medicine, 76*(5), 501-505.
- Gurses, A. P., Carayon, P., & Wall, M. (2009). Impact of performance obstacles on intensive care nurses' workload, perceived quality and safety of care, and quality of working life. *Health services research, 44*(2p1), 422-443

- Gurses, A.P., Kim, G., Martinez, E.A. Martinez, Marsteller, J., Bauer, L., Lubomski, L.H., Pronovost, P.J., & Thompson, D., (2012) Identifying and categorising patient safety hazards in cardiovascular operating rooms using an interdisciplinary approach: a multisite study. *British Medical Journal of Quality and Safety*, 21, 810–818
- Gwent, Kilem., 2010. *Handbook of Inter-rater Reliability*, (2) Advanced Analytics, LLC
- Harper, D. A., & Lawson, H. M. (2003). *The cultural study of work*. Lanham, MD: Rowman & Littlefield.
- Harkins, D. (2009). Trauma is a team sport. *Journal of Trauma Nursing*, 16(2), 61-63.
- Hartmann, D. P. (1977). Considerations in the choice of interobserver reliability estimates. *Journal of Applied Behavior Analysis*, 10, 103-116.
- Hayes, A.F., & Krippendorff, K. (2007). Answering the call for a standard reliability measure for coding data. *Communication Methods and Measures*, 1 (1), 77-89.
- Hawkins, F.H., & Orlady, H.W. (Ed.). (1993). *Human factors in flight* (2nd ed.). England: Avebury Technical, 1993.
- Heinrich, H. W. (1959). *Industrial accident prevention: A scientific approach* (4th ed.). New York: McGraw-Hill.
- Heron, M.P., Hoyert, D.L., Murphy, & S.L., et al. (2009) Deaths: Final data for 2006. *Natl Vital Stat Rep*. 57(14). National Center for Health Statistics. Hyattsville, MD
- Holden, R. J., Carayon, P., Gurses, A. P., Hoonakker, P., Hundt, A. S., Ozok, A. A., & Rivera-Rodriguez, A. J. (2013). SEIPS 2.0: A human factors framework for studying and improving the work of healthcare professionals and patients. *Ergonomics*, 56(11), 1669. doi:10.1080/00140139.2013.838643

- Hollnagel, E. (2010;2009;). *Safer complex industrial environments: A human factors approach* (1st ed.). Boca Raton, Fla: CRC Press.
- Hoonakker, P., Cartmill, R.S., Carayon, P., & Walker, J. (2010). Development of the survey to evaluate CPOE/EHR implementation in ICUs. *Technical Report*, Center for Quality and Productivity Improvement (CQPI), Madison, WI.
- House, A. E., House, B. J., & Campbell, M. B. (1981). Measures of interobserver agreement: Calculation formulas and distribution effects. *Journal of Behavioral Assessment*, 3, 37- 57.
- Hughes, A.M., Sonesh, S., Zajac, S., & Salas, E. (2015) Leveraging HFACS to understand medication error in emergency medical services (EMS): A systematic Review. *Proceedings from HFES 57th Annual Meeting*, 1688-1692.
- IOM (Institute of Medicine) (2000). *To Err is Human—Building a Safer Health System*. Washington, DC: The National Academies Press
- James J. (2013) A new, evidence-based estimate of patient harms associated with hospital care. *Journal of Patient Safety*, 9, 122-128.
- Karsh, B.T., Holden, R.J., Alper, S.J., & Or C.K. (2006). A human factors engineering paradigm for patient safety: designing to support the performance of the healthcare professional. *Quality and Safety in Healthcare*, 15(Suppl 1), i59-i65.
- Kita, J. (2012). Doctors Confess Their Fatal Mistakes | Page 4 of 7 | Reader's Digest. Retrieved May 23, 2016, from <http://www.rd.com/health/conditions/doctors-confess-their-fatal-mistakes/4/>
- Kleinman, D. L. & Serfaty, D. (1989, April). Team performance assessment in distributed decision making. Paper presented at the Simulation and Training Research

Symposium on Interactive Networked Simulation for Training, University of Central Florida, Orlando, FL

Kopp, C. (2013, March 16). Anatomy of A Mistake. Retrieved May 15, 2016, from <http://www.cbsnews.com/news/anatomy-of-a-mistake-16-03-2003/>

Krippendorff, K., (2004). *Content analysis: An introduction to its methodology*. Sage Publications, Incorporated.

