12-2007

Creating a Culture of Safety: The Influence of Medical Simulation on the Attitudes of Surgical Team Members

Diane J. Jones

Western Michigan University

Follow this and additional works at: https://scholarworks.wmich.edu/dissertations

Part of the Medicine and Health Sciences Commons

Recommended Citation

Jones, Diane J., "Creating a Culture of Safety: The Influence of Medical Simulation on the Attitudes of Surgical Team Members" (2007). Dissertations. 880.
https://scholarworks.wmich.edu/dissertations/880

This Dissertation-Open Access is brought to you for free and open access by the Graduate College at ScholarWorks at WMU. It has been accepted for inclusion in Dissertations by an authorized administrator of ScholarWorks at WMU. For more information, please contact maira.bundza@wmich.edu.
CREATING A CULTURE OF SAFETY: THE INFLUENCE
OF MEDICAL SIMULATION ON THE ATTITUDES
OF SURGICAL TEAM MEMBERS

by

Diane J. Jones

A Dissertation
Submitted to the
Faculty of The Graduate College
in partial fulfillment of the
requirements for the
Degree of Doctor of Philosophy
Interdisciplinary Health Studies

Western Michigan University
Kalamazoo, Michigan
December 2007
CREATING A CULTURE OF SAFETY: THE INFLUENCE OF MEDICAL SIMULATION ON THE ATTITUDES OF SURGICAL TEAM MEMBERS

Diane J. Jones, Ph.D.

Western Michigan University, 2007

The objective of this study was to determine change in individual and aggregate attitudes of the members of two subspecialty surgery teams (cardiac and vascular) in six domains (teamwork climate, safety climate, job satisfaction, stress recognition, perception of management, and working conditions) following medical team simulation and debriefing exercises. The study was a pre-post partially randomized controlled, quasi-experimental study that took place in a 545-bed hospital with a high-volume cardiovascular surgery program. Participants included interdisciplinary team members caring for cardiac and vascular surgery patients.

Main outcome measures included change in responses to safety attitudes questionnaires (SAQs) administered before, 1 week after, and 6 weeks after medical team simulation and debriefing exercises. Statistically significant changes were observed in several domains and occurred at various points during the study period. Consistent with the design of team simulation, the greatest positive effect was observed in the teamwork climate domain. An apparent spill-over positive effect was also observed in control group members, which implies the potential for team simulation exercises to have an effect beyond individual participants, extending change into the broader culture of an organization. Additionally, SAQ scores were analyzed to identify domains at risk for

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
adverse effects on patient safety, creating further opportunity for interventions to improve the organization’s safety culture.
ACKNOWLEDGMENTS

I dedicate this dissertation work to my niece, Jennifer Bechard, who has been an inspiration to my entire family as she battles with noncommunicating hydrocephalus. As a patient, she has endured 60 operations during the last 10 years to revise lumbar-peritoneal shunts to treat her condition. Her tenacity is remarkable and her positive outlook despite her suffering helped keep everything in perspective for me personally as I completed this project. I pray every day that those caring for her recognize the importance of their teamwork in keeping her safe.

I owe special thanks to my mother (Joan), sisters and brothers (Dan, Donna, Dave, Therese, Denise, Debbie, and John), nieces and nephews (Paula, Erin, Chris, Leah, Ryan, Doug, Mike, Bernadette, Jennifer, Genevieve, Lauren, Andrew, Tony, Jacquie, and Christian). Their support and understanding as I missed family events to complete this manuscript are examples of their unwavering love and encouragement they have provided for me my entire life. By their influence and direction, I have achieved all that I have and all that I am.

I have also been fortunate to have an adopted family. The Stelzer and Hance families have been an integral part of my life and have provided me with love, encouragement, and support for many years. Dr. Paul and Mary Delle (Hance) Stelzer, your generosity and gentle spirits touch everyone fortunate enough to know you. You accepted me into your family when I so desperately missed my own, and truly opened my eyes to helping others as a way of life; Paul, I will be forever indebted to you for your
Acknowledgments—Continued

willingness to train me to be the cardiothoracic PA that I am, and will always admire the way you share your faith with your patients.

My work family has also been integral to my ability to complete this dissertation. The surgeons I have worked for throughout my career have always been encouraging to my personal and professional growth and I will always be indebted for their support. My PA and nurse colleagues helped alleviate some of my clinical and administrative responsibilities so I could devote time to this project; they, too, will be relieved when it is finally finished. They along with all of the other participants who agreed to take part in this study have been committed to patient safety and have allowed me to expose all of us in order to improve the care we provide for our patients. This willingness says a lot about their character and commitment to what really matters: the patient.

I would like to thank my dissertation committee, chaired by Dr. Nickola Nelson, Dr. Mark Cowen, Dr. William Hamman, and Dr. Kieran Fogarty. Their commitment to education and research is inspiring; the countless hours they spent guiding me through this study are greatly appreciated. I would also like to thank Dr. Bryan Sexton who has committed a great deal of his professional life to improving safety in both aviation and medicine. His passion for his work and all of the data is very evident; his willingness to guide me through my first use of his SAQ survey instrument is a testament to his commitment to safety.

To Dr. Hamman’s team at the Michigan Center of Excellence for Simulation Research center in Battle Creek, their simulation work is very labor intensive and
Acknowledgments—Continued

consuming. It will no doubt contribute greatly to the formation and training of medical teams that will create a safer environment for patients.

My classmates Dr. Eric Vangsnes and Dr. Lori Pearl-Kraus preceded me in completing their dissertations; their support and encouragement in helping me stay on course to finish will never be forgotten. Their friendship has made this process much more enjoyable.

Finally, I would like to thank my PA colleague, study coordinator, and friend, LaWaun Hance. Her support and calming influence through prayer and example have made an incredible difference in my life and in my ability to complete this project.

Each person named herein has been a blessing to my life for which I thank God. Through them, he has shown himself faithful to me. I first came to know his love for me by example through my father, the original D.J. Jones, who formed my image of what a Father is. Without my Father’s constant grace and abiding love, I simply would not be.

Diane J. Jones
# TABLE OF CONTENTS

ACKNOWLEDGMENTS ................................................................................................ ii
LIST OF TABLES............................................................................................................. xii
LIST OF FIGURES ....................................................................................................... xiii

CHAPTER

I. INTRODUCTION ................................................................................................. 1
   Statement of the Problem......................................................................................... 1
   Significance of the Research.................................................................................. 6
   Setting for the Research ....................................................................................... 8
   Research Questions and Hypotheses ..................................................................... 9
   Definitions of Variable Levels Relevant to Research Questions ....................... 10
      Independent Variables .................................................................................... 10
      Dependent Variables ...................................................................................... 11

II. LITERATURE REVIEW .................................................................................. 12
   Institutional Safety Culture Tools ....................................................................... 13
   Error Reporting .................................................................................................. 14
   Punitive Versus Nonpunitive Environments and PEERS .................................. 14
   Culture of Safety Survey ................................................................................... 17
   Medication Safety Program .............................................................................. 18
   Computerized Physician Order Entry (CPOE) .................................................. 18
   Patient Safety Rounds ...................................................................................... 21
# Table of Contents—Continued

## CHAPTER

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patient Safety Rounds</td>
<td>21</td>
</tr>
<tr>
<td>Patient Safety Rounds</td>
<td>22</td>
</tr>
<tr>
<td>Team Oriented Medical Simulation</td>
<td>24</td>
</tr>
<tr>
<td>In-situ® Simulation and Debriefing</td>
<td>25</td>
</tr>
<tr>
<td>Debriefing</td>
<td>27</td>
</tr>
<tr>
<td>Culture and Attitude Domains</td>
<td>29</td>
</tr>
<tr>
<td>Teamwork Climate</td>
<td>30</td>
</tr>
<tr>
<td>Safety Climate</td>
<td>32</td>
</tr>
<tr>
<td>Job Satisfaction</td>
<td>33</td>
</tr>
<tr>
<td>Stress Recognition</td>
<td>34</td>
</tr>
<tr>
<td>Perception of Management</td>
<td>36</td>
</tr>
<tr>
<td>Working Conditions</td>
<td>37</td>
</tr>
<tr>
<td>Surgical Teams and Job Classifications</td>
<td>38</td>
</tr>
<tr>
<td>Cardiac Surgery Teams</td>
<td>38</td>
</tr>
<tr>
<td>Vascular Surgery Teams</td>
<td>39</td>
</tr>
<tr>
<td>Interdisciplinary Team Composition</td>
<td>40</td>
</tr>
<tr>
<td>Summary</td>
<td>51</td>
</tr>
</tbody>
</table>

## III. METHODS ................................................................. 53

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preliminary Steps</td>
<td>58</td>
</tr>
<tr>
<td>Participants</td>
<td>59</td>
</tr>
<tr>
<td>Eligibility Criteria</td>
<td>59</td>
</tr>
</tbody>
</table>

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
Table of Contents—Continued

CHAPTER

Recruitment ................................................................................................ 59
Sample Size ................................................................................ 61
Procedure .......................................................................................................... 63
Randomization ........................................................................................... 63
Simulation and Debriefing ........................................................................ 65
Settings and Location of Data Collection ................................................ 66
Instrument ......................................................................................................... 67
Survey Domains ......................................................................................... 69
Survey Administrative Guidelines ............................................................ 72
Reliability .................................................................................................... 72
Analysis ............................................................................................................. 73
Data Management and Processing ............................................................ 73
Calculation of Domain Scores .................................................................. 73
Data Analysis ................................................................................................ 77

IV. RESULTS ............................................................................................................... 79

Research Questions and Hypotheses .............................................................. 79
Descriptive Analyses of Sample Characteristics ........................................... 81
Operating Room Safety Attitudes Questionnaire Descriptives ................. 83
Intensive Care Unit Safety Attitudes Questionnaire Descriptives ............. 87
Nonparametric Test of Normality ................................................................. 90
Analysis by Research Question ................................................................. 92

vii

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
# Table of Contents—Continued

## CHAPTER

<table>
<thead>
<tr>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group Equivalence at Baseline</td>
<td>92</td>
</tr>
<tr>
<td>Research Question Analysis</td>
<td>95</td>
</tr>
<tr>
<td>Operating Room Safety Attitude Questionnaire Analyses</td>
<td>96</td>
</tr>
<tr>
<td>Teamwork Climate in the OR With Group as the Unit of Analysis</td>
<td>96</td>
</tr>
<tr>
<td>Teamwork Climate in the OR With Individual as the Unit of Analysis</td>
<td>100</td>
</tr>
<tr>
<td>Additional Analyses of Low OR Teamwork Climate Scores at Baseline</td>
<td>101</td>
</tr>
<tr>
<td>General Estimated Equation of Teamwork Climate Scores in OR</td>
<td>102</td>
</tr>
<tr>
<td>Post-hoc Analyses on OR Teamwork Questions With Paired t Tests</td>
<td>104</td>
</tr>
<tr>
<td>Safety Climate in the OR With Group as the Unit of Analysis</td>
<td>105</td>
</tr>
<tr>
<td>Climate in the OR With Individual as the Unit of Analysis</td>
<td>107</td>
</tr>
<tr>
<td>Post-hoc Analyses on ICU Safety Climate Questions With Paired t Test</td>
<td>109</td>
</tr>
<tr>
<td>Job Satisfaction in the OR With Group as the Unit of Analysis</td>
<td>110</td>
</tr>
<tr>
<td>Job Satisfaction in the OR With Individual as the Unit of Analysis</td>
<td>112</td>
</tr>
<tr>
<td>Stress Recognition in the OR With Group as the Unit of Analysis</td>
<td>114</td>
</tr>
<tr>
<td>Stress Recognition in the OR With Individual as the Unit of Analysis</td>
<td>116</td>
</tr>
<tr>
<td>Post-hoc Analyses on OR Stress Recognition Questions With Paired t Tests</td>
<td>117</td>
</tr>
<tr>
<td>Perception of Management in the OR With Group as the Unit of Analysis</td>
<td>118</td>
</tr>
</tbody>
</table>

viii

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
Table of Contents—Continued

CHAPTER

Perception of Management in the OR With Individual as the Unit of Analysis ............................................................. 120

Post-hoc Analyses on Perception of Management Climate Questions With Paired t Test .............................................. 121

Working Conditions in the OR With Group as the Unit of Analysis .... 123

Working Conditions in the OR With Individual as the Unit of Analysis ........................................................................ 124

Post-hoc Analyses on OR Working Conditions Climate Questions With Paired t Test .................................................. 127

Intensive Care Unit Safety Attitude Questionnaire Analyses .......... 127

Teamwork Climate (ICU SAQ) With Group as the Unit of Analysis .................................................................................. 127

Teamwork Climate (ICU SAQ) With Individual as the Unit of Analysis ............................................................................. 129

Additional Analyses of Low Teamwork Climate Scores at Baseline .................................................................................... 131

General Estimated Equation of Teamwork Climate Scores in ICU ...... 132

Post-hoc Analyses on ICU Teamwork Climate Questions With Paired t Test ................................................................. 133

Safety Climate (ICU SAQ) With Group as the Unit of Analysis ............ 134

Safety Climate (ICU SAQ) With Individual as the Unit of Analysis .................................................................................. 136

Post-hoc Analyses on Safety Climate Questions With Paired t Test..... 138

Job Satisfaction (ICU SAQ) With Group as the Unit of Analysis....... 138

Job Satisfaction (ICU SAQ) With Individual as the Unit of Analysis .................................................................................. 140
## Table of Contents—Continued

### CHAPTER

- Post-hoc Analyses on Job Satisfaction Questions With Paired *t* Test .... 142
- Stress Recognition (ICU SAQ) With Group as the Unit of Analysis ..... 142
- Stress Recognition (ICU SAQ) With Individual as the Unit of Analysis ................................................................................................... 144
- Post-hoc Analyses on ICU Stress Recognition Questions With Paired *t* Test ................................................................. 146
- Perception of Management (ICU) With Group as the Unit of Analysis ................................................................................................... 147
- Perception of Management (ICU SAQ) With Individual as the Unit of Analysis ................................................................................................... 149
- Post-hoc Analyses on ICU Perception of Management Questions With Paired *t* Test ................................................................. 151
- Working Conditions (ICU SAQ) With Group as the Unit of Analysis ................................................................................................... 152
- Working Conditions (ICU SAQ) With Individual as the Unit of Analysis ................................................................................................... 153
- Post-hoc Analyses on ICU Working Conditions Questions With Paired *t* Test ................................................................. 155
- Quality of Communication and Collaboration ................................................. 156
- Answers to Open-Ended Questions ................................................................. 158
- Chapter Summary .............................................................................................. 162

### V. DISCUSSION .......................................................................................................... 163

- Chapter Preview ................................................................................................. 163
- Summary of Research Design and Results ...................................................... 163
- Discussion of Results ......................................................................................... 167
### Table of Contents—Continued

#### CHAPTER

<table>
<thead>
<tr>
<th>Topic</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Teamwork Climate</td>
<td>167</td>
</tr>
<tr>
<td>Safety Climate</td>
<td>173</td>
</tr>
<tr>
<td>Job Satisfaction</td>
<td>175</td>
</tr>
<tr>
<td>Stress Recognition</td>
<td>176</td>
</tr>
<tr>
<td>Perception of Management</td>
<td>179</td>
</tr>
<tr>
<td>Working Conditions</td>
<td>181</td>
</tr>
<tr>
<td>Quality of Communication and Collaboration</td>
<td>182</td>
</tr>
<tr>
<td>Broader Implications of Results</td>
<td>183</td>
</tr>
<tr>
<td>Study Limitations</td>
<td>185</td>
</tr>
<tr>
<td>Randomization</td>
<td>185</td>
</tr>
<tr>
<td>Study Design</td>
<td>188</td>
</tr>
<tr>
<td>Sample Size</td>
<td>189</td>
</tr>
<tr>
<td>Survey Instrument Tool</td>
<td>191</td>
</tr>
<tr>
<td>Recommendation for Future Studies</td>
<td>192</td>
</tr>
<tr>
<td>Conclusions</td>
<td>194</td>
</tr>
</tbody>
</table>

**REFERENCES** ................................................................. 202

**APPENDICES**

<table>
<thead>
<tr>
<th>Appendix</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>A.</td>
<td>Human Subjects Institutional Review Board Letters of Approval</td>
<td>213</td>
</tr>
<tr>
<td>B.</td>
<td>Results Summary Tables</td>
<td>216</td>
</tr>
<tr>
<td></td>
<td>Table Title</td>
<td>Page</td>
</tr>
<tr>
<td>---</td>
<td>------------------------------------------------------------------------------</td>
<td>-------</td>
</tr>
<tr>
<td>1</td>
<td>Comparison of Punitive and Nonpunitive Environments and Reporting Systems</td>
<td>15</td>
</tr>
<tr>
<td>2</td>
<td>Participants by Location and Job Category</td>
<td>62</td>
</tr>
<tr>
<td>3</td>
<td>Likert Scale Conversion</td>
<td>74</td>
</tr>
<tr>
<td>4</td>
<td>Definition of Domain Scores</td>
<td>75</td>
</tr>
<tr>
<td>5</td>
<td>Percent Response Rates Based on Time of Survey Distribution</td>
<td>83</td>
</tr>
<tr>
<td>6</td>
<td>Descriptive Analyses of OR Questionnaire Respondents</td>
<td>84</td>
</tr>
<tr>
<td>7</td>
<td>Central Tendency of Baseline OR SAQ Climate Scores</td>
<td>85</td>
</tr>
<tr>
<td>8</td>
<td>Descriptive Analyses: ICU Questionnaire Respondents</td>
<td>88</td>
</tr>
<tr>
<td>9</td>
<td>Central Tendency of Baseline ICU SAQ Climate Scores</td>
<td>89</td>
</tr>
<tr>
<td>10</td>
<td>OR Effect of Simulation on Individual Teamwork Scores</td>
<td>103</td>
</tr>
<tr>
<td>11</td>
<td>ICU Effect of Simulation on Individual Teamwork Scores</td>
<td>133</td>
</tr>
<tr>
<td>12</td>
<td>ICU SAQ All Participants – Communication and Collaboration</td>
<td>157</td>
</tr>
<tr>
<td>13</td>
<td>OR SAQ All Participants – Communication and Collaboration</td>
<td>158</td>
</tr>
<tr>
<td>14</td>
<td>Baseline Recommendations for Improving Safety</td>
<td>159</td>
</tr>
<tr>
<td>15</td>
<td>One Week Recommendations for Improving Safety</td>
<td>160</td>
</tr>
<tr>
<td>16</td>
<td>Six Week Recommendations for Improving Safety</td>
<td>161</td>
</tr>
</tbody>
</table>
# LIST OF FIGURES

1. Study Design ............................................................................................................ 55
2. Teamwork Climate Across Groups OR SAQ – All Participants ......................... 97
3. Teamwork Climate Across Groups OR SAQ – By Experimental Condition.... 98
4. Teamwork Climate Across Groups OR SAQ – Randomized-Experimental and Control Only .................................................................................................... 99
5. Safety Climate Across Groups OR SAQ – All Participants ..................................105
6. Safety Climate Across Groups OR SAQ – By Experimental Condition .......... 106
7. Safety Climate Across Groups OR SAQ – Randomized-Experimental and Control Only ......................................................................................................107
8. Job Satisfaction Across Groups OR SAQ – All Participants...............................110
9. Job Satisfaction Across Groups OR SAQ – By Experimental Condition .......... 111
10. Job Satisfaction Across Groups OR SAQ – Randomized-Experimental and Control Only ......................................................................................................112
11. Stress Recognition Across Groups OR SAQ – All Participants .........................114
12. Stress Recognition Across Groups OR SAQ – By Experimental Condition ....... 115
13. Stress Recognition Across Groups OR SAQ – Randomized-Experimental and Control Only ....................................................................................................116
14. Perception of Management Across Groups OR SAQ – All Participants .......... 119
15. Perception of Management Across Groups OR SAQ – By Experimental Condition............................................................................................................. 119
16. Perception of Management Across Groups OR SAQ – Randomized-Experimental and Control Only ..............................................................................120
17. Working Conditions Across Groups OR SAQ – All Participants .......................124
List of Figures—Continued

18. Working Conditions Across Groups OR SAQ – By Experimental Condition... 125
19. Working Conditions Across Groups OR SAQ – Randomized-Experimental and Control Only .............................................................................. 125
20. Teamwork Climate Across Groups ICU SAQ – All Participants............... 128
21. Teamwork Climate Across Groups ICU SAQ – By Experimental Condition... 129
22. Teamwork Climate Across Groups ICU SAQ – Randomized-Experimental and Control Only .............................................................................. 130
23. Safety Climate Across Groups ICU SAQ – All Participants...................... 135
24. Safety Climate Across Groups ICU SAQ – By Experimental Condition...... 136
25. Safety Climate Across Groups ICU SAQ – Randomized-Experimental and Control Only .............................................................................. 136
26. Job Satisfaction Across Groups ICU SAQ – All Participants ...................... 139
27. Job Satisfaction Across Groups ICU SAQ – By Experimental Condition........ 139
28. Job Satisfaction Across Groups ICU SAQ – Randomized-Experimental and Control Only .............................................................................. 140
29. Stress Recognition Across Groups ICU SAQ – All Participants................ 143
30. Stress Recognition Across Groups ICU SAQ – By Experimental Condition.... 144
31. Stress Recognition Across Groups ICU SAQ – Randomized-Experimental and Control Only .............................................................................. 145
32. Perception of Management Across Groups ICU SAQ – All Participants ........ 147
33. Perception of Management Across Groups ICU SAQ – By Experimental Condition ........................................................................................................ 148
34. Perception of Management Across Groups ICU SAQ – Randomized-Experimental and Control Only .............................................................................. 149
35. Working Conditions Across Groups ICU SAQ – All Participants............. 152
36. Working Conditions Across Groups ICU SAQ – By Experimental Condition ................................................................. 153

37. Working Conditions Across Groups ICU SAQ – Randomized-Experimental and Control Only ........................................ 154
CHAPTER I
INTRODUCTION

This is a study of individual and aggregate attitudes of the members of two subspecialty surgery teams (cardiac surgery and vascular surgery) in two environments (operating room [OR] and intensive care unit/progressive care unit [ICU]) regarding teamwork climate, safety climate, job satisfaction, stress recognition, and working conditions after being exposed to medical team simulation and debriefing. The study used a pre-post randomized experimental design and took place in a 545-bed hospital in the Midwest with a high-volume cardiovascular surgery program.

This chapter provides an overview of the problem and discusses the significance of the research. It also introduces the research questions the study was designed to answer and posits a hypothesis for each. The independent and dependent variables are also defined.

Statement of the Problem

Primum non nocere (first do no harm) has always been a guiding principle of healthcare professionals, but recent studies have shown that, due to medical error, patients actually are placed at risk by the very institutions charged with their wellbeing (Nieva & Sorra, 2003). Patient safety has catapulted to the forefront of endeavors within the healthcare industry, particularly since the publication of the Institute of Medicine’s 1998 report citing as many as 98,000 deaths per year due to medical error (Kohn,
One way to improve patient safety is to change the underlying culture of an institution (Espin, Levinson, Regehr, Baker, & Lingard, 2006; Leape, n.d.; Nieva & Sorra, 2003; Pizzi, Goldfarb, & Nash, 2001). Before a culture can be changed intentionally, however, key components must be identified and defined. Helmreich and Merritt (2001) defined culture as

> a complex framework of national, organizational and professional attitudes and values within which groups and individuals function. The power of culture often goes unrecognized since it represents “the way we do things here”—the natural and unquestioned mode of viewing the world. However, the reality and strength of culture becomes salient when we work with a new group (whether in a new country, a new organization, or a new profession) and interact with people who have well-established norms and values. (p. 1)

According to the Advisory Committee on the Safety of Nuclear Installations (Health & Safety Commission, 1993), safety culture can be defined as “the product of individual and group values, attitudes, perceptions, competencies, and patterns of behavior that determine the commitment to, and the style and proficiency of, an organization’s health and safety management” (p. 18). This statement further defined a positive safety culture as one that is “characterized by communications founded on mutual trust, by shared perceptions of the importance of safety and by confidence in the efficacy of preventive measures” (p. 18).