Landis, J.R., & Koch, G.G. (1977). The measurement of observer agreement for categorical data, *Biometrics*, 33, 159-174.

Larouzee, J., Guarnieri, F., & Besnard, D. (2014). Le modèle de l'erreur humaine de James Reason. *Papiers de Recherche du CRC, MINES ParisTech*. 44p.

Lenné, M.G., Ashby, K., & Fitzharris, M. (2008) Analysis of general aviation crashes in Australia using the human factors analysis and classification system. *The International Journal of Aviation Psychology*, 18(4), 340-352.

Li, W., Harris, D., & Yu, C. (2008). Routes to failure: Analysis of 41 civil aviation accidents from the Republic of China using the human factors analysis and classification system. *Accident, Analysis and Prevention*, 40, 426-434.

Li, W., & Harris, D. (2006). Pilot error and its relationship with higher organizational levels: HFACS analysis of 523 accidents. *Aviation, Space, and, Environmental Medicine*, 77(10), 1056-1061.

Mackenzie, C. F., & Xiao, Y. (2003). Video techniques and data compared with observation in emergency trauma care. *Quality and Safety in Health Care*, 12(suppl 2), ii51-ii57.

- Manser, T. (2009). Teamwork and patient safety in dynamic domains of healthcare: a review of the literature. *Acta Anaesthesiologica Scandinavica*, 53(2), 143-151.
- Memtsoudis, S. G., Vougioukas, V. I., Stundner, O., & Poultsides, L. A. (2015). Perioperative outcomes in orthopedic surgery. *The Scientific World Journal*, doi:<http://dx.doi.org.ezproxy.libproxy.db.erau.edu/10.1155/2015/648284>
- O'Connor, P., & Walker, P. (2011) Evaluation of a human factors analysis and classification system as used by simulated mishap boards. *Aviation, Space, and Environmental Medicine*, 82(1), 44-48.
- O'Connor, P., Walliser, & Philips, E., (2010). Evaluation of a human factors analysis and classification system used by trained raters. *Aviation, Space, and Environmental Medicine*, 81 (10), 957-960.
- O'Connor, P. (2008). HFACS with an additional layer of granularity: Validity and utility in accident analysis. *Aviation, Space, and Environmental Medicine*, 79(6), 599-606.
- Ofri, D. (2013). My Near Miss. Retrieved May 23, 2016, from http://www.nytimes.com/2013/05/29/opinion/addressing-medical-errors.html?_r=0
- Olsen, N.S. (2011). Coding ATC incident data using HFACS: Inter-coder consensus. *Safety Science*, 49,1365-1370.
- Olson N.S., Shorrock S.T., 2010. Evaluation of the HFACS-ADF safety classification system: Inter-coder consensus and intra-coder consistency. *Accident Analysis and Prevention*, 42, 437-444.

- Orlando Health. (2016). *Orlando Regional Medical Center Services*. Retrieved March 11, 2016 by: <http://www.orlandohealth.com/facilities/orlando-regional-medical-center/services>.
- Overconfidence and nursing malpractice. (2012, November 12). Retrieved May 15, 2016, from <http://www.medleague.com/overconfidence-and-nursing-malpractice/>
- Palmer, G., Abernathy, J., Swinton, G., Allison, D., Greenstein, J., Shappell, S., Juang, K., Reeves, S.T. (2013). Realizing Improved Patient Care through Human-centered Operating Room Design. *Anesthesiology*, *119* (5),1066-1077.
- Parker, S. E. H., Laviana, A. A., Wadhera, R. K., Wiegmann, D. A., & Sundt III, T. M. (2010). Development and evaluation of an observational tool for assessing surgical flow disruptions and their impact on surgical performance. *World journal of surgery*, *34*(2), 353-361.
- Patterson, J.M., & Shappell, S.A., (2010). Operator error and system efficiencies: Analysis of 508 mining incidents and accidents from Queensland, Australia using HFACS. *Accident, Analysis & Prevention*, *42*, 1379-1385.
- Pennathur, P. R., Thompson, D., Abernathy III, J. H., Martinez, E. A., Pronovost, P. J., Kim, G. R., ... & Gurses, A. P. (2013). Technologies in the wild (TiW): human factors implications for patient safety in the cardiovascular operating room. *Ergonomics*, *56*(2), 205-219.
- Rashid, H.S.J., Place, C.S., & Braithwaite, G.R. (2010) Helicopter maintenance error analysis: Beyond the third order of the HFACS-ME. *International Journal of Industrial Ergonomics*, *40*, 636-647.
- Reason, J., (1990). *Human error*. New York: Cambridge University Press.

- Reever, M. M., & Lyon, D. S. (2002). Emergency room communication issues: Dealing with crisis. *Advanced Emergency Nursing Journal*, 24(4), 62-66.
- Reinach, S., & Viale, A., (2006). Application of a human error framework to conduct train accident/incident investigations. *Accident Analysis & Prevention*. 38, 396-406.
- Roberts, N. K., Williams, R. G., Schwind, C. J., Sutyak, J. A., McDowell, C., Griffen, D., ... & Wohltmann, C. (2014). The impact of brief team communication, leadership and team behavior training on ad hoc team performance in trauma care settings. *The American Journal of Surgery*, 207(2), 170-178.
- Rousson, V., Gasser, T., & Seifert, B. (2002). Assessing intrarater, interrater and test-retest reliability of continuous measurements. *Statistics in Medicine*, 21, 3431-3446.
- Sarcevic, A. (2010, April). Who's scribing? documenting patient encounter during trauma resuscitation. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1899-1908). ACM.
- Sarcevic, A., Marsic, I., Waterhouse, L. J., Stockwell, D. C., & Burd, R. S. (2011). Leadership structures in emergency care settings: A study of two trauma centers. *International Journal of Medical Informatics*, 80(4), 227-238. doi:10.1016/j.ijmedinf.2011.01.004
- Schraagen, J. M., Schouten, T., Smit, M., Haas, F., van der Beek, D., van de Ven, J., & Barach, P. (2010). Assessing and improving teamwork in cardiac surgery. *Quality and Safety in Health Care*, 19(6), e29-e29

- Shappell, S., Detwiler, C., Holcomb, K., Hackworth, C., Boquet, A., & Wiegmann, D. A. (2007). Human error and commercial aviation accidents: an analysis using the human factors analysis and classification system. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 49(2), 227-242.
- Shappell, S., & Wiegmann, D. Applying Reason: The human factors analysis and classification system (HFACS). *Human Factors and Aerospace Safety 2001*; 1:59-86.
- Shappell, S. & Wiegmann, D. (2006). Developing a Methodology for Assessing Safety Programs Targeting Human Error in Aviation. Report number: DOT/FAA/AM-06/24. Washington, D.C.: Office of Aerospace Medicine/Federal Aviation Administration.
- Sharpe, D. (2015). Your Chi-square test is statistically significant. Now what? *Practical Assessment, Research & Evaluation*, 20(8), 1-10.
- Shohamy, E., Gordon, C., & Kraemer, R., 1992. The Effect of Rater's Background and Training on the Reliability of Direct Writing Tests. *The Modern Language Journal*, 76(1), 27-33.
- Shouhed, D., Blocker, R., Gangi, A., Ley, E., Blaha, J., Margulies, D., ... & Gewertz, B. L. (2014). Flow disruptions during trauma care. *World journal of surgery*, 38(2), 314-321.
- Spiess, B.D., Rotruck, J., McCarthy, H., Suarez-Wincosci, O., Kasirajan, V., Wahr, J., & Shappell, S. (2015). Human Factors analysis of a near-miss event: Oxygen supply failure during cardiopulmonary bypass. *Journal of Cardiothoracic and Vascular Anesthesia*, 29 (1), 204-209.