This is in contrast to the traditional safety culture within healthcare, which is sometimes referred to as “blame and shame.” Such a culture discourages open discussion of error and seeks to make individual performance its scapegoat (Nieva & Sorra, 2003). This punitive atmosphere is generally maladaptive to measures that could proactively reduce errors simply through freedom and willingness to acknowledge the potential for error.
The safety culture of an institution is shaped by the aggregate perceptions of individual team members. Safety culture assessment surveys can identify potential deficiencies in an institution’s unique culture and thus can aid assessment, monitoring and improvement of patient safety within an organization (Firth-Cozens, 2003; Kaissi, Johnson, & Kirschbaum, 2003; Nieva & Sorra, 2003; Pizzi et al., 2001; Pronovost & Sexton, 2005; Pronovost et al., 2003; Sexton, Thomas, & Helmreich, 2000; Thomas, Sexton, & Helmreich, 2003). Sexton et al. (2000) reviewed research regarding the use of survey items for different purposes related to safety in healthcare. Irwin (1991) indicated that survey items can be designed in a manner to lead to increased understanding of error. Other research has shown that survey items can predict performance (Helmreich, Foushee, Benson, & Russini, 1986) and are sensitive to training interventions (Gregorich, Helmreich, & Wilhelm, 1990; Helmreich et al., 1986; Irwin, 1991). For example, a positive response to teamwork items on a safety attitudes questionnaire corresponds to positive team function, which correlates to good outcomes in patient safety. Sexton et al. (2000) stated that

attitudes regarding the recognition of stressor effects indicate the degree to which individuals will place themselves in error-inducing conditions, and items regarding hierarchy and teamwork indicate the abilities of team members to manage both threats and errors in a team environment. (p. 745)

Sexton et al. (n.d.) reported that one such survey developed by Sexton, the Safety Attitudes Questionnaire (SAQ), gives “a snapshot of culture in a given area and SAQ data can help to assess the effectiveness of interventions to improve safety” (p. 15). The items on this questionnaire were designed for healthcare teams, but were derived from the Flight Management Attitudes Questionnaire, which is widely used in commercial aviation to assess flight crew member attitudes about such topics as teamwork, speaking
up, leadership communication, and collaborative decision making. Additionally, "when used in a pre-intervention/post-intervention methodology, the SAQ factors have demonstrated sensitivity to quality improvement interventions" (Sexton, n.d., p. 16).

A quality improvement intervention currently being studied by faculty of Western Michigan University’s College of Aviation translates aviation safety culture research to health care settings. It involves simulating a typical but challenging medical team problem-solving situation, followed by a debriefing exercise in which participants actively process what happened during the simulation (Rutherford & Hamman, 2005b). Two primary goals of the simulation project are to influence the culture of how medical teams work together and to determine what interpersonal skills are needed for successful teamwork in healthcare. The intervention creates an opportunity for dialogue that normally would not occur in most organizations, but could help shape perceptions of individual team members regarding safety issues.

Culture in aviation and medicine are thought to be similar because employment in both arenas requires “high levels of interpersonal collaboration, communication and coordination” (Helmreich & Merritt, 2001, p. 1). Research in both the aviation industry and in medicine has the “potential to increase safety and save lives” (Helmreich & Merritt, 2001, p. 4). Simulation in the airline industry has been shown to improve safety practices, so there are broadening attempts to apply this type of approach to medical teams.

Although, in theory, this type of exercise should change the way medical teams interact, a paucity of research exists to support the efficacy of such an approach in medicine for changing practice, or even for changing attitudes among team members that
might lead to changes in practice (W. Rutherford & W. Hamman, personal communication, March 23, 2006). The primary purpose of the current investigation was to determine whether simulation techniques could change self-reported attitudes on the SAQ at a healthcare institution whose surgical teams have participated in simulation and debriefing exercises.

Existing attitudes using various safety culture assessment surveys of operating room personnel have been reported (Flin, Fletcher, McGeorge, Sutherland, & Patey, 2003; Gaba, Howard, & Jump, 1994; Grogan et al., 2004; Sexton et al., 2000). Marshall, Flanagan, Joseph, and Bujor (2006) conducted a study in Melbourne, Australia that assessed attitudes before and after a simulation-based team training program using the operating room SAQ that was used in the current study. Preliminary data from the study did not show that the simulation produced a statistically significant change on attitudes toward teamwork or safety climate as measured with the SAQ (Marshall et al., 2006); however, senior surgeons, managers, and senior clinicians were not involved in the team training in Marshall’s study. This exclusion of key team members could have been the reason for a lack of effect in changing teamwork or safety climate (Marshall et al., 2006).

Marshall, Flanagan, Joseph, and Bujor (2007) administered the SAQ prior to simulation-based team training and 3 months posttraining. The initial SAQ was sent to all staff. There was a 54% return rate on the initial SAQ and a 12% response rate at the time of writing on the follow-up SAQ. The low response rate on the follow-up questionnaire may also have contributed to the inability of the research to detect significant change in teamwork or safety climate. The simulation-based team training was designed after results from the initial SAQ were analyzed. The training participants included “surgical
and anaesthetic specialist nursing staff” (nurse anesthetists and surgical nurses), “anaesthetic specialists and trainees” (anesthesiologists and anesthesia residents), “theatre technicians” (scrub technologists), and “surgical trainees” (surgical residents) (Marshall et al., 2007). No control group was used. Marshall et al. indicated belief that a control group would not be appropriate within any single institution because the organization itself is the study group as opposed to individuals or subgroups. They also recommended that “participation rates by surgeons should be a major focus in the design of future studies” (Marshall et al., 2007).

Many researchers have studied the dynamics of the physician-physician or physician-nurse team (Baggs et al., 1999; Flin et al., 2003; Gaba et al., 1994; Helmreich & Merritt, 2001; Helmreich & Schaefer, 1994; Lingard et al., 2005; Makary et al., 2006a, 2006b; Reader, Flin, Mearns, & Cuthbertson, 2007; Thomas et al., 2003), but fewer studies have examined disciplines outside of these teams (Edmondson, 2003; France et al., 2005; Grogan et al., 2004; Marshall et al., 2006; Undre, Sevdalis, Healey, Darzi, & Vincent, 2006). No studies were found that compared members of cardiac and vascular surgery teams both in the OR and the ICU. Neither were studies found that involved physician assistants, respiratory therapists, radiology technicians, laboratory personnel or the other support staff included in this study.

Significance of the Research

The Pew Health Professions Commission issued a report in 1998 regarding the structure and regulation of the healthcare profession. Among recommendations for establishing competencies for various health professionals was a specific edict for
improving interdisciplinary teamwork. The commission cited a lack of research as an impediment to capitalizing on the unique contributions of interdisciplinary team members to cost and quality of healthcare (O’Neil & Pew Health Professions Commission, 1998). The interdisciplinary teamwork focus of the current study addressed this gap.

This study was designed to determine whether attitudes of a cardiac surgery and vascular surgery interdisciplinary team would change as measured by responses to the OR SAQ and ICU SAQ at baseline (before), 1 week after, and 6 weeks after being exposed to medical team simulation and debriefing. This effect was measured by comparing results for the participants in the experimental simulation approach with survey results gathered at the same time points for control team members who did not take part in the simulation and debriefing exercises.

The Pew Commission also recommended that “states should require that their regulated health care practitioners demonstrate their competence in the knowledge, judgment, technical skills and interpersonal skills relevant to their jobs throughout their careers” (O’Neil & Pew Health Professions Commission, 1998, p. xix). Most simulation projects in medicine tend to focus only on technical skills, whereas Hamman and Rutherford’s simulation research (Hamman et al., 2007) seeks to determine what interpersonal skills are needed for successful teamwork in healthcare. It was expected that the results of this study would help those researchers develop and evaluate a program aimed at promoting continuing competence in teamwork skills. Cooperation between the study hospital and Western Michigan University (WMU) was expected to fulfill another Pew Commission recommendation to have local institutions and educational programs.
describe and demonstrate how interdisciplinary skills are being incorporated into practice (O’Neil & Pew Health Professions Commission, 1998).

Use of a safety culture assessment provides objective insight as to the effectiveness of current measures and future interventions, such as the simulation and debriefing exercise employed in this study. A repeated measure sampling method was used. By sampling attitudes at three points in time, and comparing the results for teams of randomly assigned (when feasible) experimental and control team members, inferences could be drawn about the effect of the simulation and debriefing intervention on team members’ attitudes as expressed on the questionnaires, both immediately after the simulation and 6 weeks later. This extended use of the questionnaires also has practical implications for complementing and expanding prior efforts to build awareness of, and thus to promote a culture of patient safety at the participating hospital.

Setting for the Research

The setting for this research was a 545-bed teaching hospital in a metropolitan area in the Midwest, which employs more than 4,000 individuals. Caring for approximately 2,000 cardiac and vascular surgery patients annually, the hospital was named one of the nation’s Top 100 Cardiovascular Hospitals for 2005. This ranking was awarded by Solucient, a healthcare information products company based in Evanston, Illinois, that provides scorecard-type ranking of hospitals based on specific performance measures.

Prior to the current study the hospital had many safety-promoting tools in place, including an anonymous Potential Error and Event Reporting System (PEERS), the
Keystone ICU Safety Project (a joint project of the Michigan Hospital Association and Johns Hopkins University), a yearly safety culture survey, a medication safety program, computerized physician order entry, and patient safety rounds. The hospital also was currently implementing the American Hospital Association’s quality and patient safety agenda. All of these tools were directed toward enhancing the culture of safety, but they did not address the enhancement of teamwork skills directly as did the simulation and debriefing project and related survey research, which is reported here.

Research Questions and Hypothesis

The purpose of this study was to determine change in attitudes of members of two subspecialty surgery teams (cardiac surgery and vascular surgery) in two settings (OR and ICU) regarding teamwork climate, safety climate, job satisfaction, stress recognition, perception of management, and working conditions after exposure to medical simulation and debriefing by the Michigan Center of Excellence in Simulation Research. The study was designed to answer two primary questions. The research hypotheses below are stated as alternative hypotheses according to the anticipated results prior to conducting the study. Inferential statistical analysis procedures were used to test the null hypothesis form of each research question.

1. Do pre-post-post (baseline, 1 week, and 6 week) attitudes of surgical team members (as measured by an operating room safety attitude questionnaire)

---

1 The Michigan Center of Excellence in Simulation Research is a grant project funded by the Michigan Economic Development Corporation–Tri-Corridor. The Center is located at Western Michigan University’s College of Aviation in Battle Creek, Michigan.
change after being exposed to medical simulation and debriefing, compared with changes at pre-post-post points for members of control group teams?

Research Hypothesis 1: Exposure to team medical simulation and debriefing will result in a change in attitudes regarding teamwork climate, safety climate, job satisfaction, stress recognition, perception of management, and working conditions among members of cardiac and vascular surgery teams as measured by an operating room safety attitude questionnaire.

2. Do pre-post-post (baseline, 1 week, and 6 week) attitudes of surgical team members (as measured by an intensive care unit safety attitude questionnaire) change after being exposed to medical simulation and debriefing, compared with changes at pre-post-post points for members of control group teams?

Research Hypothesis 2: Exposure to team medical simulation and debriefing will result in a change in attitudes regarding teamwork climate, safety climate, job satisfaction, stress recognition, perception of management, and working conditions among members of cardiac and vascular surgery teams as measured by an intensive care unit safety attitude questionnaire.

Definitions of Variable Levels Relevant to Research Questions

Independent Variables

The independent variables (experimental or control, group membership, and time) were manipulated and controlled within this study to provide information about how the effects of simulation and debriefing might affect self-reported attitudes regarding safety climate on an SAQ.
**Dependent Variables**

Changes in safety attitudes of cardiac and vascular surgery team members within the study hospital were measured with the SAQ at three points in time (baseline, 1 week, and 6 weeks). Subscale scores were available in each of six domains: (1) teamwork climate, (2) safety climate, (3) job satisfaction, (4) stress recognition, (5) working conditions, and (6) perception of management on both the OR and ICU versions of the instrument.
CHAPTER II

LITERATURE REVIEW

In 2000, the Institute of Medicine (IOM) published a manuscript entitled "To Err is Human" (Kohn et al., 2000), which reported that at least 44,000 Americans die each year as a result of medical errors and that the number might actually be as high as 98,000 deaths per year. The editors further stated that these numbers are higher than the number of deaths each year from motor vehicle accidents, breast cancer, or AIDS (Kohn et al., 2000). These statistics instantly grabbed the attention of the public despite the fact that studies on medical error have been in the English literature for more than 50 years (Zeman, 1956). The IOM report has been the topic of headlines and television talk shows on a routine basis since its publication.

Other factors such as cost containment, information technology and the transfer of physician-driven decision making to patients and corporations have also prompted promotion of a culture of safety in medicine (The Leapfrog Group, 2004; Millenson, 2002; Small & Barach, 2002;). The IOM report merely accelerated this effort. Some hospitals had tools to develop a culture of safety in place prior to the IOM report; others were forced to develop systems in response to it (Mutter, 2003; Small & Barach, 2002).

Much debate has ensued as to whether the statistics published by the IOM were exaggerated (Leape, 2000; McDonald, Weiner, & Hui, 2000). Some considered the report to be primarily political and lacking scientific merit (Brennan, 2000; Hayward & Hofer, 2001). Regardless, the IOM report heightened interest in patient safety related to medical
errors and has sparked debate and research directed at reducing the incidence of errors in medicine. The recommendations in the IOM report were laid out in a four-tiered approach (Kohn et al., 2000):

1. Establish a national focus to enhance knowledge about patient safety issues;
2. Create mandatory and voluntary error reporting;
3. Raise standards and expectations relating to safety; and
4. Create safety systems within health care organizations through implementation of safe practices.

These recommendations were based on the theory that healthcare organizations need to convert to a culture that believes errors result from system failures, rather than from practitioner failures (Kohn et al., 2000; Nieva & Sorra, 2003). Reason (2000) expressed agreement with the IOM and reported that the person approach to human error, which blames individuals, impedes healthcare safety.

Institutional Safety Culture Tools

It is important to review safety tools currently in place at the study hospital in order to identify the existing culture. These tools include an anonymous Potential Error and Event Reporting System (PEERS) along with a nonpunitive reporting policy, a yearly culture of safety survey, a medication safety program, computerized physician order entry, and patient safety rounds. The study hospital is also currently implementing the American Hospital Association’s quality and patient safety agenda, and is actively participating in the Keystone ICU Safety project, which is a joint project of the Michigan Hospital Association and Johns Hopkins University.
Error Reporting

The reporting of human errors in industries such as aviation, nuclear power technology, oil, gas, and nautical has occurred for several years. Reporting systems developed within these industries have only recently been considered as potential models for reporting medical errors (Barach & Small, 2000; Chassin, 1998; McIntire, 2003; Schmidt, Figlock, & Schmorrow, 2000; Shalala, Herman, & Eisenberg, 2000). The aviation industry has led the way in human error reporting, and many medical error reporting systems are based on the National Aeronautics and Space Administration (NASA) reporting system (Chiles, 2001; Karpf, 2000; Shalala et al., 2000).

Punitive Versus Nonpunitive Environments and PEERS

Table 1 provides a comparative analysis of the characteristics in punitive versus nonpunitive environments and reporting systems. Early error reporting systems enforced a punitive environment that focused on individual performance problems (Bates et al., 2001; Cohen, 2000; ISMP, 1999; Kohn et al., 2000). In such cases, an error made by an individual leads to efforts to identify the culprit, and punishment is then carried out. This punitive environment was in force at the study hospital prior to the 2000 implementation of a confidential/anonymous error reporting system and the 2001 adoption of a nonpunitive policy on error reporting.

When the punitive environment was in place, any error discovered was described in a written incident report, followed by a corrective action imposed on each individual involved. Conlon (2002) reported that this type of punitive environment can be devastating to employees and can limit process improvements. In a punitive environment
individuals are less likely to give detailed information about an incident in order to protect themselves (Cohen, 2000). Without this detailed information, there is no way to get at the root of a problem and to help others avoid the same mistakes.

Table 1

Comparison of Punitive and Nonpunitive Environments and Reporting Systems

<table>
<thead>
<tr>
<th></th>
<th>Punitive</th>
<th>Nonpunitive</th>
</tr>
</thead>
<tbody>
<tr>
<td>Environment Characteristics</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Focus</td>
<td>Individual Performance</td>
<td>System Processes</td>
</tr>
<tr>
<td>Collaboration</td>
<td>Hierarchical</td>
<td>Interdisciplinary</td>
</tr>
<tr>
<td>Feedback</td>
<td>Corrective</td>
<td>Constructive and Proactive</td>
</tr>
<tr>
<td>Basis for Change</td>
<td>Administrative Mandate</td>
<td>Collaborative Suggestions</td>
</tr>
<tr>
<td>Error Reporting</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Error Reporting</td>
<td>Public</td>
<td>Anonymous</td>
</tr>
<tr>
<td>Error Review</td>
<td>Public</td>
<td>Confidential</td>
</tr>
<tr>
<td>Type of Reports</td>
<td>Incidents Only</td>
<td>Near Misses and Incidents</td>
</tr>
</tbody>
</table>

Many authors posit that promotion of patient safety and reduction of medical errors relies on establishing a nonpunitive environment that places a greater value on solving system errors rather than on punishing providers (Asplin, Knopp, Tintinalli, & Waeckerle, 2000; Bates et al., 2001; Bogner, 1994; Cohen, 2000; Finkelstein, 2005; ISMP, 1999; Kohn et al., 2000; Rosenthal & Sutcliffe, 2002; Rozich & Resar, 2001; Wears et al., 2000). A key characteristic of a nonpunitive environment is an anonymous
or confidential reporting system (Fernald et al., 2004; Kerr, 2003; Suresh, 2004). These systems have provisions for reporting near misses as well as actual error incidents, both of which are evaluated to identify possible system breakdowns. There are no reprisals for reporting errors or near misses; in fact, reporting is encouraged as a mechanism to generate opportunities to focus on process improvement.

Administrators at the participating institution began to examine their organization’s error reporting system prior to the IOM manuscript in 2000. Many fundamental questions were examined, including the purpose of reporting and whether to foster a punitive or nonpunitive environment. According to a representative from the institution’s Center for Healthcare Improvement, a literature review regarding error reporting systems was conducted.

Based on this review, the leaders at the institution decided that a nonpunitive culture with an anonymous reporting system would be most beneficial in determining the root cause of errors, which could then be used to change system flaws that allow errors. Several recent reports on the experiences of similar organizations in selecting types of reporting systems and punitive versus non-punitive culture support this decision (Fernald et al., 2004; Mutter, 2003; Wong, Helsinger, & Petry, 2002).

The administrators of the study hospital collaborated with their parent organization to create a new reporting system known as Potential Error/Event Reporting System (PEERS). The developers modeled this system after NASA’s voluntary reporting system (Conlon, 2002). PEERS is a corporate Infonet, web-based application that provides ready access and a “drop down list” of selections for ease of use (access to the Infonet was initially restricted to in-hospital system computers). The PEERS reporting
system gives employees the opportunity to detail their experiences with errors and near misses. It also requests that the individual reporting the incident offer suggestions to avoid the problem in the future.

The PEERS system differentiates between near misses and errors; if an error occurs, managers are automatically notified. In the case of sentinel events (errors with harmful outcomes), managers, risk management, the attending physician, physician assistant (PA), nurse, pharmacist and any other involved parties identified in the report hold a confidential discussion about the situation. Nonpunitive root cause analysis conducted during the discussion results in system changes to address the cause of the error or event. This interdisciplinary discussion is imperative in improving quality and outcomes (Australian Council for Safety and Quality in Health Care, 2002; Bogner, 1994; Reason & Hobbs, 2003; Spath, 2000).

With PEERS implemented to allow confidential error reporting, the parent organization of the study hospital had one of the key components of a non-punitive environment now in place. By having a formal policy, employees have a written contract that supports a nonpunitive environment. The purpose of the policy is to have a quality improvement program that supports the institution’s core values, searches for root causes of errors, ensures no reprisals for error reporting, and develops a culture of open communication (Conlon, 2002).

Culture of Safety Survey

According to its safety program plan, the study hospital surveys employees and physicians on an annual basis as part of an effort to create a culture of safety. The goal of
the survey is to evaluate the ability of employees to openly discuss patient safety issues and to provide an opportunity for employees to give input about ways to improve patient safety. Questions on the survey ask about the ability to discuss patient safety in specific units within the hospital, with managers and supervisors, and asks about communications, event reporting and the hospital overall. This survey is analyzed and serves as a guide to plan for the upcoming year’s patient safety program.

**Medication Safety Program**

The study hospital’s medication survey program evaluates medication safety issues as reported in PEERS. A medication survey committee submits a work-plan and recommends action based on research and investigation of the issues. Follow-up results are reported back to the committee.

**Computerized Physician Order Entry (CPOE)**

Computerized systems for medical order entry have been in existence for several years, but demand for their use has recently blossomed. One organization calling for CPOE system implementation is The Leapfrog Group (2003).

Shortly after the Institute of Medicine released the alarming statistics regarding medical errors, the Business Roundtable (coalition of ~150 company CEOs) decided it was time for payors to demand action from the healthcare industry. In a press release the Business Roundtable (2000) introduced the “Leapfrog Initiative” to encourage large employers to “recognize and reward health plans and hospitals that make breakthrough improvements in patient safety and quality with preferential use and other market reinforcements” (p. 1).
This initiative, now called the “Leapfrog Group,” has targeted three areas to improve patient safety. CPOE tops the list. Other components call for evidence-based hospital referral (EHR) and ICU physician staffing (IPS).

According to the Leapfrog Group (2007), more than one million serious medication errors occur in hospitals every year. The group categorizes these errors as administration of the wrong drug, drug overdoses, and overlooked drug interactions and allergies. Many researchers have studied the incidence of medication errors and the effect of computerized physician order systems on reducing medication errors (Bates, Boyle, Vander Vilet, Schneider, & Leape, 1995; Bates, Culen, et al., 1995b; Bates, Teigh, et al., 1999; Bates, 2003; Classen, Pestonik, Evans, Burke, & Battles, 2005; Gawande & Bates, 2000; Leape et al., 1995; Oren, Shaffer, & Guglielma, 2003).

Bates et al. (2001) reported that CPOE not only reduces medication errors substantially but can also improve the quality and efficiency of medication use. CPOE involves a practitioner entering orders on a computer rather than handwriting orders. The benefits of CPOE as outlined by The Leapfrog Group (2007) include:

1. Prompts that warn against the possibility of drug interaction, allergy or overdose;
2. Accurate, current information that helps physicians keep up with new drugs as they are introduced into the market;
3. Drug-specific information that eliminates confusion among drug names that sound alike;
4. Improved communication between physicians and pharmacists; and
5. Reduced healthcare costs due to improved efficiencies.
Despite the fact that CPOE systems have been shown to reduce serious prescribing errors by more than fifty percent, it is estimated that less than 5% of hospitals currently utilize this type of system (iHealth Beat, January 18, 2002). Many hospitals lack the necessary funds (~$5 million) to implement such a system (iHealthBeat, May 24 2002). Another obstacle to implementation of CPOE systems is cultural because many providers resist the idea of ordering prescriptions by a computer rather than by hand (The Leapfrog Group, 2007). The Leapfrog Group (2007) has been advised by national experts in health care to implement CPOE as one of its Safety Standards. In order to meet this standard, hospitals must:

1. Assure that physicians enter hospital medication via a computer system that includes prescribing-error prevention software;

2. Demonstrate that their inpatient CPOE system can alert physicians of at least 50% of common, serious prescribing errors, using a testing protocol; and

3. Require that physicians electronically document a reason for overriding an interception prior to doing so.

The use of CPOE systems is an expensive but important factor in reducing medical errors. These systems can ultimately save money and decrease length of stay (The Leapfrog Group, 2007), despite requiring significant financial and personnel investments to adopt them. The study hospital is one of a small number of hospitals in the country that have a fully implemented CPOE since the mid 1990s. Only five hospitals in the study hospital’s state have CPOE in place, three of which belong to the study hospital’s parent organization.
**Patient Safety Rounds**

Patient safety rounds were implemented at the study hospital in 2004. According to the hospital safety management plan, these rounds rotate with a different patient care unit or department targeted each week. The multidisciplinary rounds consist of a team, which includes a representative from Executive Management, patient safety team members and staff of the unit where the rounds are conducted. Feedback from the rounds are documented and reviewed by the patient safety leadership team.