- Smith, M. J., & Sainfort, P. C. (1989). A balance theory of job design for stress reduction. *International journal of industrial ergonomics*, 4(1), 67-79.
- Smith, L., & Coleman, V. (2010). *Child and family-centered healthcare: Concept, theory, and practice*. Houndmills, Basingstoke: Palgrave Macmillan.
- Stampalia, H. J. (2006). Inadequate Staffing [Scholarly project]. Retrieved 2016, from [https://www.qu.edu/prebuilt/pdf/SchoolLaw/LawReviewLibrary/12_25QLR173\(2006-2007\).pdf](https://www.qu.edu/prebuilt/pdf/SchoolLaw/LawReviewLibrary/12_25QLR173(2006-2007).pdf)
- Steinemann, S., Berg, B., Skinner, A., DiTulio, A., Anzelon, K., Terada, K., ... & Speck, C. (2011). In situ, multidisciplinary, simulation-based teamwork training improves early trauma care. *Journal of surgical education*, 68(6), 472-477.
- Surgeon cut off testicle 'by mistake' at Bury hospital. (2010, April 19). Retrieved May 23, 2016, from http://news.bbc.co.uk/2/hi/uk_news/england/manchester/8630592.stm
- Swayze, S. C., & James, A. (2013, March). The Unfamiliar Catheter. Retrieved May 17, 2016, from <https://psnet.ahrq.gov/webmm/case/294>
- The Joint Commission. (2016, January 26). Retrieved May 04, 2016, from http://www.jointcommission.org/sentinel_event_policy_and_procedures/
- Thiels, C.A., Lal, T.M., Nilenow, J.M., Pasupathy, K.S., Blocker, R.C., Aho, J.M., ... Bingener, J. (2015). Surgical never events and contributing human factors. *Surgery*, 158, 515-21.
- Tired German doctor accidentally kills patient in the UK. (2009, May 04). Retrieved May 28, 2016, from <http://www.thelocal.de/20090504/19045>
- Wallace B., & Ross, A. (2006). *Beyond human error: Taxonomies and safety science*. Boca Raton, FL: CRC/Taylor & Francis.

- Webster, H. (2014, October 27). Is Surgery Safer at a Academic hospital. *U.S. News & World Report*. Retrieved September 28, 2016, from <http://health.usnews.com/health-news/patient-advice/articles/2014/10/27/is-surgery-safer-at-a-teaching-hospital>
- Weigle, S.C. (1998). Using FACETS to model rater training effects. *Language Testing*, 15(2) 263-287.
- Wickens, C. D., Gordon, S. E., Liu, Y., & Lee, J. (1998). An introduction to human factors engineering. Chicago.
- Wiegmann, D.A., & Shappell, S.A. (2001). Human error analysis of commercial aviation accidents: Application of the human factors analysis and classification system (HFACS). *Aviation, Space, and Environmental Medicine*, 72(11),1006-1016.
- Wiegmann, D.A., & Shappell, S.A., (2003). *A human error approach to aviation accident analysis: The human factors analysis and classification system*. Burlington, VT: Ashgate.
- Wiegmann, D., Eggman, A., Elbardissi, A., Parker, S.H., & Sundt, T.M. (2010). Improving cardiac surgical care: A work systems approach. *Applied Ergonomics*, 41 (5), 701-712. [PMCID: PMC2879339](https://pubmed.ncbi.nlm.nih.gov/2879339/)
- Woods, D. D. (2010). *Behind human error*. Ashgate Publishing, Ltd.
- Yates, D., Moore, Moore, D., McCabe, G. (1999). *The Practice of Statistics* (1st Ed.). New York: W.H. Freeman.

APPENDICES

Appendix A: Related publication - Cohen et al., 2015

REVIEW ARTICLE

Evaluating the Reliability of the Human Factors Analysis and Classification System

Tara N. Cohen; Douglas A. Wiegmann; Scott A. Shappell

- INTRODUCTION:** This paper examines the reliability of the Human Factors Analysis and Classification System (HFACS) as tool for coding human error and contributing factors associated with accidents and incidents.
- METHODS:** A systematic review of articles published across a 13-yr period between 2001 and 2014 revealed a total of 14 peer-reviewed manuscripts that reported data concerning the reliability of HFACS.
- RESULTS:** Results revealed that the majority of these papers reported acceptable levels of interrater and intrarater reliability.
- CONCLUSION:** Reliability levels were higher with increased training and sample sizes. Likewise, when deviations from the original framework were minimized, reliability levels increased. Future applications of the framework should consider these factors to ensure the reliability and utility of HFACS as an accident analysis and classification tool.
- KEYWORDS:** HFACS, human error, error analysis.

Cohen TN, Wiegmann DA, Shappell SA. Evaluating the reliability of the Human Factors Analysis and Classification System. *Aerosp Med Hum Perform.* 2015; 86(8):728–735.