**Medical Simulation**

Historically, medical education has been accomplished by the training of students using the “see one, do one, teach one” training concept (Dunn, 2004). This adage is regarded as “no longer viable nor ethically acceptable” (Morgan & Cleave-Hogg, 2005, p. 202). Safety proponents advocate a focus on training medical students on patient safety while ensuring competence using new and cost-effective techniques (Al-Assaf, Bumpus, Carter, & Dixon, 2003; Dunn, 2004; Ziv, Ben-David, & Ziv, 2005).

Training of medical personnel was also targeted for revamping by the IOM report on medical errors, including recommendations that the medical community mimic aviation safety measures by using simulation training to reduce medical errors (Kohn et al., 2000). Simulation training has been used in aviation for more than 50 years and has proven effective in enhancing the culture of safety within the aviation industry, but it is unknown whether this approach can effect culture change within a healthcare environment (W. Hamman, personal communication, January 22, 2007). Because of the
IOM report, a recent proliferation of medical simulation technology has occurred, but research on validity and integration into medical training remains limited.

Rutherford and Hamman (2005a) proposed that the aviation experience can offer a model for enhancements of safety cultures within healthcare, but noted that the simulation process requires significant modification to achieve success in healthcare. The early years of aviation simulation focused on technical skills, but improvements in safety were not realized until simulation exercises included team skills. To evaluate and implement various simulation techniques for application in healthcare, Rutherford and Hamman developed the Michigan Center of Excellence for Simulation Research. The Center is designed to focus its efforts on applying simulation to teams of medical professionals, studying how teams work together during procedures, and characterizing behavioral skills of highly effective healthcare team members. From these efforts, they hope to develop simulation tools that address efforts uniquely within the healthcare industry to improve patient safety.

Simulation Basics

Medical simulations involve the use of mannequins to provide teaching, learning, and performance evaluation (Dunn, 2004). These mannequins have pulses, breathe, and exhibit appropriate physiologic and pharmacologic responses to drugs and inhaled gases. Current medical simulators are similar to flight simulators used in aviation in that the goal is for individuals to learn from mistakes made during simulated exercises (Gaba, Howard, Fish, Smith, & Sowb, 2001). The intended outcome is for translation of this experience to clinical situations to result in reduced errors in real life (Ziv et al., 2005).
Simulation helps protect patients from the “learning curve” associated with mastering new skills by allowing students the opportunity to practice procedures repeatedly prior to performing them on a real patient (Patow, 2005). Patow (2005) reported that students who used simulation as part of their education “were the most confident and competent they have ever had for clinical training” (p. 2).

According to Salas, Wilson, Burke, and Priest (2005), the key components of medical simulations include the following: “performance history/skill inventory, tasks/competencies, training objectives, events/exercises, measures/metrics, performance diagnostics, and feedback and debrief” (p. 363). Salas and colleagues further stated that in order to be effective, simulations must be implemented appropriately with the following guidelines:

Understand the training needs and requirements; instructional features such as performance measurement and feedback must be embedded within the simulation; craft scenarios based on guidance from the learning outcomes; create opportunities for assessing and diagnosing individual and/or team performance within the simulation; guide the learning; focus on cognitive/psychological simulation fidelity; form a mutual partnership between subject matter experts and learning experts; and ensure that the training program worked. (p. 363)

Equipment required to implement these guidelines include personnel trained to run the simulation and debriefings; a mannequin; software; monitor screens; clinical equipment, such as clinical gases; and audio and video recording equipment (Dunn, 2004; Rutherford & Hamman, 2005a).

Patow (2005) noted that simulation technology is an expensive undertaking because technology is still improving. This necessitates an ongoing investment to keep up with new equipment and software, which can seem prohibitive, but when considering the
risk to patients without simulation training, cost effectiveness becomes more readily apparent.

*Team Oriented Medical Simulation*

Interdisciplinary teams are increasingly recognized as an important part of improving patient safety. Historically, the focus of education in many health disciplines has been technical. This focus is no longer adequate, given the explosion of technical information and the need to integrate multiple professionals into one interdisciplinary team. Now viewed as imperative, teamwork skills are defined as “the cognitive, behavioural, and attitudinal actions that members need to function effectively as part of an interdependent team” (Burke, Salas, Wilson-Donnelly, & Priest, 2004). Individual disciplines need to collaborate with other disciplines in simulations to contribute to creating a culture of safety (Dunn, 2004; Hamman, 2004; Hamman et al., 2007; Kohn et al., 2000).

Teamwork across disciplines is not always an easy task but can be changed with education as Kayes, Kayes, and Kolb (2005) stated:

Teamwork is prevalent in organizations, yet it has pitfalls such as social loafing, groupthink, overdependence on a dominant leader, over-commitment to goals, and diffusion of responsibility. Such negative factors can be overcome and team effectiveness improved when teams intentionally focus on learning. (p. 330)

Patow (2005) identified development of the capacity to “work closely with others in proficiently functioning teams” as possibly the “most important contribution of simulation to medical education and patient safety” (p. 3).

A good example of this type of team training is the Anesthesia Crisis Resource Management training (ACRM). This was developed based on principles of Crew
Resource Management (CRM) training in aviation with the goal of "training single-discipline crews to work in teams" (Gaba et al., 2001). Anesthesiology has the most experience with medical simulation and ACRM has now been applied to other areas of medicine including the emergency rooms, critical care units, and cardiac arrest teams.

ACRM involves each discipline working in separate teams, but the curriculum provides for combined team training where all team members are involved in the simulations (Gaba et al., 2001). The combined team training allows all team members to hear and discuss views from others' disciplines of the same situation, along with playing the role of another team member's discipline. Both methods are complementary. Ideally team members would participate in both types of training (Gaba et al., 2001). This training is consistent with the IOM (Kohn et al., 2000) recommendations, which stated that organizations should

establish interdisciplinary team training programs—including the use of simulation for trainees and experienced practitioners for personnel in areas such as the emergency department, intensive care unit, and operating room; and incorporating proven methods of managing work in teams as exemplified in aviation. (p. 157)

Gaba et al. (2001) indicated that this type of training should be mandated in medicine, as it is in the airline industry. Further, health care systems must bear the cost of training personnel despite other financial demands.

*In-situ® Simulation and Debriefing*

According to Hamman et al. (2007), the aviation CRM model cannot be translated directly to medicine, but should be used as a guide for development of the medical team
model. Aviation simulation techniques applied in the medical model can help determine
the model of team/crew resource management that should be implemented in healthcare.

*In-situ*® simulation is based on the philosophy of the aviation industry’s
Advanced Qualification Program (AQP) (Hamman et al., 2007). It is

1. Systematically-developed based on a detailed needs analysis;
2. Proficiency based;
3. Mission oriented;
4. Focused on team, rather than individual performance;
5. Focused on simultaneously training both teamwork and technical skills;
6. Presented using scenarios that are based on real-world events; and
7. Validated using both quantitative and qualitative methods. (p. 2)

The *In-situ*® simulation and debriefing technique allows participants to respond
in their particular environment and therefore respond to system issues within their
organization. Event sets, as described by Hamman et al. (2007), bring out issues such as
equipment failure, supply availability, and responsiveness of code personnel unique to
any given institution. These event sets comprise an event trigger, distractors, and
supporting events. As described by Hamman (2004):

The event trigger is the condition or conditions under which the event is fully
activated. The distractors are conditions inserted within the event set timeframe
that are designed to divert the team’s attention from other events that are
occurring or about to occur. Supporting events are other events taking place
within the event set that are designed to further the training objectives. (p. i76)

The *In-situ*® simulation project is designed such that the
physical simulator in the health care domain can only present 20% of the
challenges that are faced by the team. The remaining 80% of the challenges to the
team in the event sets come from the environment that the simulation takes place
in. (Hamman, 2007, pp. 2–3)
Debriefing

“To learn from their experience, teams must create a conversational space where members can reflect on and talk about their experience together” (Kayes et al., 2005). This is accomplished during the debriefing part of simulation, which is the most important part of simulation training (Dunn, 2004; Graber, 2003; Sinz, 2005; Ziv, Wolpe, Small, & Glick, 2003). The debriefing facilitator plays a key role in this training and therefore emphasis should be placed on the quality of the facilitator and not the quantity of facilitators (Dunn, 2004). The most educated or intelligent person is not necessarily the best facilitator; a person with “a warm, open, accepting personality with an extroverted flair may work best since the student deficiencies must be approached with compassion mixed with creative ability” (Dunn, 2004, p. 78). According to Dunn (2004) instructors should first master the CRM training prior to becoming a debriefing facilitator.

Rutherford and Hamman (personal communication, May 24, 2006) depicted the basics of facilitating simulation debriefings as:

1. Keeping the discussion team-centered;
2. Encouraging team members to participate actively and do most of the talking;
3. Adapting the level of facilitation to the capabilities of each healthcare participant;
4. Balancing the dual role as instructor and facilitator;
5. Reinforcing good team performance following team analysis;
6. Showing by your attitude that you are interested in what the team members have to say;
7. Avoiding lectures and long speeches;
8. Not interrupting or leaving a topic while the team still has something to say;
9. Using the highest level of facilitation possible to maximize team self-
discovery; and
10. Adapting the level of facilitation to accommodate varying team needs throughout the session.

Dunn (2004) added that successful debriefing relies on capturing “the essential components of reinforcing, self-critiquing, and discussion among team members” (p. 80).

Facilitators should start the debriefing with a short introduction, during which they clarify their role, provide rationale for why the debriefing should be team-centered, and tell the participants how long the debriefing will last. The facilitators should then explain the format of the debriefing and develop an agenda (Rutherford & Hamman, personal communication, May 24, 2006).

Another important aspect of the debriefing session is the videotape analysis, which encourages participants to analyze the simulation. Rutherford and Hamman (personal communication, May 24, 2006) stated that the guidelines for facilitating the use of video include:

1. Indexing of important events during the simulation;
2. Not showing video segments you don’t intend to discuss;
3. Efficient use of the video equipment;
4. Introduction of each video segment, reminding the team that they will analyze what they see; and
5. Pausing the video for comments and to discuss important aspects of team performance.
They also stated that debriefing sessions should not be shortened for teams that are already functioning well. At the end of the debriefing the facilitator should concisely summarize the session by reiterating important issues, bearing in mind the ultimate goal of understanding team performance within a specific health care system and remembering to focus on “what is right” rather than “who is right” (Rutherford & Hamman, personal communication, May 24, 2006).

Culture and Attitude Domains

One important concept regarding medical simulation debriefings and team leadership in general can be summed up in the following statement: “It is wise to spend at least as much time ‘managing’ attitudes as it is managing core competencies. If the team is mentally on board, the quality of work improves because each individual’s desire to learn is increased” (Durand, 2006, p. B4).

The survey used in the current study examined six attitudinal domains that were developed out of focus groups, literature review, and round-table discussions with industry experts (Sexton, Thomas, & Grillo, 2003). These six domains are important in creating a culture of safety. A primary goal of the current study was to determine whether the proposed simulation project could influence any of these domains, but particularly teamwork. The simulation and debriefing intervention creates an opportunity for dialogue that normally would not occur in most organizations and could help shape perceptions of individual team members regarding each of these domains (teamwork climate, safety climate, job satisfaction, stress recognition, perception of management, and working conditions).
Teamwork Climate

Teamwork as defined by Sexton et al. (2003) is the “perceived quality of collaboration between personnel.” The IOM stated that in order to improve safety, interdisciplinary teams must learn how to work in teams and suggested use of the aviation industry’s CRM training to emphasize better communication across disciplines (Kohn et al., 2000). Lencioni (2002) stated that

teams, because they are made up of imperfect human beings, are inherently dysfunctional. But that is not to say that teamwork is doomed. Far from it. In fact, building a strong team is both possible and remarkably simple. But it is painfully difficult. (p. vii)

Baker, Day, and Salas (2006) stated that, in medicine, knowledge and skill of a task are not enough and that “teamwork depends on each team member being able to anticipate the needs of others; adjust to each other’s actions, and have a shared understanding of how a procedure should happen” (p. 1579). The simulation and debriefing exercises in this study examined how interdisciplinary cardiac and vascular surgery personnel worked together as coordinated teams. Dunn (2004) stated that some interdisciplinary medical teams are dysfunctional because

social barriers such as hierarchy, modesty, and hubris sometimes stand in the way of explicit communication about team structure and process. Poor communication habits often lead to incomplete understanding or splintered goals among team members. Lack of respect or trust between specialties at times leads to poor teamwork or even conflict. (p. 43)

According to Dunn (2004) the objectives for team members are threefold:

1. Appreciate the advantages and disadvantages of two team models: the high-performance team and the work group.
2. Learn how to attain the high-performance team models.
3. Learn the team skills necessary to achieve role clarity, communicate effectively, optimize support, utilize resources effectively, and maintain a global view. (p. 43)

Sexton (2004) recommended that team members also “use language to foster team perspective” including first person plural pronouns such as “we/let’s” (p. 20). Studies have shown that 73% of aviation accidents occur on the first day of crew pairings and 44% are on the first flight of a new crew (National Transportation Safety Board, 1994; Sexton, 2004). This is thought to be a result of the crew’s limited familiarity with each other and is reflected in the crew rarely referring to themselves at that point in the first person plural, such as we, our, or us (Sexton, 2004).

Dunn (2004) concluded that a simulated crisis can improve teamwork, and interdisciplinary teams can learn to improve teamwork with the use of video reviews during the debriefing exercises. The SAQ was designed to determine participant attitudes regarding teamwork defined as perceived quality of collaboration between personnel by posing the following statements for consideration:

1. It is easy for personnel in this ICU/OR to ask questions when there is something that they do not understand.
2. I have the support I need from other personnel to care for patients.
3. Nurse input is well received in this ICU/OR.
4. In this ICU/OR, it’s difficult to speak up if I perceive a problem with patient care.
5. Disagreements in the ICU/OR are appropriately resolved (i.e. not who is right, but what is best for the patient).
6. The physicians and nurses here work as a well coordinated team.

Safety Climate

Safety climate is defined by Makary, Sexton, Freischlag, Millman, et al. (2006) as the “perceptions of a strong and proactive organizational commitment to safety” (p. 629). A commitment to a safety climate is part of an organization’s culture. Organizational culture is defined by Uttal (1983) as “shared values (what is important) and beliefs (how things work) that interact with an organization’s structures and control systems to produce behavioural norms (the way we do things around here)” (p. 66). National cultures are a result of shared values, whereas organizational cultures are a result of shared practices (Reason, 1997). Helmreich and Merritt (2001) stated that “the elements that form professional, national and organizational cultures can come together to define a safety culture or can create an unsafe operating environment” (p.133).

Reason (1997) stated that the “ideal safety culture is the engine that continues to propel the system towards the goal of maximum safety health, regardless of the leadership’s personality or current commercial concerns” (p. 195). Reason further reported that the other components of a safety culture include the ability to move forward despite many factors that threaten to distract from the course; collection of data about errors and near misses with a reporting system; and existence of a reporting culture that does not encourage a “blame and shame” mentality; organizations must also be flexible and have an environment that encourages learning from their reporting system.
As outlined above, the current study hospital has many tools in place, such as an anonymous error reporting system (PEERS) and a nonpunitive policy. These are in line with Reason’s (1997) stipulations of what an ideal safety culture should entail.

In this study, use of the SAQ determined participant attitudes regarding the hospital’s safety climate defined as perceptions of a strong and proactive organizational commitment to patient safety by posing the following statements for participants to consider:

1. The culture in this ICU/OR makes it easy to learn from errors of others.
2. Medical errors are handled appropriately in this ICU/OR.
3. I know the proper channels to direct questions regarding patient safety.
4. I am encouraged by my colleagues to report any patient safety concerns.
5. I receive appropriate feedback about my performance.
6. I would feel safe being treated here as a patient.
7. In this ICU/OR it is difficult to discuss errors.

**Job Satisfaction**

Job satisfaction is reflected as “positivity about the work experience” (Sexton et al., 2006, p. 3). In a study of mine workers in India, negativity affectivity and job dissatisfaction were two factors that contributed to accidents and injuries in the mines studied (Paul, Maiti, Dasgupta, & Forjuoh, 2005). Physicians and pilots share a common bond in their job satisfaction ratings. Helmreich and Merritt’s (2001) data showed that pilots and physicians in aviation and medicine have a high regard for their work. Their
report of a survey in 19 countries showed that pilots tended to be satisfied with their jobs, but that the most satisfied physicians were only equal to the lowest group of pilots.

A survey of over 700 hospitals in five countries reported that more than 40% of hospital nurses are dissatisfied with their jobs and concluded that this could be resolved with managerial intervention to yield the outcome of preservation of patient safety (Aiken et al., 2001). Holden (2006) reported that nursing job dissatisfaction and patient safety would improve if the culture of nursing organization were transformed into a learning organization. This type of model would mean that "every member of the nursing organization would be encouraged to reach his or her greatest potential, the welfare of the team would become paramount, and a shared vision of where the organization needs to go would emerge, thus maximizing productivity, safety, and job satisfaction for all healthcare team members" (p. 34).

The job satisfaction of the study hospital’s employees was determined in this study by eliciting responses to the following statements on the SAQ:

1. This hospital is a good place to work.
2. I am proud to work at this hospital.
3. Working in this hospital is like being part of a large family.
4. Morale in the ICU/OR is high.
5. I like my job.

Stress Recognition

Stress recognition, as defined by Sexton et al. (2003), is the “acknowledgment of how performance is influenced by stressors.” Humans make more errors when stressed;
thus, believing that one is able to function effectively when stressed is a defense mechanism that is inconsistent with safety. Denial of this effect on performance relates to organizational practices; pilots are more realistic than surgeons about the effect of stress on safety perhaps because pilots have been trained through crew resource management to recognize the effects of stress and that their organization limits flight time because of the known effects of stress on safety (Helmreich & Merritt, 2001). One study of Alaska carriers, for example, revealed a higher accident rate in pilots who were less likely to consider fatigue a factor when scheduling flights (Conway, Mode, Berman, Martin, & Hill, 2005). A balance on the effects of stress recognition is imperative because underestimation leads to accidents, whereas overestimation may lead to dysfunction due to the uneconomical waste of resources (Health & Safety Commission, 1993).

Grogan et al. (2004) studied the impact of airline crew resource management training on the attitudes of hospital employees at Vanderbilt University. The training curriculum included lectures on crew resource management concepts and utilized simulated scenarios for role playing. The training program improved “attitudes toward fatigue management, team building, communication, recognizing adverse events, team decision making, and performance feedback” (Grogan et al., 2004, p. 843).

The current study examined the domain of stress recognition by asking participants to react to the following statements:

1. When my workload becomes excessive, my performance is impaired.
2. I am more likely to make errors in hostile or tense situations.
3. Fatigue impairs my performance during emergency situations.
4. I am less effective at work when fatigued.
The perception of management is the “approval of managerial action” (Sexton et al., 2003). Employees’ attitudes regarding safety at work have been linked to the commitment of management to safety (Garcia, Boix, & Canosa, 2004). Helmreich and Merritt (2001) reported that in initial surveys of pilots, few pilots (15 out of 1,200 in one study) indicated that management was doing a good job and even fewer reported that they trusted them. This might be an indicator of professional culture, but more studies need to investigate perception of management across disciplines in both aviation and medicine (Helmreich & Merritt, 2001). Helmreich and Merritt also hypothesized that this could be a factor of organizational culture because “subjective perceptions rather than ‘hard facts’ shape an organizational culture and, by studying the perceptions of our professional groups, we gain an understanding of how these groups perform their work in the full organizational context” (p. 113).

Perceptions of management among hospital employees at the participating institution were measured in the current study with the following statements on the SAQ:

1. Hospital management does not knowingly compromise the safety of patients.
2. Hospital administration supports my daily efforts.
3. I am provided with adequate, timely information about events in the hospital that might affect my work.
4. The levels of staffing in this clinical area are sufficient to handle the number of patients.
Working Conditions

Working conditions, as defined by Sexton et al. (2003), are characterized as the “perceived quality of the ICU work environment and logistical support (staffing, equipment, etc.).” Working conditions are an important part of the safety culture of organizations. Maritime accidents in Japan are five times higher than in other industries, and many of the accidents are a result of the ergonomic conditions of the working environment (Hisamune, Amagai, Kimura, & Kishida, 2006). Working conditions were also a factor in the study of Alaska aviation accidents from 1990–2002. Pilots involved in accidents differed from other pilots in experience and in working conditions that involved working longer hours (Conway et al., 2005).

Working conditions have been evaluated at the study hospital on a yearly basis through a Gallup survey. This survey allows employees to give feedback on management and institutional issues with results compiled by departments and job classifications. The statements corresponding to working conditions on the SAQ used in this study are similar in nature to those on the Gallup survey:

1. All the necessary information for diagnosis and therapeutic decisions is routinely available to me.
2. This hospital constructively deals with problem physicians and employees.
3. Trainees in my discipline are adequately supervised.
4. This hospital does a good job of training new personnel.
Surgical Teams and Job Classifications

The operating room (OR) and the surgical intensive care unit (SICU) are two examples of "high-risk" areas in hospitals (Kaissi et al., 2003). Interdisciplinary teamwork in these areas is crucial to reducing error, as indicated by data that suggest that 70–80% of medical errors are related to communication (W. Hamman, personal communication, March 20, 2007; Runy, 2007; Williamson, Webb, Sellen, Runciman, & Van der Walt, 1993). Cardiac and vascular surgery teams comprise multiple disciplines that come together to care for patients in the OR and SICU. The personnel within the various job classifications on these teams have varied backgrounds and education. Interdisciplinary communication and teamwork training are not traditionally included in the formal education of team members.

Cardiac Surgery Teams

The cardiac surgery team at the study hospital consists of cardiac surgeons, cardiac physician assistants (PA), cardiac PA residents, clinical nurse specialists (CNS), a nurse practitioner (NP), a dedicated operative nursing and certified surgical technologist team, and perfusionists. Each of these roles is explained below.

Effective team performance in cardiac surgery is widely recognized as crucial because patients who undergo cardiac surgery typically have no other option, and because outcomes among these patients range from complete recovery to death (Friedman & Bernell, 2006). Team composition is important in determining team performance, but this varies from institution to institution. Historically, cardiac surgeons were regarded as prima donnas and were able to declare who would be on a given team.
This is no longer the case in most institutions, as reflected by a cardiac surgeon comment reported in a qualitative study by Friedman and Bernell (2006):

In the old days, when cardiac surgeons used to rule the earth, we were the prima donnas of many hospitals. But today, we are lucky to get good people. And I've been blessed with the people I have, but it's not like I get to select—hospitals will provide you with whoever they have. (p. 226)

Another cardiac surgeon’s comments from the same report concurred:

Unless there is a particularly difficult case or a case where the patient may know somebody who knows an anesthesiologist, or where we particularly want a specific physician because of their expertise, we will request that, but this doesn’t happen very often. (p. 226)

Cardiac surgeons at the study hospital occasionally request specific team members for particularly difficult cases, but this behavior may be met with resistance by some of the teams’ members. In some cases, team members express openness about the possibility of being “bumped” from a given case, indicating that they understand that if they do get “bumped” they should not take it personally. Some members of the team have indicated a preference for the surgeon to communicate openly about team composition preferences regarding the more complex cases.

**Vascular Surgery Teams**

The vascular surgery team at the study hospital consists of vascular surgeons, general surgery residents, a nurse practitioner, and a dedicated operative nursing and certified surgical technologist team. The vascular team differs from the cardiothoracic surgery team by having general surgery residents instead of PAs. However, within the last 2 years the cardiothoracic surgery PAs have performed endoscopic vein harvest (EVH) for the vascular surgery team when needed. This crossover occurred because EVH
is a fairly new, highly technical skill that has a steep learning curve. Most surgeons have not been trained in the technique, but it is gradually becoming a standard of care. The cardiothoracic PAs at the study hospital were very experienced in EVH technique at the point when the vascular surgeons decided to incorporate it into their practice. This resulted in a natural evolution for the PAs to provide this service for the vascular surgery team.

The vascular surgeons rarely have control over team composition in the operating room. They occasionally request the assistance of one of their staff surgical colleagues on a difficult case instead of relying on a general surgery resident or nurse first assistant.

Interdisciplinary Team Composition

Physicians. The physicians involved in this study include cardiac and vascular surgeons, along with general surgery residents and anesthesiologists. The anesthesiologists are part of a dedicated team for all cardiac surgery cases, but these anesthesiologists also cover the rest of the operating room when needed. The study physicians have extensive training in their given field (3 to 8 years of residency and specialty training following medical school) and also have been involved in the quality and safety programs within the institution. Attending staff physicians are ultimately responsible for ensuring that any given patient remains safe and receives quality care. In theory, as a result of patient safety and quality improvement projects within the hospital, physicians encourage open communication among all team members.

Physicians have been shown to have greater perceptions of communication with nurses than nurses do with physicians (Miller, 2001; Reader et al., 2007; Sexton et al.,
Miller (2001) suggested that physicians are “more comfortable with their communication skills in general and with their problem-solving skills in particular than nurses were with the nurses’ skills” (p. 349). Miller (2001) indicated that “physicians had less fear of repercussion or misunderstanding when speaking with nurses than nurses did when speaking with physicians” (p. 348). This is important to determine in a particular unit especially when a low level of interdisciplinary collaboration by nurses in a given unit is a significant predictor of negative outcomes (Baggs, Ryan, Phelps, Richeson, & Johnson, 1992; Miller, 2001).

*Physician extenders.* At the study hospital, there are physician extenders in the following job classifications: physician assistants, nurse anesthetists, nurse practitioners, and clinical nurse specialists. Physician assistants are trained in the medical model, whereas nurses are trained in the nursing model. It is important to review the beginning of the PA profession in order to understand the relationship between nurses and PAs and the history of their interdisciplinary collaboration. Even though many PAs and nurses are unfamiliar with this history, the antagonistic relationships that often prevail are deeply rooted and have been documented for over 40 years (Holt, 1998; Hudson, 1961; Ingles, 1968).

Hudson (1961) published an article entitled *Expansion of Medical Professional Services with Nonprofessional Personnel.* This article is thought to mark the conception of the role of the physician assistant (PA). Hudson referred to these assistants as *externes* to contrast with the medical term *intern.* He envisioned that the employer physician would extend his or her “moral and legal responsibility” (p. 840) to cover the assistant’s extension of the physician’s arms, legs, and mind. It would be left to the physicians to
judge how much responsibility their assistants safely could assume. Forty-six years later
the relationship between a PA and his or her supervising physician remains a critical
observed that the role of the *externe* could potentially be filled by nurses, but he believed
that “nursing would not find the proposal of a medicine-nurse hybrid consistent with their
present goals for nurse education” (p. 841).

The first formal training program for this *externe* role, which was renamed the
physician assistant, occurred 4 years after Hudson published his article on the *externe*
role. The first training program was initiated by Dr. Eugene Stead, Jr., at Duke
University. Stead developed an early version of this role in 1957, working with the
supervisor of medical and surgical nursing at Duke Nursing School, Thelma Ingles (Holt,
1998). Ingles was preparing to take a sabbatical year and Stead had suggested that she
take it in the medical school. She took his suggestion and became the first nurse at Duke
to do so. She functioned as a medical student in a clinical clerkship and used her
experience for giving seminars on specific topics in clinical care. After her sabbatical,
Ingles returned to the nursing school to create a Master of Science in Nursing program
modeled on her experience with Stead (Holt, 1998).

However, the National League of Nurses (NLN), which was the accrediting body
for training programs at the time, did not approve of Duke’s clinical nursing
specialization program and withheld accreditation. The NLN criticized the use of
physicians as instructors.

Duke made one more attempt to include nurses in their vision of this new role in
1961 when senior nursing students “wanted more authority and responsibility than they
saw possible in the nursing role” (Holt, 1998, p. 255) so they approached their supervisor, Ruby Wilson, with the statement: “We don’t want to become like the rest” (Holt, 1998, p. 255). Based on this assertion, Thelma Ingles came up with the idea that this group of graduating seniors might become the staff of a particular unit requiring advanced skills that were different from those taught in traditional nursing programs. They set up the Hanes Ward, which was directed by Wilson. It quickly achieved status as the best nursing care unit in North Carolina. It did not survive long, however, because the director of nursing never fully supported the concept. Additionally, hospital administration failed to uphold the agreement that the nursing program would keep profits earned from the unit (Holt, 1998).

History shows that Stead persisted in his determination to develop the PA concept and in 1964 was ready to proceed with establishing the PA training program. Because of the failure of Thelma Ingles’ master’s program to accord the support it needed to survive, he decided to look elsewhere than the nursing profession to develop this new kind of professional that physicians were seeking. It is reported that Stead’s motto at this time was a quote by George Washington: “We are not to look back unless it is to derive useful lessons from past errors, and for the purpose of profiting by dear bought experience” (Howard, 1968, p. 1). At this point, Stead began focusing on recruiting male students; the program was established within the Department of Medicine with ties neither to hospital administration nor to Duke Nursing School (Holt, 1998).

Hudson (1961) also had been concerned about reactions from the American Nurses’ Association (ANA) to his concept of the *externe* role. These concerns became realized when Stead implemented the PA program. Five years after the program started,
ANA president Dr. Hildegard Peplau accused the American Medical Association (AMA) of creating the new role to sabotage the ANA’s efforts to define nurses as independent associates of physicians, rather than task-oriented “physician’s handmaidens” (Peplau, 1970, p. 252).

In enthusiastic support of the PA concept, the vice-president of the AMA, Ernest B. Howard, proposed during the 1969–70 Lowell Lecture at Boston University’s School of Medicine that 100,000 nurses be recruited for PA training. ANA President Dorothy Cornelius deplored the unilateral action of the AMA and warned that “it is not the prerogative of one profession to speak for another” (Peplau, 1970, p. 691). The nurses, having defined themselves as independent professionals, were skeptical of the intentions of the physicians and did not want to give up their independence by becoming providers of medical care under the direction of physicians (Peplau, 1970). They questioned whether a subordinate employee, such as a PA, could have a higher status than an independent professional, such as a nurse (Holt, 1998).

The nursing profession also took issue with the PA salary structure, fearing that the salary differential was really a ploy by the AMA to obliterate the ANA by luring nurses out of their own profession. In a plea to state-level nursing associations, ANA president Dr. Hildegard Peplau stated her hope that nurses would reject the loss of their identity as nurses. She indicated that it was time to recognize and call attention to the traditional and continuing complementarity of nursing and medicine—two important, autonomous but interrelated services, both meeting public needs (Holt, 1998).

Throughout the 1960s, the ANA did not describe physician assistants as a threat to the nursing profession, and in fact Peplau many times exclaimed, “Nursing care is not
medical care” (Peplau, 1970). Thelma Ingles (1968), who had attempted to participate in
developing the new role as one that could include nurses, observed distinctions in the
duties of PAs, by indicating that there was

nothing which I could interpret as interfering with the nurse’s role. As a matter of
fact, I often saw the assistants doing procedures which nurses have rebelled
against for many years, tasks which have been seen as medical rather than
nursing: taking histories, drawing blood, collecting specimens for gastric analysis,
and doing basal metabolism rates, electrocardiography, and skin-testing for
allergies. (p. 1060)

Training for PAs typically involves 2 full-time years of postbaccalaureate
education. Of the 136 PA training programs listed on the American Academy of
Physician Assistants’ website, 113 offer a master’s degree or master’s degree option. The
remaining programs confer bachelor’s or associate degrees, with a handful of nondegree
programs offering a certificate of completion. Following the 26-month program, PAs take
a generalist national certification exam which must be maintained by the same continuing
medical education requirements as physicians. The exam was originally developed and
administered by the United States National Board of Medical Examiners, but now has its
own national commission. PAs must certify as generalists every 6 years by examination,
regardless of their specialty.

NP training is typically completed in a 2-year time frame but does not require full
time enrollment. A master’s degree is typically awarded at completion and the certifying
exam is based on a specific area of study. There currently is no re-certifying exam
required; continuing medical education requirements vary by state.

The study hospital, consistent with history in the antagonistic coexistence
between PAs and nurses, had a tradition of competition between PAs and nurses,
specifically, advanced practice nurses such as NPs and clinical nurse specialists. This
rocky history notwithstanding, it became the specific intention of a new group of cardiothoracic PAs (of which the author was a member) to change this counterproductive atmosphere. The goal was to have the PAs and advanced practice nurses portrayed to staff, and to each other, as partners rather than competitors.

Cardiac PAs at the study hospital are responsible for covering the operating room (assisting on all cases), and providing in-house medical management coverage for all cardiac surgery patients 24 hours a day and 7 days a week. The study hospital has sponsored a Cardiothoracic Surgery PA Residency since 2000 for 1-year postgraduate training of PAs in cardiac and thoracic surgery. The staff PAs are largely responsible for the training of these residents throughout the year. There are currently 13 staff and 2 Cardiothoracic Surgery PA residents who cover the service.

Clinical nurse specialists (CNS) provide preoperative and late phase postoperative care of the cardiothoracic and vascular surgery patients. The CNS designation is awarded by completion of a master’s degree. A formal, centrally administered certification exam has recently been introduced. Continuing medical education requirements vary by state. CNS make daily work rounds, writing progress notes and orders on patients, and perform minor procedures, such as removal of chest tubes and temporary pacing wires. The CNS team is specifically responsible for education of all patients. They coordinate all factors pertaining to discharging the patient, including arranging for visiting nurses, placing patients in extended care facilities, and arranging follow-up appointments with the surgeons’ office. They collect data for state and national databases for cardiac and thoracic surgery. A recently assigned function now also requires the CNS team to serve
as case managers, obtaining proper authorization from insurance companies for the hospital.

Two NPs are members of the teams in the study hospital, one cardiac and one vascular, both with critical care specialization. The cardiac surgery NP monitors and manages the cardiothoracic patients in the intensive care unit during the day and does history and physicals and other tasks unable to be performed by the CNSs. The vascular surgery NP provides pre- and postoperative care for all of the vascular surgery patients as well.

Most physician extender categories work with a variety of physician specialists, but the certified registered nurse anesthetist (CRNA) trains and practices only in anesthesia. Their 2-year full-time training requires extensive supplemental clinical hours. CRNAs must pass a certifying exam and participate in a recertifying process every 2 years consisting of 850 clinical hours and 40 hours of continuing medical education credits (American Association of Nurse Anesthetists, 2007). The two CRNAs who participated in the study are involved in the vascular surgery service but do not help on the cardiac team as primary providers. However, they are responsible for responding to in-house emergencies, such as cardiac arrest codes 24 hours a day.

_Nurses._ The nurses in this study included operating room (OR) nurses, surgical intensive care (SICU) nurses, and progressive care (PCU) nurses. The surgical intensive care nurses (SICU) and progressive care unit (PCU) nurses provide care for both the cardiac surgery and vascular surgery patients and are included in the ICU SAQ. The SICU and PCU nurses function at varying levels of experience but generally have similar educational backgrounds. The majority are bachelor’s prepared (BSN) as opposed to
holding 2-year degrees as registered nurses (RN). These nurses are responsible for the minute-by-minute bedside care of the patients, monitoring all vital signs, administering medication, assessing complaints, and communicating with the surgeon and providers regarding care plans. They interact extensively with the patient, the patient’s family, and with all providers caring for the patient.

Nurses in the OR on both the cardiac and vascular surgery teams function at differing levels of experience and hold varying educational degrees; however, most are bachelor’s prepared. The operating room has a highly focused, team-specific assignment process, which allows these nurses to integrate fully into the team. The OR nurses monitor the patient from a pre-op holding area to post-op recovery, assessing needs and communicating with the surgical team and the patient’s family. They have dual roles of “circulating” to oversee the progress of the surgery, or “scrubbing” to hand instruments to and assist the surgeon. In both roles, they are responsible for maintaining sterility and decorum, and for communicating with various contacts, such as family and other providers who may phone into the room.

**Blood bank medical technologists.** Medical technologists have a bachelor’s degree and usually major in medical technology or one of the life sciences (United States Department of Labor, 2006a). The blood bank medical technologists in this study are responsible for the preparation and distribution of all blood products within the study hospital. This involves handling specimens and analyzing the tests necessary to match and distribute the blood and blood products for safe transfusion. The blood bank is located on the first floor of the institution and normally only has phone contact with the providers ordering the blood products.
Certified surgical technologists. In this study the surgical technologists (CST) were members of the vascular surgery team. CSTs assist in the operating room under the supervision of a circulating nurse and the surgeon. The role of the CST in the operating room is to prepare instruments prior to the operation; during the operation they are responsible for supplying the surgeon and the rest of the operating room team with sterile instruments and supplies. CSTs do not function in the circulating role because they have limited knowledge of the patient’s physiologic process.

The training of CSTs occurs in community colleges, vocational schools, hospitals or universities. These formal programs are usually 9 to 24 months in length (United States Department of Labor, 2006b) and offer a certificate of completion or associate’s degree.

Patient care technician. Only one patient care technician fully participated in this study. This individual works on the progressive care unit, taking care of both cardiothoracic and vascular surgery patients. Patient care technicians attend courses at community colleges and receive a certificate upon completion. They are trained to record vital signs, collect samples for laboratory testing, and record patients’ conditions and treatments. They also help patients in ambulation, bathing, dressing, and personal hygiene. Some patient care technicians are also trained to place intravenous catheters and urinary catheters.

Perfusionist. Perfusionists operate the heart-lung machine for patients who are undergoing open heart surgery. They also operate special blood conserving devices known as cell-savers for both the cardiac and vascular surgery services and other surgical
specialties as well. They monitor the patient’s vital signs during cardiac surgery and keep the surgeon and anesthesiologist aware of the patient’s condition, along with administering any blood or medication that may be necessary while the patient is supported by the heart-lung machine. They collaborate closely with the entire operating room surgical team. The majority of perfusionists at the study hospital were respiratory therapists prior to becoming perfusionists.

Twenty-one programs in the United States train about 140 perfusionists per year. The programs may award a certificate, a bachelor’s degree, or master’s degree, and they vary in length and educational prerequisites. The American Board of Cardiovascular Perfusion awards perfusionist certification and requires yearly continuing medical education (Mariani, 2002–2003).

**Radiographers.** Radiographers perform radiographic testing for diagnostic evaluation of the cardiothoracic and vascular surgery patients. They often have only brief interaction with the providers and nurses when taking portable x-rays, which can be vital in the care of these patients. Radiography programs are 1-4 years in length and these programs award a certificate, bachelor’s degree or master’s degree.

**Respiratory therapists.** Respiratory therapists (now increasingly known as “respiratory care practitioners”) are responsible for assisting in the pulmonary care of cardiothoracic and vascular surgery patients. This involves managing the ventilator (mechanical breathing machine), providing breathing treatments, and recommending changes in respiratory care based on a patient’s condition and response treatment. During cardiac arrest codes and similar emergencies, they manage the patient’s airway until a
nurse anesthetist arrives to intubate the patient for mechanical ventilation. They also setup and operate intra-aortic balloon pumps (devices which assist the failing heart), after insertion by the surgeon or PA. Respiratory therapists at the study hospital are also responsible for obtaining electrocardiograms during the evening.

There are two levels of respiratory therapists including the certified respiratory therapist and the registered respiratory therapist (American Association for Respiratory Care, 2002). Both require an associate or bachelor’s degree and certification exam for a designation of Certified Respiratory Therapist (CRT). To become a Registered Respiratory Therapist (RRT) the therapist must take two additional exams (American Association for Respiratory Care, 2002).

Unit clerk. Unit clerks are responsible for receptionist and clerical duties in the SICU and PCU. They answer telephones, compile and organize admitting and discharge papers for patients, along with keeping track of new information that needs to be compiled in a patient’s chart. In emergency situations, unit clerks play a vital role in communicating patient and provider needs to satellite units, such as pharmacy, radiology, and respiratory care. Skills necessary for this role include keyboarding skills, knowledge of medical terminology, filing skills, and knowledge of hospital admitting and registration policies and procedures (Employment Development Department, 1998).

Summary

The IOM report (Kohn et al., 2000) heightened awareness by the public and also by medical institutions regarding medical errors. According to the report, the person approach to human error, which blames individuals, impedes healthcare safety. As a
result, the IOM recommended a change in institutional culture to one that views errors to be the result of system failures, rather than practitioner failures. The IOM has encouraged medicine to mimic aviation safety measures by using simulation techniques to reduce errors. In theory, this type of exercise should also change the way interdisciplinary medical teams interact, but there is little research to support the efficacy of such an approach in medicine for changing practice, or even for changing attitudes among interdisciplinary team members that might lead to changes in practice.

The study hospital had implemented many safety tools to improve its safety culture prior to the IOM report and had been considering the addition of simulation programs to continue updating its various safety programs. For this study, the hospital partnered with Western Michigan University’s School of Aviation to examine the effectiveness of an In-situ® simulation process, which allows participants to respond within their particular environment thereby bringing system issues to the forefront.

The interdisciplinary design of the study is unique, as similar designs were not found in the literature. Because interdisciplinary communication and teamwork training are not traditionally included in the formal education of medical professionals, understanding the educational background and role of each team member might facilitate team communication and function.
CHAPTER III

METHODS

This pre-post partially randomized controlled, quasi-experimental study examined the effects of a medical simulation and debriefing exercise on the attitudes associated with a culture of safety in a hospital. Participants were randomly assigned either to the experimental or control condition, which determined whether or not they completed a simulation exercise and debriefing session. The OR and ICU Safety Attitudes Questionnaires (Sexton et al., 2000) were used to measure all participants’ perceptions of their hospital’s culture of safety at three points—the week prior to the time when the simulations were scheduled, approximately 1 week after the simulations, and 6 weeks later.

The main objective of the study was to determine whether the attitudes of surgical team members as measured by the OR and ICU Safety Attitudes Questionnaires change after being exposed to a medical simulation and debriefing exercise as compared with control group members. A second objective was to learn whether any differences between the experimental and control groups could be detected after more than 1 month (6 weeks) following the investigation.

The results were analyzed to answer questions about whether medical simulation and debriefing would: (a) result in a change in attitudes to increase the number of positive responses regarding teamwork climate, safety climate, job satisfaction, stress recognition, perception of management, and working conditions among cardiac and
vascular surgery employees as measured by the OR SAQ; and (b) result in a change in attitudes to increase the number of positive responses regarding teamwork climate, safety climate, job satisfaction, stress recognition, perception of management, and working conditions among cardiac and vascular surgery employees as measured by the ICU SAQ. These assessments were made using various groups as the unit of analysis and were based on Sexton’s criteria (described in this chapter).

Other statistical tests were done using the individual participant as the unit of analysis. These included analysis of variance (ANOVA), general estimating equations (GEE), and paired t tests of those with low teamwork climate baseline scores and comparing specific questions at the item level.

A study flowchart was created (Figure 1) to guide the research team through the study process. The flowchart was constructed to be consistent with the Consolidated Standards of Reporting Trials (CONSORT). Designed by journal editors, trialists and methodologists to improve the quality of randomized controlled trials, the CONSORT statement includes a checklist of essential items to include when reporting randomized trials and a diagram for documenting the flow of participants through a trial (Altman et al., 2001; Moher, Schulz & Altman, 2001). The sections of this chapter include descriptions of the following: preliminary steps taken within the institution, participants and their roles, procedure for randomization and simulation, the SAQ survey instrument and its reliability, and analysis.
Figure 1
Study Design

1. WMU HSIRB approval and Study Hospital IRB approval
2. Finalize agreement with Johns Hopkins University for survey study tool
3. Informational meetings, letters and phone calls to recruit and consent participants
4. Coordinate simulation details with Michigan Center of Excellence in Simulation Research
5. Obtain surveys from Johns Hopkins Univ.
6. Study packets for participants assembled by Study Coordinator with matched serial numbers using Excel spreadsheet. Administer baseline surveys.
7. Randomly assign order of simulations (i.e., which team assigned on a given day)
8. Assign Surgeons to days based on availability by office managers who are blinded to simulation assignment

Cardiac Team
- O.R. – Day 1
- SICU – Day 6
- PCU – Day 4

Vascular Team
- O.R. – Day 2
- SICU – Day 3
- PCU – Day 5

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
Figure 1—Continued

Assess potential (consented) participants for eligibility (n=118)

Excluded (n=10)
- Did not return surveys (n=8)
- Unable to participate (n=2)

Determine participant schedules for assigned days of simulation

Randomize or assign eligible participants (n=108) to Experimental or Control condition based on number of participants required for Simulation and availability to be randomized

58 Participants required for Simulation (−5 technical participants who repeated*) (n=53)

Unique role (e.g. MDs), or refused randomization’ ∴ Assigned to Simulation (n=31)

+1 Substitution from Randomized Control group to replace call-in (n=32)

55 Participants available to serve as Controls

Randomized to Simulation: (n=22)

-1 call-in on Simulation Day (crossed over to Assigned Control) (n=21)

Not available on day of simulation (off, off-shift, vacation) ∴ Assigned to Control (n=25)

+1 Experimental cross-over who called in sick on Simulation Day (n=26)

Randomized to Control: (n=30)

-1 Substitution to Assigned Experimental to replace call-in (n=29)
Follow-up surveys distributed to all Experimental and Control participants one week and six weeks after completion of Simulation exercises.

5 Assigned-Experimental withdrew: Did not return baseline survey (n=4); Only returned baseline survey (n=1)

27 available for analysis

1 Excluded from Analysis due to non-unit based staff: 1 Med Tech

26 Analyzed

45 Experimental Group Participants Analyzed

All Randomized-Experimental Retained

21 available for analysis

2 Excluded from Analysis due to non-unit based staff: 2 Radiographers

19 Analyzed

38 Control Group Participants Analyzed

4 Assigned-Control withdrew: Only returned baseline survey

22 available for analysis

6 Excluded from Analysis due to non-unit based staff: 4 Radiographers 2 Med Techs

16 Analyzed

2 Randomized-Control withdrew: Only returned baseline survey

27 available for analysis

5 Excluded from Analysis due to non-unit based staff: 4 Radiographers 1 Med Tech

22 Analyzed

* Three technical individuals each volunteered to participate in a second simulation to preserve simulation fidelity when patient needs unexpectedly prevented scheduled (assigned) individuals from participating.

† Two nurses refused to commit an unscheduled day to cover randomized participants unless they were guaranteed to participate.
Preliminary Steps

Each of the six simulation exercises involved the participation of 8 to 11 individuals, depending on the number of team members usually involved in a given situation who carried out their normal job responsibilities related to any particular patient. In this case, however, the patient was a mannequin with a simulated emergency relating either to a cardiac or vascular diagnosis, to match the expertise of the surgical team. The mannequin patient was embedded in the regular daily operating room schedule or patient census on the nursing units. Managers of those units were aware of the timing of simulations so they could schedule appropriate staffing to accommodate the 4-hour simulation and debriefing sessions (30-minute introduction, 45-minute simulation exercise, 15-minute break, and subsequent 2.5 hours of debriefing). This ensured adequate coverage for normal patient workloads and made it possible to avoid interference with actual patient care.

Because this process required a significant coordination of effort and commitment of several personnel, it was first necessary to engender widespread support of the proposed exercises. Introductory presentations were given to hospital administrators, physicians and nursing leadership. Following approval of the overall concept by these individuals, preliminary meetings were held with multiple nursing and ancillary personnel managers who were involved with the coordination of the simulation and debriefing exercises.
Participants

Participants were members of the Cardiac Surgery team and the Vascular Surgery team at a 545-bed hospital in a Midwest suburban area. Team members included attending surgeons, general surgery residents, anesthesiologists, physician assistants, advanced practice nurses (nurse practitioners, nurse anesthetists, and clinical nurse specialists), nurses (operating room, surgical intensive care unit, progressive care unit), perfusionists, respiratory therapists, patient care technicians, operating room technologists, radiology technicians, blood bank technicians, and unit clerks.