Several methodologies have been developed for analyzing human error associated with incidents and accidents across a variety of industries. In recent years, one methodology, the Human Factors Analysis and Classification System (HFACS) has received considerable attention.^{9,14,21,24} The use of HFACS involves the identification and subsequent classification of causal factors into categories based on their presumed underlying etiology. Although the methodology has been widely implemented, the results of research investigating the reliability of the taxonomy are mixed, with some studies showing very high levels of reliability,^{14,31} while others have shown moderate or lower levels.^{19,20}

Reliability, in the present context, involves the degree to which results from an instrument, such as a framework for classifying accident causal factors like HFACS, are consistent or replicable.⁶ Reliability is crucial for ensuring consensus and consistency and is a vital foundation for establishing the validity of such analyses.²⁸ The purpose of this paper, therefore, was to review the literature regarding HFACS to summarize previous findings and identify factors that either enhance or detract from the system's reliability. Implications for ensuring the reliability of HFACS as an accident analysis tool will also be discussed.

The Human Factors Analysis and Classification System

HFACS is a methodology developed by Wiegmann and Shappell²⁷ based in part upon James Reason's "Swiss cheese" model of accident causation.²³ Within this model, Reason proposed that mishaps occur when various components of a system fail to properly interact with one another. Essentially, these failures act as "holes" within the different layers of the "cheese," ultimately leading to breakdowns in productive systems.

Reason's model describes accident causation based on four levels: unsafe acts, preconditions for unsafe acts, unsafe supervision, and organizational influences. This model has become particularly useful in accident investigation as it directs investigators to explore the active and latent failures associated with a given event. However, the model is primarily descriptive and

From Embry-Riddle Aeronautical University, Daytona Beach, FL, and the University of Wisconsin-Madison, Madison, WI.

This manuscript was received for review in December 2014. It was accepted for publication in April 2015.

Address correspondence to: Scott A. Shappell, Department of Human Factors, Embry-Riddle Aeronautical University, 600 S. Clyde Morris Blvd., Daytona Beach, FL 32119; shappe88@erau.edu.

Reprint & Copyright © by the Aerospace Medical Association, Alexandria, VA.

DOI: 10.3357/AMHP.4218.2015

Appendix B: Related publication - Ergai et al., 2016

Safety Science 82 (2016) 393–398



Contents lists available at ScienceDirect

Safety Science

journal homepage: www.elsevier.com/locate/ssci

Assessment of the Human Factors Analysis and Classification System (HFACS): Intra-rater and inter-rater reliability



Awatef Ergai^e, Tara Cohen^a, Julia Sharp^c, Doug Wiegmann^b, Anand Gramopadhye^d, Scott Shappell^{a,*}

^a Department of Human Factors, Embry-Riddle Aeronautical University, United States

^b Department of Industrial and Systems Engineering, University of Wisconsin – Madison, United States

^c Department of Mathematical Sciences, Clemson University, United States

^d College of Engineering and Science, Clemson University, United States

^e College of Engineering, Northeastern University, United States

ARTICLE INFO

Article history:

Received 5 March 2015

Received in revised form 11 July 2015

Accepted 21 September 2015

Keywords:

HFACS

Inter-rater reliability

Intra-rater reliability

Classification

Human-error

ABSTRACT

The Human Factors Analysis and Classification System (HFACS) is a framework for classifying and analyzing human factors associated with accidents and incidents. The purpose of the present study was to examine the inter- and intra-rater reliability of the HFACS data classification process.

Methods: A total 125 safety professionals from a variety of industries were recruited from a series of two-day HFACS training workshops. Participants classified 95 real-world causal factors (five causal factors for each of the 19 HFACS categories) extracted from a variety of industrial accidents. Inter-rater reliability of the HFACS coding process was evaluated by comparing performance across participants immediately following training and intra-rater reliability was evaluated by having the same participants repeat the coding process following a two-week delay.

Results: Krippendorff's Alpha was used to evaluate the reliability of the coding process across the various HFACS levels and categories. Results revealed the HFACS taxonomy to be reliable in terms of inter- and intra-rater reliability, with the latter producing slightly higher Alpha values.

Conclusion: Results support the inter- and intra-rater reliability of the HFACS framework but also reveal additional opportunities for improving HFACS training and implementation.

© 2015 Elsevier Ltd. All rights reserved.