Eligibility Criteria

All employees who provided direct and indirect care for the cardiac and/or vascular surgery patient populations were candidates for inclusion in the study. To further qualify for inclusion, both full-time and part-time staff had to work in (or have positions with significant work commitments to) the surgical intensive care unit (SICU), operating room (OR), or progressive care unit (PCU) for at least 1 month prior to enrollment in the study. Individuals were not admitted to the study if they had worked in the SICU, OR, or PCU for less than 1 month prior to study enrollment.

Recruitment

The study design and recruitment process was presented to both the Human Subjects Institutional Review Board (HSIRB) at Western Michigan University and the Institutional Review Board (IRB) at the study hospital for approval prior to recruitment of individuals for study enrollment. Employees eligible for the study attended an
informational meeting and were given a packet containing a letter explaining the study and two copies of an informed consent. The consent form was provided in duplicate to allow one copy for participants to sign and return via self-addressed, interdepartmental envelope, and one to keep, to allow thorough review of the consent form and to provide contact information for later reference.

These informational meetings were based on the same format and scheduling as in-service meetings, which are standard on all units to introduce staff to new procedures and equipment. The informational meetings included a PowerPoint presentation, a review of the consent form, and a question-and-answer period. Two hundred thirty packets were distributed during the 26 informational meetings. All consent forms were accompanied by a self-addressed interdepartmental envelope to avoid the potential for perceived coercion to sign the form in the presence of the investigators. Another mechanism to prevent perceived coercion or potential embarrassment of individuals who did not wish to participate was a check box included on the consent form to decline participation. This allowed attendees at the informational sessions to return a consent form even if they were choosing not to enroll in the study.

Not all eligible employees were able to attend an informational session, but in that case they were able to consent based on the descriptive letter in the informational packet. Participants who signed and returned a consent form but did not attend a session were contacted to verify their understanding of the study. All participants were provided with contact information for the investigators and the study coordinator to have any additional questions answered. The participants were also given contact information for the hospital IRB, WMU HSIRB, investigators, study coordinator, and the Michigan
Center of Excellence in Simulation Research in case they had any questions they wanted to ask outside the informational meeting.

Sample Size

Sample size was estimated by the simulation investigators based on their previous experience with experimental In-situ® simulation. Approximately 10 participants are typically involved in each simulation exercise. Sample size was also based on practical concerns created by the number of team members in each job classification. The research team had determined that ideally each member of the cardiac and vascular surgery team should be involved in only one simulation in order to provide the opportunity for participation to as many employees as possible. The investigators determined how many people from each job category are normally in a given cardiac and vascular surgery team by unit and job classification (Table 2). Based on these figures, it was determined that six simulations would expose enough personnel to evaluate whether simulations and debriefing might potentially change attitudes and be beneficial for the study hospital. Based on the number of consented individuals and staffing considerations for simulations, the experimental sample size was 58 participants and the control group was 50 participants. So, practicality ultimately determined sample size rather than any power calculations. The only other study currently in the literature that has examined the influence of simulation based training on climate or safety and teamwork is that published by Marshall (2007), who stated that there is no way to calculate the likely amplitude of effect due to no other studies available, which prevents accurate power calculations.
Table 2  
Participants by Location and Job Category

<table>
<thead>
<tr>
<th>Job Classification</th>
<th>OR Cardiac</th>
<th>OR Vascular</th>
<th>SICU Cardiac</th>
<th>SICU Vascular</th>
<th>PCU Cardiac</th>
<th>PCU Vascular</th>
</tr>
</thead>
<tbody>
<tr>
<td>Surgeon</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Anesthesiologist</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>CRNA</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Resident / PA / NP</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Cardiac CNS</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>OR Nurse</td>
<td>3</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>SICU Nurse</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>PCU Nurse</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Perfusionist</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Med. Technologist</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Respiratory Therapist</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Radiology</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Patient Care Tech</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Unit Clerk</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td><strong>TOTALS</strong></td>
<td><strong>10</strong></td>
<td><strong>8</strong></td>
<td><strong>11</strong></td>
<td><strong>9</strong></td>
<td><strong>10</strong></td>
<td><strong>10</strong></td>
</tr>
</tbody>
</table>


If at any time during a simulation or debriefing an actual patient had required medical care by any of the participants, the medical simulation would have stopped.
Every effort was made by hospital administration to have adequate clinical staff available to cover for the participants randomized to the exercises.

Procedure

Randomization

Randomization occurred in approximately half of the cases as noted above in Figure 1. Exceptions to randomization occurred because some participants who were vital to the simulation and debriefing could not be randomized. Reasons for this were as follows: some providers needed advanced notification for scheduling purposes (surgeons, anesthesiologists); some categories of nurses or support personnel included too few consented participants in the job classification (advanced practice nurses, unit clerks); some managers were concerned about staffing on a particular day (surgical intensive care unit nurses on one simulation day, respiratory therapists and blood bank personnel); and two surgical intensive care unit nurses declined to increase the staffing pool to allow randomization but agreed to come in as simulation participants on a day off. These necessities led 31 participants to be assigned based on availability in order to fill critical roles.

Sequence generation. The following randomization sequence occurred when possible. The list of consented participants by job classification working on each day of the simulation and debriefing exercises was provided to the study hospital’s Quality Institute data manager. Experimental and control groups were randomized by job
classification using a computer generated randomization list written by the statistician. Each participant had an equal probability of being assigned to the groups.

_Allocation concealment._ The simulation dates were determined by the researchers and these dates were given to the office managers of the cardiothoracic and vascular surgery offices. The office managers assigned the surgeons to a date based on availability. The study coordinator remained blinded to the surgeon assignment while randomly determining which simulation would occur on a given day (Cardiac OR, Vascular OR, Cardiac SICU, Vascular SICU, Cardiac PCU, or Vascular PCU). These assignments remained concealed until all remaining nonsurgeon participants were randomized to experimental or control groups.

_Implementation._ Determination of whether a participant would be randomized to the experimental or control group in the OR, SICU or PCU was generated by the computer program written by the study hospital's Quality Institute data manager and was based on participant availability on assigned days. The experimental and control groups were quantitatively matched by job classifications. The randomization list was then given to the study coordinator, who informed managers of each department of the dates and participants to be involved. The managers were previously engaged in collaborative planning and were asked to keep this information confidential.

_Blinding._ The investigators were not involved in the randomization process. They were involved in the debriefing sessions but were blind to the matching of the pre-post follow-up survey questionnaire identifying codes. The study coordinator removed the names associated with the identifying codes prior to giving the investigators the
matched identifying codes. A single master list was maintained by the study coordinator in a secure file at the study hospital to prepare the matching surveys and to monitor their return. The survey scanning process was completely objective and was accomplished by the Johns Hopkins University Quality and Safety Research Group in Baltimore, MD.

**Simulation and Debriefing**

Following the completion of the pre-intervention survey, a series of six simulations and follow-up debriefing sessions were conducted by investigators from the Michigan Center of Excellence in Simulation Research at Western Michigan University. In accordance with their HSIRB protocol application, their *In-situ*® simulation protocol (Hamman et al., 2007) was followed for all six simulation and debriefing sessions.

The six simulations consisted of clinical scenarios in either the OR, surgical intensive care unit (SICU), or progressive care unit (PCU). Various team members with no prior knowledge of the scenario contents were summoned to respond to the scenario. A 30-minute briefing session was held prior to the simulation to give a brief description of the study and to introduce the participants to the mannequin. The mannequin used pre-programmed conditions to simulate a patient scenario requiring intervention from the specialized teams summoned to respond. The participants in the simulation scenario were expected to respond as they would if the mannequin were a live patient, including ordering tests and procedures. Responses to the interventions on the mannequin were controlled by the investigators using headsets or by flashing index cards to indicate changes in the situation from moment to moment to observe reactions of team members.
Video taping of the scenarios were performed by personnel from the Michigan Center of Excellence in Simulation Research.

The debriefing sessions occurred immediately following a 15-minute break after each simulation scenario and were attended by all scenario participants. The debriefings were conducted by investigators from the Michigan Center of Excellence in Simulation Research using video-taped segments of the simulation exercise to help generate discussion and illustrate points for potential improvement. The individuals who conducted the debriefings from the Michigan Center of Excellence in Simulation Research were M.D. pilots with extensive experience in team debriefings in both the fields of aviation and medicine.

Settings and Location of Data Collection

The Safety Attitudes Questionnaire (Sexton et al., 2006; described in detail in the next section) was used to collect data for this study. The questionnaires were distributed to consented participants at their units of employment and were accompanied by self-addressed interdepartmental return envelopes. All returned questionnaires enrolled the individual participant codes in a drawing for gift certificates to a café. The participant codes were known only to the study coordinator, who conducted the drawing for each survey phase.

Additional data were collected throughout the simulation and debriefing sessions by the Michigan Center of Excellence in Simulation Research (Hamman et al., 2007). The Center has received funding for the design of performance markers in professional standards, safety science, and risk management centered on simulation-based tools.
Additionally, the Center’s goals are to evaluate simulation effectiveness and determine the level of simulation fidelity required for effective skill transfer. Researchers from the Center measure performance in small group trials utilizing expertise and simulation facilities from collaborating universities and healthcare organizations.

The Center’s study hypothesis holds that there are context-sensitive behaviors that affect individuals’ effectiveness in team creation and function. The project proposes to identify effective/ineffective behaviors in simulated healthcare scenarios, classify them, devise and implement scenario training to impart the skills, and then to test the hypothesis by comparing “trained” teams with control “groups.” This co-occurring and integrated study was described clearly in the informational sessions and in the informed consent process so that participants were aware that data collected would be used for both studies.

Instrument

The instrument for surveying the participants was the Safety Attitudes Questionnaire (SAQ) developed by Dr. Bryan Sexton at the University of Texas Center of Excellence for Patient Safety Research and Practice. The questionnaire was adapted from aviation safety assessment surveys to measure provider attitudes relevant to patient safety (Sexton et al., 2006). In developing the SAQ for healthcare professionals, Sexton and his colleagues retained 25% of the items original to the Flight Management Attitudes Questionnaire because they applied well to healthcare settings. The rest of the SAQ items were developed after conducting pilot studies in four USA critical care sites. Multiple psychometric analyses were performed on the items included in 10,843
completed questionnaires. Items shown to have weak association were deleted to achieve a satisfactory model fit, which was then tested for reliability. This scale reliability testing (Cronbach's α 0.79) combined with multilevel factor analyses demonstrated that the SAQ is reliable and formally validated (Colla, Bracken, Kinney, & Weeks, 2005; Kaissi et al., 2003; Marshall et al., 2006; Pronovost & Sexton, 2005; Sexton et al., 2006; Sexton, Thomas, & Pronovost, 2005).

Two versions of this survey were used: the SAQ–ICU Version and the SAQ–OR Version. Each questionnaire has 30 core questions that provide the factor structure. These core questions have been validated for determining change in the six domains: teamwork climate, safety climate, job satisfaction, stress recognition, perception of management, and working conditions. These six domains have been shown to correlate with patient safety. The ICU questionnaire has 35 additional items and the OR version 29 additional items. These questions were retained because nurse managers and executives found them to be informative (J. Sexton, personal communication, July 1, 2007).

The ICU version of the questionnaire is a single page (double-sided) questionnaire with 65 items; the OR version has 59 items. Both questionnaires ask for demographic information (age, sex, job status, and ethnic group). Each of the questions is answered using a 5-point Likert scale (Disagree Strongly, Disagree Slightly, Neutral, Agree Slightly, Agree Strongly).

The questionnaires have an open-ended section for comments, as well as a section that addresses questions on collaboration and communication. In the comments section respondents are asked, "What are your top three recommendations for improving
patient safety in this clinical area?" In the collaboration and communication section, respondents are asked to use a 5-point Likert scale (Very Low, Low, Adequate, High, Very High) to describe the quality of communication and collaboration they have experienced with each type of provider in their clinical area.

Survey Domains

To answer the research questions, the analysis of the survey was divided into six domains as previously described by Sexton et al. (2006) including teamwork climate, safety climate, job satisfaction, stress recognition, perception of management, and working conditions. The teamwork climate domain detects the perceived quality of collaboration between personnel and includes the following statements:

1. It is easy for personnel in this ICU/OR to ask questions when there is something they do not understand.
2. I have the support I need from other personnel to care for our patients.
3. Nurse input is well received in this ICU/OR.
4. In this ICU/OR, it's difficult to speak up if I perceive a problem with patient care.
5. Disagreements in the ICU/OR are appropriately resolved.
6. The physicians and nurses here work as a well coordinated team.

The safety climate domain reveals perceptions of a strong and proactive organizational commitment to patient safety and includes the following statements:

1. The culture in this ICU/OR makes it easy to learn from errors of others.
2. Medical errors are handled appropriately in this ICU/OR.
3. I know the proper channels to direct questions regarding patient safety.
4. I am encouraged by my colleagues to report any patient safety concerns.
5. I receive appropriate feedback about my performance.
6. I would feel safe being treated here as a patient.
7. In this ICU/OR it is difficult to discuss errors.

The job satisfaction domain examines how positive the work experience is and includes the following statements:

1. Morale in the ICU/OR is high.
2. Working in this hospital is like being part of a large family.
3. I am proud to work at this hospital.
4. This hospital is a good place to work.
5. I like my job.

Stress recognition items acknowledge how the participants feel their performance is influenced by stressors and include the following statements:

1. When my workload becomes excessive, my performance is impaired.
2. I am more likely to make errors in hostile or tense situations.
3. Fatigue impairs my performance during emergency situations.
4. I am less effective at work when fatigued.

Perception of management statements indicate participant approval of managerial actions and include the following statements:

1. Hospital management does not knowingly compromise the safety of patients.
2. Hospital administration supports my daily efforts.
3. I am provided with adequate, timely information about hospital events.
4. The levels of staffing in this ICU/OR are sufficient to handle the number of patients.

The working condition domain examines the perceived quality of work environment and logistical support and includes the following statements:

1. All the necessary information for diagnosis and therapeutic decisions is available.

2. This hospital constructively deals with problem physicians and employees.

3. Trainees in my discipline are adequately supervised.

4. This hospital does a good job of training new personnel.

Each consented participant was given an initial pretest survey and then randomized (when possible) to an experimental or control group. One week and 6 weeks after the simulation and debriefing sessions were completed, each participant was given the survey again. Each phase of the survey (pre, 1 week post, and 6 weeks post) was coded to allow individuals to be appropriately matched so that pre- and postsimulation surveys could be treated as repeated measures on the same individual. The investigators remained blind to the identifying codes; only the study coordinator had access to the identifying codes. Participants were informed of this identifying procedure.

The pretest and follow-up survey identifying codes were tracked by the study coordinator using an Excel spreadsheet. The investigators were involved in the debriefing sessions but were blind to the matching of the pre-post test questionnaire identifying codes. The study coordinator removed the names associated with the identifying codes prior to giving the investigators the matched surveys. A single master
list was maintained in a secure fashion by the study coordinator in order to prepare the matching surveys and to monitor their return.

Dr. Sexton’s current affiliation is with Johns Hopkins University (JHU) Quality and Safety Research Group, with whom a contract was reviewed and signed by the investigators and the study hospital. In signing this contract, the investigators agreed to follow administrative guidelines to make an effort to elicit an 80% or higher response rate from the sample population. The study hospital also agreed to pay a fee for the surveys, tracking worksheets, visual inspection and scanning of the surveys, conversion of the data to Statistical Product and Service Solutions (SPSS), Version 15 for later analysis by the investigators and benchmarking data and analysis.

Survey Administrative Guidelines

The SAQ project manager at JHU outlined a central survey administration process timeline (S. Grillo, personal communication, April 5, 2006) used to standardize administration of the surveys (Sexton et al., 2003). This process timeline guided the investigators and research coordinator to achieve an acceptable response rate from the participants.

Reliability

Many researchers have concluded that the Safety Attitudes Questionnaire—ICU Version and the Safety Attitudes Questionnaire—OR version are reliable and valid and demonstrate good psychometric properties (Colla et al., 2005; Kaissi et al., 2003; Marshall et al., 2006; Pronovost & Sexton, 2005; Sexton et al., 2005; Sexton et al., 2006). Based on the findings in assessing the effectiveness of the SAQ, Sexton et al.
(2006) stated that the questionnaire can be used to benchmark with other institutions, to implement interventions aimed at improving attitudes regarding safety, and to measure the effectiveness of those interventions.

Analysis

Data Management and Processing

Research assistants from JHU visually inspected surveys as described by Sexton et al. (2006). Surveys were then scanned and converted into SPSS Version 15 file for analysis. The Likert scales described above were used to score each of the items. Negatively worded items were reverse-scored to match the positively worded items.

Calculation of Domain Scores

The calculation used to determine participant attitudes for each of the six domains is as follows (J. Sexton & S. Grillo, personal communication, January 9, 2007):

To convert the 5-point Likert scale item scores to a 100-point scale score (e.g., teamwork climate) for an individual respondent:

1. Reverse score all negatively worded items (two in 30 core questions)
   a. In this ICU, it is difficult to speak up if I perceive a problem with patient care.
   b. In this ICU, it is difficult to discuss errors.

2. Calculate the mean of the set of items from the scale.

3. Subtract 1 from the mean.

4. Multiply the result by 25.
The equation is as follows: \((\text{Mean of teamwork item scores} - 1) \times 25 = \text{Teamwork Climate Score for a Respondent}\)

Calculation of the percent of respondents who are positive (i.e., percent agreement), is accomplished by looking at the percentage of respondents who had a scale score of 75 or higher. A score of 75 on the scale score indicates the same thing as “agree slightly” on the original 5-point Likert scale (1 = Disagree Strongly, 2 = Disagree Slightly, 3 = Neutral, 4 = Agree Slightly, 5 = Agree Strongly). A summary of the Likert Scale conversion to the 100-point scale is demonstrated in Table 3.

Table 3

Likert Scale Conversion

<table>
<thead>
<tr>
<th>Likert Scale</th>
<th>Conversion</th>
<th>100 Point Scale</th>
</tr>
</thead>
<tbody>
<tr>
<td>5 = Strongly Agree</td>
<td>Mean (\bar{X} - 1) \times 25</td>
<td>100</td>
</tr>
<tr>
<td>4 = Agree Slightly</td>
<td></td>
<td>75</td>
</tr>
<tr>
<td>3 = Neutral</td>
<td></td>
<td>50</td>
</tr>
<tr>
<td>2 = Disagree Slightly</td>
<td></td>
<td>25</td>
</tr>
<tr>
<td>1 = Disagree Strongly</td>
<td></td>
<td>0</td>
</tr>
</tbody>
</table>

Responses to individual SAQ items were grouped into six summary scores as outlined below in Table 4 (J. Sexton, personal communication, July 20, 2007).

Sexton (J. Sexton, personal communication, July 20, 2007) outlines his interpretive considerations for the results of the SAQ as follows:
Table 4

**Definition of Domain Scores**

<table>
<thead>
<tr>
<th>Factor: Definition</th>
<th>Example Items</th>
</tr>
</thead>
</table>
| **Teamwork climate:** perceived quality of collaboration between personnel       | • Disagreements in this clinical area are appropriately resolved (i.e., what is best for the patient).  
• Our doctors and nurses work together as a well coordinated team.               |
| **Safety climate:** perceptions of a strong and proactive organizational commitment to safety | • I would feel safe being treated in this clinical area.  
• Medical errors are handled appropriately in this clinical area.                |
| **Job satisfaction:** positivity about the work experience                          | • I like my job.  
• This hospital is a good place to work.                                        |
| **Stress recognition:** acknowledgment of how performance is influenced by stressors | • I am less effective at work when fatigued.  
• When my workload becomes excessive, my performance is impaired.                |
| **Perceptions of management:** approval of managerial action                       | • Hospital management supports my daily efforts in this clinical area.  
• Hospital management does not knowingly compromise the safety of patients.     |
| **Working conditions:** perceived quality of the work environment and logistical support (staffing, training, etc.) | • Trainees in my discipline are adequately supervised.  
• This hospital deals constructively with problem personnel.                      |
1. **Results are expressed as percentages rather than means** *(Likert scores are converted to a 100 point scale as described above).*
   
a. Percents are better predictors of subsequent outcomes.
   
b. Percents do not mask variability in small groups the way means do.

2. **A difference of 10 points in a scale is significant if representative response rates of at least 60% are attained** *(e.g., 10 points of improvement within an area, or 10 points of difference between areas is significantly different).*

3. **Culture target of 80% percent or higher.** There is a threshold of excellence at 80%. For example, when four out of five people agree that teamwork climate is good, the consensus is strong and stable over time. Further improvement from 80% to 100% is not associated with the same improvements in outcomes as improvement from 50% to 70%.

4. **Danger Zone** *(needs improvement).* There is a threshold of risk below 60%, for example, when fewer than 3 out of 5 people agree that safety climate is good, the lack of consensus in assessments of frontline caregivers indicates increased potential for substandard clinical and operational outcomes. When at least 60% of respondents have positive domain scores *(≥ 75)* that domain is considered safe.

5. **Culture goals.**
   
a. Improve by 10 points or more on an item or overall domain score
   
b. *Get out of the Danger Zone*
Data Analysis

Statistical analyses were performed using the Statistical SPSS, version 15.0 for Windows (SPSS, Inc., Chicago, IL) and SAS® Version 9.00 (SAS Institute, Cary, NC). Responses to the survey questions were analyzed according to the six domains (teamwork climate, safety climate, job satisfaction, stress recognition, perception of management, and working conditions) in the experimental and control groups.

Statistical methods. Statistical methods used to compare groups for primary outcomes included summarizing the descriptive data using percentages, frequencies, means, medians and modes. The distribution of the experimental and control groups were examined using skewness and kurtosis to determine whether the sample met the assumptions for conducting parametric tests. Additional objective testing of normality was conducted on the experimental and control groups on all six domain climate scores at baseline for both the OR and ICU SAQ using the Shapiro-Wilk test. The Levene test was used to assess homogeneity of variance on both the OR and ICU SAQ. Descriptive data were examined for both the OR and ICU questionnaires using Sexton’s criteria as outlined above.

Inferential statistics were used to assess the association between the independent variables (experimental or control and time) and the dependent variables (baseline, 1 week, and 6 week SAQ domain climate scores, which include teamwork climate, safety climate, job satisfaction, stress recognition, perception of management, and working conditions). Repeated measures ANOVA were used to assess individual scale scores of all participants, those with low baseline teamwork climate scores, as well as the
experimental and control groups individually. When the SAQ domain scale scores were statistically significant, a breakdown of individual domain questions was analyzed using a paired $t$ test. General estimated equations (GEE) analyses were performed to look at the effect of simulation on individual survey scores over time. For these, focus was restricted to participants in the randomized-experimental and randomized-control groups (see Figure 1). GEE were used to examine the binary outcomes of whether or not a survey score was at or above 75.

Quality of communication and collaboration. Participants were asked to use a Likert scale (Very Low, Low, Adequate, High, Very High) to describe the quality of communication and collaboration they have experienced with each type of provider in their clinical area. The baseline, 1 week, and 6 week responses were described as percentage above adequate (4 or 5 on the Likert scale).

Coding of responses to open-ended questions. The study coordinator and the researcher independently coded responses to the open-ended comment section of the questionnaire that asked, “What are your top three recommendations for improving patient safety in the OR/SICU?” The responses were then combined and the top 10 recommendations are reported for each phase of the survey (baseline, 1 week, and 6 weeks).
CHAPTER IV

RESULTS

The OR and ICU Safety Attitudes Questionnaires developed by Sexton et al. (2006) were used to study individual and aggregate attitudes of the members of two subspecialty surgery teams (cardiac surgery and vascular surgery) in two environments (OR and ICU) in the six domains of teamwork climate, safety climate, job satisfaction, stress recognition, perception of management, and working conditions after being exposed to medical team simulation and debriefing. The study took place in a single 545-bed hospital in the Midwest with a high-volume cardiovascular surgery program.

The study was designed to be a randomized controlled experiment, but the logistics of running such a complex project in situ made it necessary to assign some personnel (e.g., physicians), so that the initial set of inferential analytical results are better characterized as the results of a pre-post partially randomized controlled, quasi-experimental design. The chapter begins with a presentation of the results of descriptive analysis. The two sets of inferential statistic analysis come next, and qualitative data are considered last.

Research Questions and Hypotheses

The purpose of this study was to determine change in safety attitudes of the members of two subspecialty surgery teams (cardiac surgery and vascular surgery) after exposure to medical simulation and debriefing by the Michigan Center of Excellence in
Simulation Research. The study was designed to answer two primary questions by testing the null hypothesis form of each research question. The primary questions are stated below along with the research hypotheses in their alternative forms:

1. Do pre-post1-post6 (baseline, 1 week, and 6 week) attitudes of surgical team members (as measured by an OR SAQ) change after being exposed to medical simulation and debriefing, compared with changes at pre-post1-post6 points for members of control group teams?

   Research Hypothesis 1a: Exposure to team medical simulation and debriefing will result in a change in attitudes regarding teamwork climate, safety climate, job satisfaction, stress recognition, perception of management, and working conditions among members of cardiac and vascular surgery teams as measured by an OR SAQ.

2. Do pre-post1-post6 (baseline, 1 week, and 6 week) attitudes of surgical team members (as measured by an ICU SAQ) change after being exposed to medical simulation and debriefing, compared with changes at pre-post1-post6 points for members of control group teams?

   Research Hypothesis 2a: Exposure to team medical simulation and debriefing will result in a change in attitudes regarding teamwork climate, safety climate, job satisfaction, stress recognition, perception of management, and working conditions among members of cardiac and vascular surgery teams as measured by an ICU SAQ.
Descriptive Analyses of Sample Characteristics

A total of 118 individuals consented to participate in the study. Two of these individuals later were unable to participate due to medical leave, and 8 did not return baseline surveys. Of the 108 remaining participants, 52 were available to be randomized and 56 had to be assigned based on availability due to logistical concerns (e.g., surgeon schedule limitations, too few participants in a given job category, etc.). Allocation to experimental and control groups included 53 participants in the simulation and debriefing combined-experimental group (31 assigned-experimental and 22 randomized-experimental) and 55 in the control group. In the combined-experimental group 1 randomized participant called in sick on the day of the simulation and 1 control participant was reassigned to replace this person in the simulation exercise. This resulted in 32 participants in the assigned-experimental group and 21 in the randomized-experimental group. The control group consisted of 55 participants of which 30 were randomized-control and 25 were assigned-control.

After simulation and debriefing and return of the follow-up surveys, 5 participants withdrew from the assigned-experimental group: 4 had not completed the baseline survey prior to simulation (which was not recognized until after the simulation), and 1 returned only the baseline survey (no follow-up surveys). A sixth assigned-experimental participant was withdrawn later due to being non-unit based staff. In the randomized-experimental group, 2 participants were excluded from analysis because

2 “Non-unit based staff” refers to individuals in technical roles who were crucial to simulation fidelity but were unable to validly complete OR or ICU SAQ items due to their “home unit” referent. This included radiographers and medical technologists who are affiliated with all hospital units but are not based in any one particular unit that permits consistent, daily interaction with the surgical team members being evaluated by this study.
they were non-unit based staff. In the combined-control group 6 participants withdrew because they only returned a baseline survey. An additional 11 participants were excluded from analysis because they were non-unit based staff. This resulted in a final analysis pool of 83 individuals, 45 of whom were experimental (26 assigned-experimental, 19 randomized-experimental) and 38 were control (16 assigned-control, 22 randomized-control) (Figure 1).

Twenty participants (surgeons and PAs) have dual roles in the OR and ICU and could have been randomized to either clinical area. As such, those individuals were given both OR and ICU SAQs to complete at all stages of the study. They were instructed to carefully consider their roles and relationships in the specified unit (OR vs. ICU) as they completed each questionnaire. Concerns regarding this dual response condition were addressed by the participants completing both surveys, who indicated that they would have no trouble assuming alternate “mindsets” for responding to surveys pertaining to what they regarded as very different environments (OR vs. ICU). There is no way to completely rule out, however, that the responses to one of the questionnaires might not have influenced responses to the other. It also is possible that having responses of a few of the same people represented in both the OR and ICU surveys might have influenced the results. All participants who had dual roles maintained their experimental \( (n = 15) \) or control \( (n = 5) \) designation for both units to avoid confounding of experimental and control group representation. After all adjustments were made, data were analyzed for 46 participants taking the OR SAQ (26 combined-experimental and 20 control), and for 57 participants completing the ICU SAQ (34 combined-experimental and 23 control).
A total of 392 SAQ surveys were distributed to the 118 originally consented participants, 355 of which were returned. This resulted in a total project response rate of 90.6%. By SAQ type, 143 OR surveys were distributed and 138 were returned (96.5%); 249 ICU surveys were distributed and 217 were returned (87.1%). Further breakdown of response rates by baseline, 1 week, and 6 week distribution is shown in Table 5.

Table 5

<table>
<thead>
<tr>
<th>% Response Rates Based on Time of Distribution</th>
<th>OR SAQ</th>
<th>ICU SAQ</th>
<th>Non-Unit Based Staff (ICU SAQ)</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td>94.0</td>
<td>92.8</td>
<td>71.4</td>
<td>90.0</td>
</tr>
<tr>
<td>1 Week</td>
<td>97.9</td>
<td>84.1</td>
<td>93.3</td>
<td>90.4</td>
</tr>
<tr>
<td>6 Week</td>
<td>97.8</td>
<td>88.7</td>
<td>84.2</td>
<td>91.3</td>
</tr>
<tr>
<td>Overall</td>
<td>96.5</td>
<td>88.7</td>
<td>81.8</td>
<td>90.6</td>
</tr>
</tbody>
</table>

Operating Room Safety Attitudes Questionnaire Descriptives

The mean age of the sample population (n = 46) for the operating room condition was 41.5 years (range 23 to 63), with 20 males (43%) and 26 females (57%). Ethnicity of the sample population included 2.2% Hispanic, 4.3% Black, 84.8% White, 4.3%

3 As individuals dropped out, they were not given follow-up surveys at the next point of distribution. Non-unit based staff were surveyed at all points as the decision to exclude them from analysis was not made until after follow-up was complete.
Asian/Pacific Islander, 2.2% Multi-Ethnic and 2.2% Other. The mean number of years of experience was 12 and the mean number of years the participants had worked in the hospital was 12 years. These descriptives are further delineated by randomization status (assigned-experimental vs. randomized-experimental) in Table 6.

Table 6

Descriptive Analyses of OR Questionnaire Respondents

<table>
<thead>
<tr>
<th>Variable</th>
<th>Combined-Experimental Group (n = 26)</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Assigned (n = 15)</td>
<td>Randomized (n = 11)</td>
<td>Control (n = 20)</td>
<td>Total (n = 46)</td>
</tr>
<tr>
<td>Mean Age (in years)</td>
<td>38.7</td>
<td>43</td>
<td>42.9</td>
<td>41.5</td>
</tr>
<tr>
<td>Gender: Male</td>
<td>9</td>
<td>5</td>
<td>6</td>
<td>20</td>
</tr>
<tr>
<td>Female</td>
<td>6</td>
<td>6</td>
<td>14</td>
<td>26</td>
</tr>
<tr>
<td>Ethnicity (% of participants)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hispanic</td>
<td>–</td>
<td>–</td>
<td>5.0</td>
<td>2.2%</td>
</tr>
<tr>
<td>Black</td>
<td>6.7</td>
<td>9.1</td>
<td>–</td>
<td>4.3%</td>
</tr>
<tr>
<td>White</td>
<td>73.3</td>
<td>90.9</td>
<td>90.0</td>
<td>84.8%</td>
</tr>
<tr>
<td>Asian/Pacific Islander</td>
<td>13.3</td>
<td>–</td>
<td>–</td>
<td>4.3%</td>
</tr>
<tr>
<td>Multi-Ethnic</td>
<td>6.7</td>
<td>–</td>
<td>–</td>
<td>2.2%</td>
</tr>
<tr>
<td>Other</td>
<td>–</td>
<td>–</td>
<td>5.0</td>
<td>2.2%</td>
</tr>
<tr>
<td>Mean Years in Specialty</td>
<td>9.5</td>
<td>12</td>
<td>13.8</td>
<td>12</td>
</tr>
<tr>
<td>Mean Years with Organization</td>
<td>6.9</td>
<td>16.4</td>
<td>12.9</td>
<td>12</td>
</tr>
</tbody>
</table>

As described in Chapter III, the SAQ uses a Likert scale. Scores from each of the six domains are converted to climate scale scores, which are expressed as percentages.
This places the scores on the same scale so they are comparable across domains.

Measures of central tendency of the climate scores are shown in Table 7.

Table 7

*Central Tendency of Baseline OR SAQ Climate Scores*

<table>
<thead>
<tr>
<th>Experimental Condition/Domain</th>
<th>Mean</th>
<th>Median</th>
<th>Mode</th>
<th>Std. Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Assigned-Experimental</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Teamwork Climate</td>
<td>73.89</td>
<td>70.83</td>
<td>70.83</td>
<td>14.47</td>
</tr>
<tr>
<td>Safety Climate</td>
<td>67.38</td>
<td>71.43</td>
<td>71.43</td>
<td>15.79</td>
</tr>
<tr>
<td>Job Satisfaction</td>
<td>79.67</td>
<td>85.00</td>
<td>90.00</td>
<td>15.17</td>
</tr>
<tr>
<td>Stress Recognition</td>
<td>66.67</td>
<td>68.75</td>
<td>68.75</td>
<td>22.98</td>
</tr>
<tr>
<td>Perceptions of Management</td>
<td>68.33</td>
<td>75.00</td>
<td>87.50</td>
<td>19.54</td>
</tr>
<tr>
<td>Working Conditions</td>
<td>70.69</td>
<td>68.75</td>
<td>81.25</td>
<td>14.14</td>
</tr>
<tr>
<td><strong>Randomized-Experimental</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Teamwork Climate</td>
<td>70.69</td>
<td>75.00</td>
<td>75.00</td>
<td>7.06</td>
</tr>
<tr>
<td>Safety Climate</td>
<td>78.57</td>
<td>78.57</td>
<td>78.57</td>
<td>6.78</td>
</tr>
<tr>
<td>Job Satisfaction</td>
<td>80.45</td>
<td>85.00</td>
<td>90.00</td>
<td>14.04</td>
</tr>
<tr>
<td>Stress Recognition</td>
<td>57.38</td>
<td>62.50</td>
<td>62.50*</td>
<td>20.31</td>
</tr>
<tr>
<td>Perceptions of Management</td>
<td>67.61</td>
<td>68.75</td>
<td>75.00</td>
<td>11.46</td>
</tr>
<tr>
<td>Working Conditions</td>
<td>72.92</td>
<td>68.75</td>
<td>68.75</td>
<td>11.14</td>
</tr>
<tr>
<td><strong>Control</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Teamwork Climate</td>
<td>73.00</td>
<td>72.92</td>
<td>100.00</td>
<td>19.83</td>
</tr>
<tr>
<td>Safety Climate</td>
<td>76.73</td>
<td>78.57</td>
<td>82.14*</td>
<td>15.76</td>
</tr>
<tr>
<td>Job Satisfaction</td>
<td>81.06</td>
<td>82.50</td>
<td>60.00</td>
<td>13.90</td>
</tr>
<tr>
<td>Stress Recognition</td>
<td>60.31</td>
<td>60.31</td>
<td>75.00*</td>
<td>27.30</td>
</tr>
<tr>
<td>Perceptions of Management</td>
<td>71.25</td>
<td>71.25</td>
<td>62.50*</td>
<td>14.11</td>
</tr>
<tr>
<td>Working Conditions</td>
<td>68.02</td>
<td>71.88</td>
<td>75.00</td>
<td>16.21</td>
</tr>
</tbody>
</table>

* Multiple modes exist. The smallest value is shown.
Distribution of the randomized-experimental, assigned-experimental, and control climate scale scores were examined using skewness and kurtosis, with the results as follows for the OR SAQ: (a) teamwork climate scale scores in the randomized-experimental and assigned-experimental groups were skewed to the right (0.432 / 0.070), the control group was skewed to the left (-0.358), and the distributions reflected leptokurtosis in the randomized-experimental group (0.414) and platykurtosis in the assigned-experimental and control groups (-0.877 / -0.165); (b) safety climate scale scores in the randomized-experimental group were skewed to the right (0.644) with leptokurtosis (1.096), and scores for the assigned-experimental and control groups were skewed to the left (-1.001 / -0.204) with leptokurtosis in the assigned group (0.414) and platykurtosis in the control group (-0.734); (c) job satisfaction scale scores in the randomized-experimental, assigned-experimental, and control groups were skewed to the left (-1.337 / -1.502 / -0.402), with leptokurtosis in the randomized-experimental and assigned-experimental groups (1.084 / 1.115), and platykurtosis in the control group (-1.149); (d) stress recognition scale scores in the randomized-experimental, assigned-experimental, and control groups were skewed to the left (-0.903 / -0.469 / -0.600) and reflected platykurtosis (-0.786 / -0.107 / -0.222); (e) perceptions of management scale scores in the randomized-experimental, assigned-experimental, and control groups were skewed to the left (-0.507 / -0.349 / -0.068) and reflected platykurtosis (-1.109 / -1.340 / -0.322); and (f) working conditions scale scores in the randomized-experimental and assigned-experimental groups were skewed to the right (0.432 / 0.070), and the control group scores were skewed to the left (-0.358), with leptokurtosis in the randomized-experimental group (0.414) and platykurtosis in the assigned-experimental and control
groups (−0.877 / −0.165). These findings indicate a non-normal distribution of data on the OR SAQ of varying degrees and types across the six domains.

Intensive Care Unit Safety Attitudes Questionnaire Descriptives

The mean age of the sample population (n = 57) who completed the ICU SAQ was 41 years (range 23 to 60) with 17 males (30%), 39 females (68%), and one missing data point (2%). Ethnicity of the sample population included 1.8% Hispanic, 1.8% Black, 89.5% White, 3.5% Asian/Pacific Islander, 1.8% Multi-Ethnic and 1.8% Other. The mean number of years of experience was 12 and the mean number of years the participants had worked in their designated units was 9 years (Table 8).

Measures of central tendency of the climate scores are shown in Table 9. The distribution of the randomized-experimental, assigned-experimental, and control climate scale scores were examined using skewness and kurtosis as follows for the ICU SAQ: (a) teamwork climate scale scores in the randomized-experimental and control groups were skewed to the right (0.127 / 0.102) and the assigned-experimental group was skewed to the left (−1.05) and reflected leptokurtosis in the assigned-experimental group (1.78) and platykurtosis in the randomized-experimental and control groups (−0.847 / −1.24); (b) safety climate scale scores in the randomized-experimental group are skewed to the left (−0.689) with leptokurtosis (0.347), and the assigned-experimental and control groups were skewed to the right (0.121 / 0.265) with platykurtosis in the assigned-experimental group (−0.667) and leptokurtosis in the control group (0.291); (c) job satisfaction scale scores in the randomized-experimental and assigned groups were skewed to the left (−0.447 / −1.025), and the control group was skewed to the right
(0.267), with platykurtosis in the randomized-experimental and control groups (−1.123 / −0.38) and leptokurtosis in the assigned-experimental group (0.282); (d) stress recognition scale scores in the randomized-experimental, assigned-experimental, and control groups were skewed to the left (−0.627 / −0.199 / −0.591) and reflected

Table 8

Descriptive Analyses: ICU Questionnaire Respondents

<table>
<thead>
<tr>
<th>Variable</th>
<th>Combined-Experimental Group (n = 33)</th>
<th>Assigned (n = 21)</th>
<th>Randomized (n = 13)</th>
<th>Control (n = 23)</th>
<th>Total (n = 57)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean Age (in years)</td>
<td>41</td>
<td>44</td>
<td>39</td>
<td>41</td>
<td></td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>9</td>
<td>4</td>
<td>4</td>
<td>17</td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>12</td>
<td>8</td>
<td>19</td>
<td>39</td>
<td></td>
</tr>
<tr>
<td>Missing Data Point</td>
<td>–</td>
<td>1</td>
<td>–</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Ethnicity (% of participants)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hispanic</td>
<td>–</td>
<td>–</td>
<td>4.2</td>
<td>1.8%</td>
<td></td>
</tr>
<tr>
<td>Black</td>
<td>4.8</td>
<td>–</td>
<td>–</td>
<td>1.8%</td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>81</td>
<td>100</td>
<td>91.7</td>
<td>89.5%</td>
<td></td>
</tr>
<tr>
<td>Asian/Pacific Islander</td>
<td>9.5</td>
<td>–</td>
<td>–</td>
<td>3.5%</td>
<td></td>
</tr>
<tr>
<td>Multi-Ethnic</td>
<td>4.8</td>
<td>–</td>
<td>5.0</td>
<td>1.8%</td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>–</td>
<td>–</td>
<td>4.2</td>
<td>1.8%</td>
<td></td>
</tr>
<tr>
<td>Mean Years in Specialty</td>
<td>9.5</td>
<td>12</td>
<td>13.8</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td>Mean Yrs in Designated Unit</td>
<td>6.9</td>
<td>16.4</td>
<td>12.9</td>
<td>9</td>
<td></td>
</tr>
</tbody>
</table>
## Table 9

**Central Tendency of Baseline ICU SAQ Climate Scores**

<table>
<thead>
<tr>
<th>Experimental Condition/Domain</th>
<th>Mean</th>
<th>Median</th>
<th>Mode</th>
<th>Std. Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Assigned-Experimental</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Teamwork Climate</td>
<td>75.99</td>
<td>79.17</td>
<td>75.00</td>
<td>14.00</td>
</tr>
<tr>
<td>Safety Climate</td>
<td>74.15</td>
<td>75.00</td>
<td>71.43*</td>
<td>14.37</td>
</tr>
<tr>
<td>Job Satisfaction</td>
<td>78.09</td>
<td>85.00</td>
<td>90.00</td>
<td>17.28</td>
</tr>
<tr>
<td>Stress Recognition</td>
<td>60.41</td>
<td>62.50</td>
<td>68.75</td>
<td>20.95</td>
</tr>
<tr>
<td>Perceptions of Management</td>
<td>64.58</td>
<td>68.75</td>
<td>62.50*</td>
<td>19.09</td>
</tr>
<tr>
<td>Working Conditions</td>
<td>66.16</td>
<td>68.75</td>
<td>68.75</td>
<td>18.12</td>
</tr>
<tr>
<td><strong>Randomized-Experimental</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Teamwork Climate</td>
<td>60.76</td>
<td>62.50</td>
<td>62.50</td>
<td>15.93</td>
</tr>
<tr>
<td>Safety Climate</td>
<td>74.10</td>
<td>76.78</td>
<td>78.57</td>
<td>12.00</td>
</tr>
<tr>
<td>Job Satisfaction</td>
<td>60.00</td>
<td>65.00</td>
<td>65.00*</td>
<td>18.21</td>
</tr>
<tr>
<td>Stress Recognition</td>
<td>67.70</td>
<td>71.87</td>
<td>43.75*</td>
<td>20.61</td>
</tr>
<tr>
<td>Perceptions of Management</td>
<td>48.09</td>
<td>56.25</td>
<td>56.25</td>
<td>22.89</td>
</tr>
<tr>
<td>Working Conditions</td>
<td>51.04</td>
<td>50.00</td>
<td>68.75</td>
<td>18.23</td>
</tr>
<tr>
<td><strong>Control</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Safety Climate</td>
<td>73.36</td>
<td>73.21</td>
<td>75.00</td>
<td>10.80</td>
</tr>
<tr>
<td>Job Satisfaction</td>
<td>74.37</td>
<td>75.00</td>
<td>70.00*</td>
<td>12.88</td>
</tr>
<tr>
<td>Stress Recognition</td>
<td>61.45</td>
<td>62.50</td>
<td>50.00</td>
<td>24.57</td>
</tr>
<tr>
<td>Perceptions of Management</td>
<td>56.60</td>
<td>56.25</td>
<td>50.00</td>
<td>14.96</td>
</tr>
<tr>
<td>Working Conditions</td>
<td>67.19</td>
<td>65.63</td>
<td>81.25</td>
<td>16.20</td>
</tr>
<tr>
<td>Teamwork Climate</td>
<td>72.57</td>
<td>70.83</td>
<td>54.17*</td>
<td>14.43</td>
</tr>
</tbody>
</table>

* Multiple modes exist. The smallest value is shown.
leptokurtosis in the randomized-experimental group (0.391) and platykurtosis in the assigned-experimental and control groups (−0.115 / −0.001); (e) perceptions of management scale scores in the randomized-experimental and assigned-experimental groups were skewed to the left (−1.022 / −0.652) and scores for the control group were skewed to the right (0.091) and reflected leptokurtosis in the randomized-experimental and assigned-experimental groups (0.811 / 0.536) and platykurtosis in the control group (−0.313); and (f) working conditions scale scores in the randomized-experimental were skewed to the right (0.382) and in the assigned-experimental and control groups were skewed to the left (−0.856 / −0.108) with leptokurtosis in the randomized-experimental and assigned-experimental groups (0.757 / 0.392) and platykurtosis in the control group (−0.885). These findings indicate a non-normal distribution of data on the ICU SAQ.

Nonparametric Test of Normality

In view of the results of the parametric tests of normality, a nonparametric objective test of normality was conducted on the randomized-experimental, assigned-experimental, and control groups on all six domain climate scores at baseline for both the OR and ICU SAQ using the Shapiro-Wilk test. Results for the OR SAQ the results were as follows: (a) normality was achieved on the teamwork climate scales for the randomized-experimental, assigned-experimental, and control groups, D(11) = 0.903, ns⁴; D(15) = 0.958, ns; and D(20) = 0.948, ns, respectively; (b) on the safety climate scale scores normality was achieved for the randomized-experimental, assigned-experimental, and control groups, D(11) = 0.948, ns; D(15) = 0.890, ns; and D(20) =

⁴ ns – nonsignificant.
0.959, ns, respectively; (c) on the job satisfaction scale scores the randomized-experimental, assigned-experimental, and control groups all appeared to be non-normal, $D(11) = 0.838, p < 0.05$; $D(15) = 0.727, p < 0.05$; and $D(15) = 0.901, p < 0.05$, respectively; (d) on the stress recognition climate scale scores the randomized-experimental group appeared non-normal, $D(11) = 0.812, p < 0.05$, and the assigned-experimental and control groups appeared normal, $D(15) = 0.956, ns$; $D(20) = 0.953, ns$; and $D(20) = 0.953, ns$ respectively; (e) on the perceptions of management climate scale scores the randomized-experimental, assigned-experimental, and control groups appeared normal, $D(11) = 0.897, ns$; $D(15) = 0.897$; and $D(20) = 0.972, ns$, respectively; (f) and on the working conditions climate scale scores the randomized-experimental, assigned-experimental, and control groups appeared normal, $D(11) = 0.938, ns$; $D(15) = 0.947$; and $D(20) = 0.950, ns$, respectively. The OR SAQ data were normal based on the Shapiro-Wilk test with the exceptions of job satisfaction (randomized-experimental, $p = 0.030$; assigned-experimental, $p = 0.000$; and control, $p = 0.043$) and stress recognition (randomized-experimental, $p = 0.014$).

Shapiro-Wilk objective test of normality on the ICU SAQ revealed: (a) normality for teamwork climate scales for the randomized-experimental, assigned-experimental, and control groups, $D(12) = 0.945, ns$; $D(21) = 0.929, ns$; and $D(24) = 0.942, ns$, respectively; (b) on the safety climate scale scores normality was achieved for the randomized-experimental, assigned-experimental, and control groups, $D(12) = 0.946, ns$; $D(21) = 0.958, ns$; and $D(20) = 0.973, ns$, respectively; (c) on the job satisfaction scale scores the randomized-experimental and control groups appeared to be normal, $D(12) = 0.927, ns$; and $D(24) = 0.955, ns$, respectively, and the assigned-experimental
group appeared non-normal, $D(21) = 0.873, p < 0.05$; (d) on the stress recognition climate scale scores the randomized-experimental, assigned-experimental, and control groups appeared normal, $D(12) = 0.968, ns; D(21) = 0.973, ns; \text{ and } D(24) = 0.950, ns$, respectively; (e) on the perceptions of management climate scale scores the randomized-experimental, assigned-experimental, and control groups were normal, $D(12) = 0.877, ns; D(21) = 0.941, ns; \text{ and } D(24) = 0.968, ns$, respectively; (f) and for the working conditions climate scale scores the randomized-experimental, assigned-experimental, and control groups were also normal, $D(12) = 0.956, ns; D(21) = 0.923, ns; \text{ and } D(24) = 0.952, ns$, respectively. The ICU SAQ data were normal based on the Shapiro-Wilk test with the exception of job satisfaction in the assigned-experimental group ($p = 0.11$).