1. Introduction

The role that human factors plays in the genesis of accidents and incidents is indisputable. However, the approach one takes when investigating and identifying human factors with respect to accidents and incidents varies depending on an individual's workplace, their background, training, and knowledge of the field. One methodology that has proven particularly useful across domains is the Human Factors Analysis and Classification System (HFACS; Wiegmann and Shappell, 2003). Indeed the HFACS methodology has been employed in a variety of industrial settings such as aviation (Li and Harris, 2006), mining (Patterson and Shappell, 2010), maritime (Chen et al., 2013), rail (Reinach and Viale, 2006), and medicine (ElBardissi et al., 2007).

HFACS is a human error taxonomy based on Reason's well-known "Swiss Cheese" model of accident causation (Reason,

1990). The structure of HFACS is hierarchical, defining nineteen causal factor categories within four levels or tiers (see Fig. 1). The four tiers include *unsafe acts*, *preconditions for unsafe acts*, *unsafe supervision*, and *organizational influences*. Each level is dependent on the previous one and factors are assumed to progress from active to latent conditions as they progress up the hierarchy from *unsafe acts* to *organizational influences*. A brief description of the HFACS framework is presented in Table 1 (for a more complete description see Wiegmann and Shappell, 2003).

One of the fundamental applications of HFACS involves the classification of incident/accident causal factors into the HFACS causal categories. This classification or coding process is often performed on pre-existing causal factors associated with events that were not originally investigated using HFACS. Rather, the HFACS framework is applied post hoc in attempt to identify meaningful trends in the human causal factors that were not apparent in the original data structure. The reliability of this HFACS coding process impacts the subsequent validity and utility of the HFACS output. If more than one person codes the same causal factors differently, or if the coding results vary for the same person over time, the final

* Corresponding author at: Department of Human Factors, Embry-Riddle Aeronautical University, 600 S Clyde Morris Blvd, Daytona Beach, FL 32114-3900, United States. Tel.: +1 (405)640 5479.

<http://dx.doi.org/10.1016/j.ssci.2015.09.028>

0925-7535/© 2015 Elsevier Ltd. All rights reserved.

Original Article

Identifying workflow disruptions in the cardiovascular operating room

T. N. Cohen,¹ J. S. Cabrera,² O. D. Sisk,¹ K. L. Welsh,¹ J. H. Abernathy,³ S. T. Reeves,⁴
D. A. Wiegmann,⁵ S. A. Shappell⁶ and A. J. Boquet⁷

1 Doctoral Student, *2* Graduate Student, *6* Chair, *7* Associate Professor, Department of Human Factors, Embry-Riddle Aeronautical University, Daytona Beach, Florida, USA

3 Professor, *4* Professor and Chair, Department of Anesthesia and Perioperative Medicine, Medical University of South Carolina, Charleston, South Carolina, USA

5 Associate Professor, University of Wisconsin–Madison, Madison, Wisconsin, USA

Summary

The objectives of this study were to identify the frequency and nature of flow disruptions in the operating room with respect to three cardiac surgical team members: anaesthetists; circulating nurses; and perfusionists. Data collected from 15 cases and coded using a human factors taxonomy identified 878 disruptions. Significant differences were identified in frequency relative to discipline type. Circulating nurses experienced more coordination disruptions (χ^2 (2, N = 110) = 7.136, $p < 0.028$) and interruptions (χ^2 (2, N = 427) = 29.743, $p = 0.001$) than anaesthetists and perfusionists, whereas anaesthetists and perfusionists experienced more layout issues than circulating nurses (χ^2 (2, N = 153) = 48.558, $p = 0.001$). Time to resolve disruptions also varied among disciplines (λ (12, 878) = 5.186, $p = 0.000$). Although most investigations take a one-size fits all approach in addressing disruptions to flow, this study demonstrates that targeted interventions must focus on differences with respect to individual role.

Correspondence to: A. J. Boquet

Email: boque007@erau.edu

Accepted: 8 April 2016

Keywords: quality measures-patient care; root cause analysis: essential elements

Introduction


Patient safety and quality of care are of utmost concern in any healthcare setting. However, delivering optimum care can be a challenging task, especially in a complex medical venue. One of the most demanding environments in clinical medicine is the operating room. Due to high workload, advanced technology, and involvement of multiple, interdependent medical specialties, the operating room is a unique environment within which to examine threats to patient safety. Indeed, at least half of all adverse events experienced by hospitalised patients are related to surgical

procedures, and three-quarters of these events occur in the operating room [1, 2].