**Analysis by Research Question**

*Group Equivalence at Baseline*

The Kruskal-Wallis test was used to assess for differences between the three independent groups (randomized-experimental, assigned-experimental, and control) at baseline on both the OR and the ICU SAQs. None of the climate scores was significantly different on the OR SAQ: teamwork climate scale score $H(1) = 0.001, ns$; safety climate score $H(1) = 1.167, ns$; stress recognition climate score $H(1) = 0.001, ns$; job satisfaction scale score $H(1) = 0.046, ns$; perception of management climate score $H(1) = 0.300, ns$; and stress recognition climate score $H(1) = 0.338, ns$.

Some of the climate scores were significantly different at baseline when assessing the three independent groups for the ICU SAQ (randomized-experimental, assigned-experimental, and control): teamwork climate scale score $H(1) = 6.683$,
$p < 0.05$; job satisfaction scale score $H(1) = 8.397$, $p < 0.05$; and working conditions scale score $H(1) = 6.794$, $p < 0.05$. The remainder of the climate scores were not significantly affected by the independent groups: safety climate score $H(1) = 1.167$, $ns$; stress recognition climate score $H(1) = 1.035$, $ns$; perception of management climate score $H(1) = 5.939$, $ns$.

When further testing was done to evaluate the randomized-experimental and control group the climate scores were statistically significantly different at baseline in the job satisfaction and working condition domains: job satisfaction score $H(1) = 4.38$, $p < 0.05$, $H(1) = 5.91$, $p < 0.05$. The teamwork climate scale scores were still near statistical significance with the following Kruskal-Wallis test results: $H(1) = 3.82$, $p = 0.051$. The combined-experimental (randomized-experimental and assigned-experimental) and control groups were not significantly different at baseline on the ICU SAQ: teamwork climate scale score $H(1) = 0.072$, $ns$; safety climate score $H(1) = 0.222$, $ns$; stress recognition climate score $H(1) = 0.001$, $ns$; job satisfaction scale score $H(1) = 0.017$, $ns$; perception of management climate score $H(1) = 0.860$, $ns$; and stress recognition climate score $H(1) = 1.274$, $ns$.

The Mann-Whitney U test was used to test the difference between means in the randomized-experimental (no assigned-experimental participants) and control groups at baseline in both the OR and ICU to determine whether the groups were equivalent. This was also done for the combined-experimental group (randomized-experimental and assigned-experimental) compared to the control group. Differences between the randomized-experimental (no assigned-experimental participants) and control groups were not statistically significant in either the OR or ICU for: (a) teamwork climate scale
scores, \( z = -0.021, p = 0.983; z = -1.954, p = 0.151 \); (b) safety climate scale scores, \( z = -0.104, p = 0.917; z = -0.68, p = 0.543 \); (c) stress recognition climate scale scores, \( z = -0.643, p = 0.520; z = -0.658, p = 0.510 \); and (d) perception of management climate scale scores, \( z = -0.753, p = 0.452; z = -0.711, p = 0.477 \). The difference in job satisfaction climate scale scores between groups for the OR was not significant \( (z = -0.146, p = 0.884) \), but the difference was significant for the ICU \( (z = -2.093, p = 0.036) \), with the control group scoring significantly higher at baseline. Similarly, working conditions climate scale scores in the OR were not significantly different \( (z = -0.690, p = 0.490) \), but a significant difference did appear for the scores of participants in the ICU \( (z = -2.433, p = 0.015) \), with the control group scoring significantly higher at baseline. Although the data were not normally distributed, the randomization process did result in equivalent groups in the randomized-experimental, assigned-experimental, and control groups for the OR SAQ; however, the ICU SAQ respondent groups were not equivalent at baseline despite randomization when the assigned-experimental participants were removed from the combined-experimental group.

Differences in the combined-experimental groups (both randomized and assigned) compared to the control groups were not statistically significant in either the OR or ICU: teamwork climate scale scores \( (z = -0.033, p = 0.973; z = -2.68, p = 0.789) \); safety climate scale scores \( (z = -1.08, p = 0.280; z = -0.472, p = 0.637) \); job satisfaction climate scale scores \( (z = -0.214, p = 0.830; z = -0.032, p = 0.974) \); stress recognition climate scale scores \( (z = -0.033, p = 0.973; z = 0.130, p = 0.897) \); perception of management climate scale scores \( (z = -0.548, p = 0.584; z = -0.927, p = 0.354) \); and working conditions climate scale scores \( (z = -0.582, p = 0.561; z = -1.129, p = 0.259) \).
In all of these instances, the data were not normally distributed but the randomization process did result in equivalent groups in the combined-experimental and control groups in both the OR and the ICU.

*Research Question Analysis*

After testing the properties of the distributions of scores, further analyses were conducted to answer the primary research questions. Two different methods were used. In each of the six domains on the SAQ, the first set of analyses reported are based on the group as the unit of analysis, as recommended by Sexton et al. (2006). The group method was based on a calculation of percentage of respondents reporting "good" climate scale scores on the each of the domains. The operational definition of a "good" score, indicating a positive safety climate, was a score that reached a threshold ≥ 75% (points) by Sexton et al. (J. Sexton & S. Grillo, personal communication, January 9, 2007) based on prior research (see Chapter III, *Calculation of Domain Scores*).

The second method of analysis used the individual scores for the particular domain as the unit of analysis. Repeated measures ANOVA were conducted for individual climate scale scores for all participants (randomized-experimental and assigned-experimental participants, compared with control); for the randomized-experimental and control participants; and for the combined-experimental (randomized plus assigned) and control participants. These ANOVA were used to evaluate the effects of group and time on the six domain climate scores. Although assumptions of normal distribution were not met for a number of the domains and groups, ANOVA is known to
be relatively robust in the face of nonparametric data and was determined to be an appropriate statistical measure (Fields, 2005).

Finally, if there was a significant change in the safety climate, the domain questions were separated and analyzed individually using a paired $t$ test as a form of post-hoc consideration of whether particular questions might be contributing more to the change. This approach was suggested by Sexton in personal communication (August 31, 2007). (Note: A summary table of the results of all analyses is provided in the Appendix.)

Operating Room Safety Attitude Questionnaire Analyses

*Teamwork Climate in the OR With Group as the Unit of Analysis*

Forty-six surveys were analyzed at baseline with 52% of all participants (experimental and control) reporting good teamwork climate ($\geq 75\%$) in the operating room. One week following simulation and debriefing exercises, 45 surveys were analyzed with 60% of respondents reporting a good teamwork climate; at 6 weeks 45 surveys were returned with 67% of the respondents reporting a good teamwork climate (Figure 2). According to the Sexton criteria, this change in safety climate scores from baseline to 6 weeks was statistically significant (i.e., it met the criterion of $>10\%$ difference in respondents reporting a good teamwork climate). It also moved the group out of the “danger zone.”

---

5 “Danger zone” is described as a threshold of risk below 60% of respondents with positive responses in a given domain indicating potential for substandard clinical and operational outcomes; see Chapter III.
In the combined-experimental group (randomized-experimental plus assigned-experimental), 54% of the respondents reported good teamwork climate at baseline, whereas 50% of respondents in the control group reported a good teamwork climate at baseline. According to Sexton's threshold test, this was not a significant difference. Both groups were in the "danger zone" (<60%) for teamwork climate at baseline. At 1 week, both the combined-experimental group and the control group results showed 60% of respondents reporting a good teamwork climate. According to the Sexton criteria, the change in the control group response was statistically significant at 1 week, but the combined-experimental group change did not meet Sexton's criterion for significant change from 1 week to 6 weeks because it changed less than 10 percentage points. At 6 weeks 65% of the respondents in the combined-experimental group reported a good teamwork climate in the operating room, and 68% of the control group reported a good
teamwork climate. According to Sexton’s criteria, this amount of change from baseline is statistically significant for the combined-experimental group (Figure 3).

In the randomized-experimental group (i.e., without the assigned-experimental participants) 64% of the respondents reported good teamwork climate at baseline and 50% of respondents in the control group reported a good teamwork climate at baseline. As noted previously, the control group results were in the danger zone (<60%) for teamwork climate at baseline. One week after the simulation, the percentage of the randomized-experimental group with responses in the safety zone actually fell from 64% to 45% (a 19% difference that met Sexton’s criterion for significant change), whereas responses from the control group increased from 50% to 60% (a 10% improvement).
The unexpected decline in the percentage of randomized-experimental group reporting attitudes in the safety zone resolved at 6 weeks with a return to 64% of the respondents reporting a good teamwork climate in the operating room. As noted previously, these percentages are based on randomized-experimental and control participants only who responded to the SAQ at three time points. Continued improvement in scores of the control group (60% to 68%) was demonstrated from 1 week to 6 weeks post, though not reaching the criterion for significance with only an 8% improvement from the 1 week survey. These results are graphically depicted in Figure 4.

Figure 4

Teamwork Climate Across Groups
OR SAQ – Randomized-Experimental and Control Only

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
Teamwork Climate in the OR With Individual as the Unit of Analysis

A repeated measures ANOVA was conducted to test the effects of Group (three levels) and Time (three levels) on teamwork climate scores using individual scaled scores as the unit of analysis. In this set of analyses, the three levels for Group were created by dividing the members of the combined-experimental group into two levels so that the scores for the randomized-experimental and assigned-experimental participants could be compared separately with the scores of the control group members. The ANOVA results for teamwork climate present a different picture from that shown by the prior analyses that used the percentage of a group above a safety threshold as the unit of analysis. The results of the repeated measures ANOVA were not statistically significant for Group, $F(2, 41) = 0.187, p = 0.830$, but were nearing significance for a main effect of change over Time $F(2, 82) = 2.943, p = 0.058$, $\eta^2_p = 0.067$. The interaction effect of Group $\times$ Time, $F(4, 82) = 1.34, p = 0.262$, also was not statistically significant in this analysis.

Similar evaluation methods were used to assess the effects of Group (two levels) and Time (three levels) on teamwork climate scores using individual scaled scores as the unit of analysis. In this set of analyses, the two levels for Group were created by taking the randomized-experimental participants as one group and the controls as the other. The results of the repeated measures ANOVA were not statistically significant for Group, $F(1, 28) = 0.004, p = 0.948$, or Time $F(2, 56) = 1.908, p = 0.158$. The interaction effect of Group $\times$ Time, $F(2, 56) = 0.691, p = 0.505$, also was not statistically significant in this analysis.
A third evaluation repeated measures ANOVA assessed the effect of Group (two levels) and Time (three levels) on teamwork climate scores. In this analysis, the two levels for Group were the combined-experimental group (created by combining the randomized-experimental plus assigned-experimental into a single group) and the control group. The results of the repeated measures ANOVA were not statistically significant for Group, $F(1, 42) = 0.096, p = 0.758$, but was statistically significant for Time $F(2, 84) = 3.704, p = 0.029, d = 0.01$. The interaction effect of Group $\times$ Time, $F(2, 84) = 0.355, p = 0.702$, also was not statistically significant in this analysis.

**Additional Analyses of Low OR Teamwork Climate Scores at Baseline**

The intention of safety intervention is to impact those who have low scores in the various domains. When analyzing those participants who did not report a good teamwork climate in the OR at baseline ($<75\%$ for SAQ teamwork climate) analyses using repeated measures ANOVA and paired $t$ tests were statistically significant. The repeated measures ANOVA examined the interaction of Group (randomized-experimental, assigned-experimental, or control); (combined-experimental and control); and (randomized-experimental and control) and Time. In the first analysis (randomized-experimental, assigned-experimental, and control) the results of the repeated measures ANOVA were not statistically significant for Group, $F(2, 18) = 0.672, p = 0.523$, but were statistically significant for Time $F(2, 36) = 8.92, p = 0.001, \eta_p^2 = 0.331$. The interaction effect of Group $\times$ Time, $F(4, 36) = 0.517, p = 0.724$, also was not statistically significant in this analysis.
For the second group (combined-experimental and control) the results of the repeated measures ANOVA were not statistically significant for Group, $F(1, 19) = 1.156, p = 0.296$, but were statistically significant for Time $F(2, 38) = 10.70, p = 0.000, d = 0.19$. The interaction effect of Group $\times$ Time, $F(2, 38) = 0.410, p = 0.666$, also was not statistically significant in this analysis. Results for the last group (randomized-experimental and control) were not statistically significant for Group, $F(1, 12) = 1.166, p = 0.301$, but was statistically significant for Time $F(2, 24) = 5.74, p = 0.009, d = 0.64$. The interaction effect of Group $\times$ Time, $F(2, 24) = 0.013, p = 0.987$, also was not statistically significant in this analysis.

When analyzing the teamwork climate scores of those not reporting a good teamwork climate at baseline using the paired $t$ test, on average, participants' mean scores increased and were statistically significant from baseline to 6 weeks ($M = 60.87, SE = 2.22; M = 69.50, SE = 3.40, t(21) = -3.42, p = 0.003, r = 0.67$).

*General Estimated Equation of Teamwork Climate Scores in OR*

Additional analyses were generated by the Quality Institute at the participating institution to evaluate the effect of simulation on individual survey scores over time. For these, focus was restricted to participants who were randomly assigned to either intervention or control (Figure 1). Separate analyses were conducted for the two types of surveys, OR and ICU. General estimating equations (GEE) were used to examine the binary outcomes of whether or not a survey score was at or above 75 (cutoff for "good" climate score). SAS version 9.00 (SAS Institute, Cary, NC) was used to performed these analyses. The basic structure of these analyses is an intercept and slope approach. The
regression coefficient (or parameter estimate) for the intervention (compared to control) group represents differences in the baseline scores. These need to be accounted for, but are not directly related to the study question regarding the effectiveness of the intervention.

The Time parameter represents the average change in survey scores over time regardless of whether the participant was in the intervention or control conditions. Similar to baseline differences, the Time trend is important to control for but provides no information on the intervention effect. Rather, the interaction term (intervention × time) is the major parameter of interest and represents the slope or change in scores over time in the intervention group versus change in scores in the control group. As shown in Table 10, the slope of the intervention was not statistically significant in the OR subpopulation whether the team scores were examined as a continuous or binary variable.

Table 10

*OR Effect of Simulation on Individual Teamwork Scores*

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>Pr &gt; (t)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>0.0328</td>
<td>0.4076</td>
<td>0.9359</td>
</tr>
<tr>
<td>Intervention</td>
<td>0.1608</td>
<td>0.5315</td>
<td>0.7622</td>
</tr>
<tr>
<td>Time</td>
<td>0.1704</td>
<td>0.0918</td>
<td>0.0634</td>
</tr>
<tr>
<td>Intervention × Time</td>
<td>−0.1220</td>
<td>0.1546</td>
<td>0.4300</td>
</tr>
</tbody>
</table>
Post-hoc Analyses on OR Teamwork Questions With Paired t Tests

Because the repeated measures ANOVA showed an effect for Time that was nearly significant when the combined-experimental group was split into the randomized-experimental and assigned-experimental subgroups, and it was significant when the two full groups (combined-experimental and control) were compared, and because there was statistical significance with the Sexton criteria, the teamwork climate domain questions were separated and analyzed individually from baseline to 6 weeks for all participants using a paired $t$ test. The teamwork climate domain detects the perceived quality of collaboration between personnel and includes the following statements:

1. It is easy for personnel in this OR to ask questions when there is something they do not understand.
2. I have the support I need from other personnel to care for our patients.
3. Nurse input is well received in this OR.
4. In this OR, it is difficult to speak up if I perceive a problem with patient care.
5. Disagreements in the OR are appropriately resolved.
6. The physicians and nurses here work as a well coordinated team.

On average, participants’ mean scores increased from baseline to 6 weeks when responding to the statement, “The physicians and nurses here work as a well coordinated team” ($M = 3.98, SE = 0.112; M = 4.24, SE = 0.101, t(44) = -2.74, p = 0.009$). When responding to “Nurse input about patient care is well received in the OR” ($M = 4.07, SE = 0.114; M = 4.27, SE = 0.101, t(44) = -2.74, p = 0.071$) scores increased as well at 6 weeks. This was not significant, however, when the Bonferroni correction was applied.
for running multiple $t$ tests, which requires the more conservative alpha level of $p = 0.008$.

_Safety Climate in the OR With Group as the Unit of Analysis_

Forty-six surveys were completed at baseline with 61% of all respondents (randomized-experimental, assigned-experimental, and control) reporting a good safety climate ($\geq 75\%$) in the operating room. This decreased to 51% at 1 week, a change which persisted at 6 weeks (Figure 5). According to the Sexton criteria, this decrease is statistically significant (10% difference with $>60\%$ positive response rate).

_Figure 5_

Safety Climate Across Groups
OR SAQ – All Participants

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
In the combined-experimental group 58% of the respondents reported good safety climate at baseline compared to 65% of respondents in the control group. This indicates that the combined-experimental group was in the danger zone (<60%) for safety climate at baseline. Responses changed minimally at 1 week in the combined-experimental group (56%), whereas positive responses in the control group fell to 45%. This represents a significant decline, which persisted at 6 weeks. The combined-experimental group responses were essentially unchanged between the 1 week and 6 week surveys (56% vs. 54%) (Figure 6).

Figure 6

Safety Climate Across Groups
OR SAQ – By Experimental Condition

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
In the randomized-experimental group, 82% of the respondents reported good safety climate at baseline, compared to 65% of respondents in the control group. A strongly negative response was exhibited in the 1 week responses, which decreased to 45% for both the randomized-experimental and control groups. Some recovery occurred at 6 weeks for the randomized-experimental group (73%), whereas the control group remained diminished at only 47% reporting a good safety climate (Figure 7).

Figure 7

Safety Climate Across Groups
OR SAQ – Randomized-Experimental and Control Only

<table>
<thead>
<tr>
<th>% Respondents Reporting Good Safety Climate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
</tr>
<tr>
<td>Randomized-Experimental</td>
</tr>
<tr>
<td>Control</td>
</tr>
<tr>
<td>1 Week</td>
</tr>
<tr>
<td>Randomized-Experimental</td>
</tr>
<tr>
<td>Control</td>
</tr>
<tr>
<td>6 Week</td>
</tr>
<tr>
<td>Randomized-Experimental</td>
</tr>
<tr>
<td>Control</td>
</tr>
</tbody>
</table>

Safety Climate in the OR With Individual as the Unit of Analysis

A repeated measures ANOVA was conducted to test the effects of Group (three levels) and Time (three levels) on safety climate scores using individual scaled scores as
the unit of analysis. In this set of analyses, the three levels for Group were created by dividing the members of the combined-experimental group into two levels so that the scores for the randomized-experimental and assigned-experimental participants could be compared separately with the scores of the control group members. The ANOVA results for safety climate present a different picture from that shown by the prior analyses that used the percentage of a group above a safety threshold as the unit of analysis. The results of the repeated measures ANOVA were not statistically significant for Group, $F(2, 41) = 0.613, p = 0.547$, or Time $F(2, 82) = 1.04, p = 0.357$. The interaction effect of Group x Time, $F(4, 82) = 2.25, p = 0.070$, also was not statistically significant in this analysis.

Similar evaluation methods were used to assess the effects of Group (two levels) and Time (three levels) on safety climate scores using individual scaled scores as the unit of analysis. In this set of analyses, the two levels for Group were created by taking the randomized-experimental participants as one group and the controls as the other. The results of the repeated measures ANOVA were not statistically significant for group, $F(1, 28) = 0.004, p = 0.949$, but were statistically significant for Time $F(1.654, 46.32) = 4.099, p = 0.030, d = 0.08$. The interaction effect of Group x Time, $F(1.654, 46.32) = 1.79, p = 0.183$, also was not statistically significant in this analysis.

A third evaluation with repeated measures ANOVA assessed the effect of Group (two levels) and Time (three levels) on safety climate scores. In this analysis, the two levels for Group were the combined-experimental group (randomized-experimental plus assigned-experimental) and the control group. The results of the repeated measures ANOVA were not statistically significant for Group, $F(1, 42) = 0.500, p = 0.483$, or
Time $F(1.73, 72.74) = 0.610, p = 0.524$. The interaction effect of Group $\times$ Time, $F(1.73, 72.74) = 0.097, p = 0.882$, also was not statistically significant in this analysis.

**Post-hoc Analyses on ICU Safety Climate Questions With Paired t Test**

Because the repeated measures ANOVA was significant for Time ($p = 0.030$) in the analysis for the randomized-experimental and control groups, the climate domain questions were analyzed separately from baseline to 6 weeks using paired $t$ tests. The safety climate domain on the SAQ reveals perceptions of a strong and proactive organizational commitment to patient safety and includes the following statements:

1. The culture in this OR makes it easy to learn from errors of others.
2. Medical errors are handled appropriately in this OR.
3. I know the proper channels to direct questions regarding patient safety.
4. I am encouraged by my colleagues to report any patient safety concerns.
5. I receive appropriate feedback about my performance.
6. I would feel safe being treated here as a patient.
7. In this OR it is difficult to discuss errors.

Participants’ mean scores on safety climate questions did not change significantly from baseline to 6 weeks or 1 week to 6 weeks in the randomized-experimental or control groups. When comparing scores for all participants from 1 week to 6 weeks, participants’ mean scores from baseline to 6 weeks decreased when responding to the statement, “Medical errors are handled appropriately in this hospital” ($M = 4.23, SE = 0.121$; $M = 4.00, SE = 0.130$, $t(43) = -2.121, p = 0.040$). This was not
significant, however, when the Bonferroni correction was applied for running multiple t-tests, which requires the more conservative alpha level of \( p = 0.007 \).

Job Satisfaction in the OR With Group as the Unit of Analysis

Forty-six surveys were analyzed at baseline for all participants (all experimental, and control), with 76% reporting good job satisfaction in the operating room. This increased slightly at 1 week, with 80% reporting good job satisfaction. This improvement was lost by 6 weeks with only 69% of the respondents reporting good job satisfaction (Figure 8). According to Sexton's criteria, these changes were not statistically significant and did not represent decline into the danger zone.

Figure 8

Job Satisfaction Across Groups
OR SAQ – All Participants

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
In the combined-experimental group 77% of the respondents reported good job satisfaction at baseline compared to 75% of respondents in the control group. At 1 week the percentage of combined-experimental group responding in the safety zone improved slightly (from 77% to 80%), but then this proportion fell to 69% at 6 weeks, which represented a significant change between 1 week and 6 weeks, but not between baseline and 6 weeks. Similarly, changes in the percentage of control group members responding in the safety zone were not significant from baseline to 6 weeks (Figure 9).

Figure 9

Job Satisfaction Across Groups
OR SAQ – By Experimental Condition

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
In the randomized-experimental group, 82% of the respondents reported good job satisfaction at baseline and at 1 week, but this decreased to 64% at 6 weeks. This represented a significant change, but did not place the group in the danger zone.

Control group responses were not significantly different between the baseline and 6 week surveys. There was a decrease in the percentage of the group reporting positive responses at 6 weeks, but the percentage of control group with positive responses remained above the danger zone as well (Figure 10).

**Figure 10**

Job Satisfaction Across Groups
OR SAQ – Randomized-Experimental and Control Only

![Bar chart showing job satisfaction across groups.]

*Job Satisfaction in the OR With Individual as the Unit of Analysis*

A repeated measures ANOVA was conducted to test the effects of Group (three levels) and Time (three levels) on job satisfaction scores using individual scaled scores...
as the unit of analysis. In this set of analyses, the three levels for Group were created by dividing the members of the experimental group into two levels so that the scores for the randomized-experimental and assigned-experimental participants could be compared separately with the scores of the control group members. The ANOVA results for job satisfaction presented a different picture from that shown by the prior analyses that used the percentage of a group above a safety threshold as the unit of analysis. The results of the repeated measures ANOVA were not statistically significant for Group, $F(2, 41) = 0.196, p = 0.822$, or Time $F(2, 82) = 0.056, p = 0.946$. The interaction effect of Group $\times$ Time, $F(4, 82) = 0.477, p = 0.752$, also was not statistically significant in this analysis.