Over the last few years, researchers have begun investigating factors that may have an impact on surgical performance and patient safety in the cardiovascular operating room (CVOR). Specifically, the literature has focused on identifying events that may impede the ability of the team to mentally focus on critical tasks, thus threatening patient safety. Such events have been defined as flow disruptions, or 'deviations from the natural progression of an operation, thereby potentially compromising the safety of the operation' [3]. Flow

Appendix D: Related publication - Cohen et al., 2016

Article

— AMERICAN COLLEGE OF —
Medical QualityAmerican Journal of Medical Quality
1–7
© The Author(s) 2016
Reprints and permissions:
sagepub.com/journalsPermissions.nav
DOI: 10.1177/1062860616675230
ajmq.sagepub.com
**Coding Human Factors Observations
in Surgery****Tara N. Cohen, MS¹, Douglas A. Wiegmann, PhD²,
Scott T. Reeves, MD, MBA³, Albert J. Boquet, PhD¹,
and Scott A. Shappell, PhD¹****Abstract**

The reliability of the Human Factors Analysis and Classification System (HFACS) for classifying retrospective observational human factors data in the cardiovascular operating room is examined. Three trained analysts independently used HFACS to categorize observational human factors data collected at a teaching and nonteaching hospital system. Results revealed that the framework was substantially reliable overall (Study I: $k = 0.635$; Study II: $k = 0.642$). Reliability increased when only preconditions for unsafe acts were investigated (Study I: $k = 0.660$; Study II: $k = 0.726$). Preconditions for unsafe acts were the most commonly identified issues, with HFACS categories being similarly populated across both hospitals. HFACS is a reliable tool for systematically categorizing observational data of human factors issues in the operating room. Findings have implications for the development of a HFACS tool for proactively collecting observational human factors data, eliminating the necessity for classification post hoc.

Keywords

HFACS, human error, error analysis, latent failures, CVOR

The Human Factors Analysis and Classification System (HFACS) is a commonly used tool for analyzing human factors issues associated with accidents across a variety of industries, including aviation, mining, manufacturing, and health care.^{1–5} HFACS consists of 4 levels of system failure as originally described by Reason's "Swiss cheese" model of accident causation: unsafe acts, preconditions for unsafe acts, unsafe supervision, and organizational influences.⁶ Shappell and Wiegmann⁷ translated this theory into practice by developing a set of causal factors at each level of the model to facilitate the categorization and analysis of human factors associated with accidents (Table 1).

Several studies have been conducted to explore the reliability of HFACS as a tool for classifying causal factors associated with accidents. Typically, these studies involve multiple coders who independently read through a set of accident reports and then classify each finding or causal factor based on the HFACS framework. A recent review of this body of literature⁸ found that the vast majority of HFACS reliability studies have shown consistently high levels of interrater reliability (agreement between independent coders) and intrarater reliability (ability of the same coders to replicate their own classifications after some delay).

Explorations of HFACS reliability for coding data other than accident causal factors have been limited. Unlike

accident investigations that are reactive, other approaches to safety can be proactive, attempting to identify threats to safety before accidents happen. Sources of such proactive data often include the direct observation of work-related activities and the identification of conditions that potentially threaten safe performance.

Within health care and surgery in particular, an observational root cause analysis approach will be utilized to identify human factors issues that may lead to patient harm. Several observational studies have documented the myriad work system factors in the operating room that disrupt workflow and surgical team performance.^{9–13} However, these studies often involve several hours of observations that produce hundreds of data points or instances of potential human factors problems, which makes the analysis and interpretation of data arduous. Consequently, identifying targeted interventions for reducing specific threats to patient safety is difficult.

¹Embry-Riddle Aeronautical University, Daytona Beach, FL²University of Wisconsin–Madison, Madison, WI³Medical University of South Carolina, Charleston, SC**Corresponding Author:**

Scott A. Shappell, PhD, Department of Human Factors, Embry-Riddle Aeronautical University, 600 S Clyde Morris Blvd, Daytona Beach, FL.
Email: shappe88@my.erau.edu