Similar evaluation methods were used to assess the effects of Group (two levels) and Time (three levels) on job satisfaction scores using individual scaled scores as the unit of analysis. In this set of analyses, the two levels for Group were created by taking the randomized-experimental participants as one group and the controls as the other. The results of the repeated measures ANOVA were not statistically significant for Group, $F(1, 28) = 0.443, p = 0.511$, or Time $F(2, 56) = 0.054, p = 0.948$. The interaction effect of Group $\times$ Time, $F(2, 56) = 0.445, p = 0.643$, also was not statistically significant in this analysis.

A third evaluation repeated measures ANOVA assessed the effect of Group (two levels) and Time (three levels) on job satisfaction scores. In this analysis, the two levels for Group were the combined-experimental group, which was created by combining the randomized-experimental plus assigned-experimental into a single group and the control group. The results of the repeated measures ANOVA were not statistically significant for Group, $F(1, 42) = 0.324, p = 0.572$, or Time $F(2, 84) = 0.787, p = 0.459$. The
interaction effect of Group x Time, $F(2, 84) = 0.787, p = 0.459$, also was not statistically significant in this analysis.

**Stress Recognition in the OR With Group as the Unit of Analysis**

Forty-six surveys were completed at baseline for all participants (randomized experimental, assigned-experimental, and control), with only 37% of respondents fully acknowledging the effects of stress in the operating room. At 1 week post simulation this improved insignificantly to 40%, but this slight improvement was not maintained at 6 weeks (Figure 11). According to Sexton’s criteria, the OR is in the danger zone for stress recognition with fewer than 60% of respondents fully acknowledging the effects of stress.

A similar pattern was seen in comparing the combined-experimental group to the control group, with a danger zone response rate throughout the study (<60% of respondents fully acknowledged the effects of stress). A slight improvement occurred in the combined-experimental group 1 week following simulation (31% to 36%), but even this slight improvement was lost by 6 weeks (Figure 12).

A more substantial improvement was realized in the randomized-experimental group due to the very low acknowledgment of the affects of stress at baseline in this group (18%). Following simulation, a substantial improvement occurred to 36% at 1 week with continued improvement at 6 weeks to 45%. Though the group remained in the danger zone regarding recognition of the effects of stress, the improvement was highly significant (Figure 13).
Figure 11
Stress Recognition Across Groups
OR SAQ – All Participants

Figure 12
Stress Recognition Across Groups
OR SAQ – By Experimental Condition
A repeated measures ANOVA was conducted to test the effects of Group (three levels) and Time (three levels) on stress recognition scores using individual scaled scores as the unit of analysis. In this set of analyses, the three levels for Group were created by dividing the members of the combined-experimental group into two levels so that the scores for the randomized-experimental and assigned-experimental participants could be compared separately with the scores of the control group members. The ANOVA results for stress recognition presented a different picture from that shown by the prior analyses that used the percentage of a group above a safety threshold as the unit of analysis. The results of the repeated measures ANOVA were not statistically
significant for Group, \(F(2, 41) = 0.138, p = 0.871\), or Time \(F(2, 82) = 0.273, p = 0.762\). The interaction effect of Group \(\times\) Time, \(F(4, 82) = 0.441, p = 0.775\), also was not statistically significant in this analysis.

Similar evaluation methods were used to assess the effects of Group (two levels) and Time (three levels) on stress recognition scores using individual scaled scores as the unit of analysis. In this set of analyses, the two levels for Group were created by taking the randomized-experimental participants as one group and the controls as the other. The results of the repeated measures ANOVA were not statistically significant for Group, \(F(1, 28) = 0.052, p = 0.821\), or Time \(F(2, 56) = 0.013, p = 0.987\). The interaction effect of Group \(\times\) Time, \(F(2, 56) = 0.208, p = 0.812\), also was not statistically significant in this analysis.

A third evaluation repeated measures ANOVA assessed the effect of Group (two levels) and Time (three levels) on stress recognition scores. In this analysis, the two levels for Group were the combined-experimental group and the control group. The results of the repeated measures ANOVA were not statistically significant for Group, \(F(1, 42) = 0.004, p = 0.947\), or Time \(F(2, 84) = 0.333, p = 0.717\). The interaction effect of Group \(\times\) Time, \(F(2, 84) = 0.086, p = 0.918\), also was not statistically significant in this analysis.

Post-hoc Analyses on OR Stress Recognition Questions With Paired t Tests

Because the randomized-experimental group significantly improved at 1 week according to the Sexton criteria, the climate domain questions were separated and analyzed individually from baseline to 1 week for control participants using a paired \(t\)
test. The stress recognition items acknowledge how the participants feel their performance is influenced by stressors and include the following statements:

1. When my workload becomes excessive, my performance is impaired.
2. I am more likely to make errors in hostile or tense situations.
3. Fatigue impairs my performance during emergency situations.
4. I am less effective at work when fatigued.

Participants' mean scores on stress recognition questions did not change significantly from baseline to 1 week in the randomized-experimental group.

Perception of Management in the OR With Group as the Unit of Analysis

Forty-six surveys were analyzed at baseline for all participants (randomized-experimental, assigned-experimental, and control) with 52% reporting a positive perception of management in the operating room. At 1 week 56% of respondents reported a positive perception of management, and at 6 weeks 48% of the respondents reported a positive perception of management (Figure 14), placing the OR in the danger zone for perception of management (<60% positive response rate).

In the combined-experimental group 50% of the respondents reported a positive perception of management at baseline compared to 55% of respondents in the control group. Little change occurred in these response rates following simulation. All grouped response rates remained in the danger zone (<60%) throughout the study (Figure 15).

In the randomized-experimental group, a negative effect was demonstrated regarding perception of management with a decline of 45% to 36% at 1 week, which
Figure 14
Perception of Management Across Groups
OR SAQ – All Participants

Figure 15
Perception of Management Across Groups
OR SAQ – By Experimental Condition

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
persisted at 6 weeks (Figure 16). This change approached statistical significance according to Sexton's criteria.

Figure 16

Perception of Management Across Groups
OR SAQ – Randomized-Experimental and Control Only

![Bar chart showing perception of management across groups over time]

Perception of Management in the OR With Individual as the Unit of Analysis

A repeated measures ANOVA was conducted to test the effects of Group (three levels) and Time (three levels) on perception of management scores using individual scaled scores as the unit of analysis. In this set of analyses, the three levels for Group were created by dividing the members of the combined-experimental group into two levels (randomized and assigned) for comparison separately with the scores of the control group members. The ANOVA results for perception of managements presented a different picture from that shown by the prior analyses that used the percentage of a
group above a safety threshold as the unit of analysis. The results of the repeated measures ANOVA were not statistically significant for Group, \( F(2, 41) = 2.319, p = 0.111 \), or Time \( F(2, 82) = 0.132, p = 0.877 \). The interaction effect of Group \( \times \) Time, \( F(4, 82) = 3.46, p = 0.012, \eta_p^2 = 0.144 \), was statistically significant in this analysis.

Similar evaluation methods were used to assess the effects of Group (two levels) and Time (three levels) on perception of management scores using individual scaled scores as the unit of analysis. In this set of analyses, the two levels for Group were created by taking the randomized-experimental participants as one group and the controls as the other. The results of the repeated measures ANOVA were not statistically significant for Group, \( F(1, 28) = 3.69, p = 0.065 \), or Time \( F(2, 56) = 1.908, p = 0.158 \). The interaction effect of Group \( \times \) Time, \( F(2, 56) = 2.763, p = 0.072 \), also was not statistically significant in this analysis.

A third evaluation repeated measures ANOVA assessed the effect of Group (two levels) and Time (three levels) on perception of management scores. In this analysis, the two levels for Group were the combined-experimental group and the control group. The results of the repeated measures ANOVA were not statistically significant for Group, \( F(1, 42) = 0.556, p = 0.460 \), or Time \( F(2, 84) = 0.111, p = 0.895 \). The interaction effect of Group \( \times \) Time, \( F(2, 84) = 0.211, p = 0.810 \), also was not statistically significant in this analysis.

**Post-hoc Analyses on Perception of Management Climate Questions With Paired t Test**

Because the results of the repeated design ANOVA were statistically significant for Group \( \times \) Time, and the decrease in scores for the randomized group approached
statistical significance at 1 week by Sexton’s criteria, the climate domain questions were separately analyzed from baseline to 1 week using a paired \( t \) test. The perception of management statements indicate participant approval of managerial actions and include the following statements:

1. Hospital management does not knowingly compromise the safety of patients.
2. Hospital administration supports my daily efforts.
3. I am provided with adequate, timely information about hospital events.
4. The levels of staffing in this OR are sufficient to handle the number of patients.

The Bonferroni correction was applied for running multiple \( t \) tests, which requires the more conservative alpha level of \( p = 0.012 \). When comparing groups from baseline to 1 week, on average, participants’ overall mean scores increased statistically significantly in the assigned-experimental group (\( M = 70.54, SE = 4.88; M = 79.01, SE = 3.96, t (13) = -3.18, p = 0.007, r = 0.84 \)). The assigned-experimental participants’ mean scores, on average, increased and were nearing statistical significance when asked, “Hospital administration supports my daily efforts” (\( M = 3.79, SE = 0.300; M = 4.14, SE = 0.253, t(13) = 2.69, p = 0.019 \)).

When comparing groups from 1 week to 6 weeks, on average, participants’ overall mean scores decreased, which represented statistically significant change in the assigned-experimental group (\( M = 79.08, SE = 3.96; M = 71.88, SE = 5.15, t(13) = 2.93, p = 0.012, r = 0.89 \)). When comparing 1 week to 6 weeks on average, the randomized-experimental participants’ mean scores increased (\( M = 2.73, SE = 0.332; M = 3.64, SE = 0.279, t(10) = -2.47, p = 0.033 \)) when responding to the statement “I am provided with
adequate, timely information about events in the hospital that affect my work.” When asked to respond to the statement, “The levels of staffing in our ORs are sufficient to handle the number of patients,” the assigned-experimental group mean scores decreased, from 1 week to 6 weeks ($M = 4.29, SE = 0.125; M = 3.86, SE = 0.254, t(13) = 2.121, p = 0.054$). Control group mean scores decreased, nearing statistical significance when responding to the statement, “Hospital management does not knowingly compromise the safety of patients” ($M = 4.37, SE = 0.244; M = 4.00, SE = -0.276, t (18) = 2.67, p = 0.015$). When comparing baseline to 6 weeks, no statistical significance was found in any of the groups.

**Working Conditions in the OR With Group as the Unit of Analysis**

Forty-six surveys were analyzed at baseline for all participants (randomized-experimental, assigned-experimental, and control) with 48% reporting positive working conditions in the operating room. A significant improvement occurred at 1 week with 60% of respondents reporting positive working conditions, but at 6 weeks only 38% of the respondents still reported positive working conditions (Figure 17). According to the Sexton criteria, the OR is in the danger zone for working conditions.

In the combined-experimental group 46% of the respondents reported positive working conditions at baseline, compared to 50% of respondents in the control group. At 1 week the combined-experimental group experienced a substantial improvement to 72% of respondents reporting positive working conditions, but this improvement was not maintained at 6 weeks, dropping below baseline responses for the group. The
control group responses demonstrated a steady, significant decline throughout the study, ending at 37% (Figure 18).

In the randomized-experimental group 45% of the respondents reported positive working conditions at baseline. This improved to 73% at 1 week, but then fell significantly below baseline to 27% at 6 weeks (Figure 19).

**Working Conditions in the OR With Individual as the Unit of Analysis**

A repeated measures ANOVA was conducted to test the effects of Group (three levels) and Time (three levels) on working conditions scores using individual scaled scores as the unit of analysis. In this set of analyses, the three levels for Group were created by dividing the members of the combined-experimental group into two levels so
Figure 18

Working Conditions Across Groups
OR SAQ – By Experimental Condition

<table>
<thead>
<tr>
<th></th>
<th>Combined-Experimental</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td>46</td>
<td>72</td>
</tr>
<tr>
<td>1 Week</td>
<td>50</td>
<td>45</td>
</tr>
<tr>
<td>6 Week</td>
<td>38</td>
<td>37</td>
</tr>
</tbody>
</table>

Figure 19

Working Conditions Across Groups
OR SAQ – Randomized Experimental and Control Only

<table>
<thead>
<tr>
<th></th>
<th>Randomized-Experimental</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td>45</td>
<td>73</td>
</tr>
<tr>
<td>1 Week</td>
<td>50</td>
<td>45</td>
</tr>
<tr>
<td>6 Week</td>
<td>27</td>
<td>37</td>
</tr>
</tbody>
</table>

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
that the scores for the randomized-experimental and assigned-experimental participants could be compared separately with the scores of the control group members. The ANOVA results for working conditions presented a different picture from that shown by the prior analyses that used the percentage of a group above a safety threshold as the unit of analysis. The results of the repeated measures ANOVA were not statistically significant for Group, $F(2, 41) = 0.438, p = 0.648$, or Time $F(2, 82) = 0.508, p = 0.604$. The interaction effect of Group $\times$ Time, $F(4, 82) = 0.815, p = 0.519$, also was not statistically significant in this analysis.

Similar evaluation methods were used to assess the effects of Group (two levels) and Time (three levels) on working conditions scores using individual scaled scores as the unit of analysis. In this set of analyses, the two levels for Group were created by taking the randomized-experimental as one group and the controls as the other. The results of the repeated measures ANOVA were not statistically significant for Group, $F(1, 28) = 0.017, p = 0.898$, or Time $F(2, 56) = 0.618, p = 0.543$. The interaction effect of Group $\times$ Time, $F(2, 56) = 0.732, p = 0.485$, also was not statistically significant in this analysis.

A third evaluation repeated measures ANOVA assessed the effect of Group (two levels) and Time (three levels) on working condition scores. In this analysis, the two levels for Group were the combined-experimental group and the control group. The results of the repeated measures ANOVA were not statistically significant for Group, $F(1, 42) = 0.442, p = 0.510$, or Time $F(2, 84) = 0.226, p = 0.798$. The interaction effect of Group $\times$ Time, $F(2, 84) = 0.385, p = 0.682$, also was not statistically significant in this analysis.
Post-hoc Analyses on OR Working Conditions Climate Questions With Paired t Test

Because the combined-experimental group’s increase in good working condition scores at 1 week was significant according to the Sexton criteria, the climate domain questions were separately analyzed from baseline to 1 week using a paired t test. The working condition domain examines the perceived quality of work environment and logistical support and includes the following statements:

1. All the necessary information is available before the start of a procedure.
2. This hospital constructively deals with problem physicians and employees.
3. Trainees in my discipline are adequately supervised.
4. This hospital does a good job of training new personnel.

The Bonferroni correction was applied for running multiple t tests, which requires the more conservative alpha level of $p = 0.012$. In comparing groups from baseline to 1 week, when responding to the statement, “All the necessary information is available before the start of a procedure,” on average the assigned-experimental participants’ mean scores increased and were nearing statistical significance ($M = 3.79, SE = 0.300; M = 4.43, SE = 0.173, t(13) = -2.857, p = 0.013$).

Intensive Care Unit Safety Attitude Questionnaire Analyses

Teamwork Climate (ICU SAQ) With Group as the Unit of Analysis

Fifty-seven surveys were analyzed at baseline for all participants (randomized-experimental, assigned-experimental, and control) with 51% reporting good teamwork climate in the intensive care unit. This remained relatively flat at 1 week and 6 weeks (53% and 49%, respectively) (Figure 20).
In the combined-experimental group 56% of the respondents reported good teamwork climate at baseline compared to 46% of respondents in the control group. Both groups were in the danger zone (<60%) for teamwork climate at baseline and remained essentially unchanged with the exception of the control group, which dropped to 35% at 6 weeks (Figure 21).

In the randomized-experimental group, 25% of the respondents reported good teamwork climate at baseline with no change at 1 week, followed by a significant improvement at 6 weeks to 50%. The control group demonstrated decline in positive response rates from 46% to 45% to 35% at baseline, 1 week and 6 weeks, respectively. Despite the significant improvement by 6 weeks in the randomized-experimental group,
both randomized-experimental and control groups remained in the danger zone regarding good teamwork climate as measured by the ICU SAQ (Figure 22).

*Teamwork Climate (ICU SAQ) With Individual as the Unit of Analysis*

A repeated measures ANOVA was conducted to test the effects of Group (three levels) and Time (three levels) on teamwork climate scores using individual scaled scores as the unit of analysis. In this set of analyses, the three levels for Group were created by dividing the members of the experimental group into two levels so that the scores for the randomized-experimental and assigned-experimental participants could be
compared separately with the scores of the control group members. The ANOVA results for teamwork climate present a different picture from that shown by the prior analyses.

Figure 22

Teamwork Climate Across Groups
ICU SAQ – Randomized-Experimental and Control Only

![Bar graph showing teamwork climate across groups.](image)

that used the percentage of a group above a safety threshold as the unit of analysis. The results of the repeated measures ANOVA were statistically significant for Group, $F(2, 46) = 3.92, p = 0.027, \eta^2_p = 0.146$, but not for Time $F(1.77, 81.28) = 0.003, p = 0.995$. The interaction effect of Group x Time, $F(3.5, 81.28) = 1.31, p = 0.275$, was not statistically significant in this analysis.

Similar evaluation methods were used to assess the effects of Group (two levels) and Time (three levels) on teamwork climate scores using individual scaled scores as the unit of analysis. In this set of analyses, the two levels for Group were created by taking

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
the randomized-experimental participants as one group and the controls as the other. The
results of the repeated measures ANOVA were not statistically significant for Group,
\(F(1, 28) = 1.35, p = 0.256\), or Time \(F(2, 56) = 0.184, p = 0.833\). The interaction effect
of Group x Time, \(F(2, 56) = 2.08, p = 0.134\), also was not statistically significant in this
analysis.

A third evaluation repeated measures ANOVA assessed the effect of Group (two
levels) and Time (three levels) on teamwork climate scores. In this analysis, the two
levels for Group were the combined-experimental group and the control group. The
results of the repeated measures ANOVA were not statistically significant for Group,
\(F(1, 47) = 0.347, p = 0.559\), or Time \(F(2, 94) = 0.353, p = 0.703\). The interaction effect
of Group x Time, \(F(2, 94) = 0.757, p = 0.472\), also was not statistically significant in
this analysis.

Addional Analyses of Low Teamwork Climate Scores at Baseline

When analyzing those participants who did not report a good teamwork climate
in the ICU at baseline (<75% for SAQ teamwork climate) analyses using repeated
measures ANOVA and paired t tests were statistically significant. The repeated
measures ANOVA examined the interaction of Group (randomized-experimental,
assigned-experimental, or control); (combined-experimental and control); and
(randomized-experimental and control) and Time. In the first analysis (randomized-
experimental, assigned-experimental, and control) the results of the repeated measures
ANOVA were not statistically significant for Group, \(F(2, 21) = 0.1.26, p = 0.304\), but
were statistically significant for Time \(F(2, 42) = 6.39, p = 0.004, \eta^2_p = 0.233\). The
interaction effect of Group × Time, $F(4, 42) = 1.88, p = 0.131$, also was not statistically significant in this analysis.

For the second group (combined-experimental and control) the results of the repeated measures ANOVA were not statistically significant for Group, $F(1, 22) = 0.094, p = 0.763$, but were statistically significant for Time $F(2, 44) = 5.11, p = 0.010, d = 0.2$. The interaction effect of Group × Time, $F(2, 44) = 1.34, p = 0.273$, was not statistically significant in this analysis. Results for the last group (randomized-experimental and control) were not statistically significant for Group, $F(1, 17) = 1.075, p = 0.314$, or for Time $F(2, 34) = 2.856, p = 0.071$. The interaction effect of Group × Time, $F(2, 34) = 2.351, p = 0.111$, also was not statistically significant in this analysis.

When analyzing the teamwork climate scores of those not reporting a good teamwork climate at baseline using the paired $t$ test, on average, participants' mean scores increased and were statistically significant from baseline to 6 weeks ($M = 58.64, SE = 1.93; M = 65.59, SE = 2.77, t(26) = -2.55, p = 0.017, r = 0.37$).

*General Estimated Equation of Teamwork Climate Scores in ICU*

Additional analyses were generated by the Quality Institute at the participating institution to evaluate the effect of simulation on individual survey scores over time. For these, focus was restricted to participants who were randomly assigned to either intervention or control. Separate analyses were conducted for the two types of surveys, OR and ICU. General estimating equations (GEE) were used to examine the binary outcomes of whether or not a survey score was at or above 75 (cutoff for "good" climate score). SAS version 9.00 (SAS, Cary NC) was used to performed these analyses. The
basic structure of these analyses is an intercept and slope approach. The regression coefficient (or parameter estimate) for the intervention (compared to control) represents differences in the baseline scores. These need to be accounted for, but are not directly related to the study question regarding the effectiveness of the intervention.

The Time parameter represents the average change in survey scores over time regardless of whether the participant was in the intervention or control. Similar to baseline differences, the Time trend is important to control for but provides no information on the intervention effect. Rather, the interaction term (intervention x time) is the major parameter of interest and represents the slope or change in scores over time in the intervention group versus change in scores in the control group. The intervention slope was statistically significant in the ICU (Table 11) (M. Cowen, personal communication, September 13, 2007).

Table 11

**ICU Effect of Simulation on Individual Teamwork Scores**

<table>
<thead>
<tr>
<th>Effect</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>Pr&gt;(t)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>68.61</td>
<td>4.25</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Intervention</td>
<td>-7.25</td>
<td>5.98</td>
<td>0.2357</td>
</tr>
<tr>
<td>Time</td>
<td>-0.49</td>
<td>0.68</td>
<td>0.4699</td>
</tr>
<tr>
<td>Intervention x Time</td>
<td>-2.04</td>
<td>0.94</td>
<td>0.0359</td>
</tr>
</tbody>
</table>

*Post-hoc Analyses on ICU Teamwork Climate Questions With Paired t Test*

Because teamwork climate scores increased significantly at 6 weeks per the Sexton criteria, the teamwork climate domain questions were separately analyzed from
baseline to 6 weeks for the randomized-experimental, assigned-experimental, and control groups using a paired $t$ test. When comparing groups from baseline to 6 weeks, on average, the increase in teamwork climate scores from baseline to 6 weeks did not reach statistical significance with the Bonferroni correction of $p = 0.008 \ (M = 60.76, SE = 4.6; M = 70.5, SE = 3.9, t(11) = -2.21, p = 0.049)$. From baseline to 6 weeks when responding to the statement, "The physicians and nurses here work together as a well coordinated team," the randomized-experimental participants’ mean scores increased and were statistically significant ($M = 2.83, SE = 0.322; M = 4.17, SE = 0.167, t(11) = -3.752, p = 0.003, r = 0.047) . When participants were asked to respond to the statement, "It is easy for personnel in this ICU to ask questions when there is something that they do not understand," the randomized participants’ mean scores did not reach statistical significance ($M = 3.75, SE = 0.279; M = 4.17, SE = 0.241, t(11) = -2.159, p = 0.054) .

Safety Climate (ICU SAQ) With Group as the Unit of Analysis

Fifty-seven surveys were analyzed at baseline for all participants (randomized-experimental, assigned-experimental, and control); 54% reported good safety climate in the intensive care unit. At 1 week 47% of respondents reported a good safety climate, and at 6 weeks 55% of the respondents reported a good safety climate (Figure 23).

In the combined-experimental group 58% of the respondents reported good safety climate at baseline followed by a significant decline to 48% at 1 week. This recovered at 6 weeks (59%). The control group scores regarding safety climate did not
change significantly over time (Figure 24). Both combined-experimental and control groups remained in the danger zone (<60%) for safety climate at all points measured.

In the randomized-experimental group, 67% of the respondents reported good safety climate at baseline. This response rate dropped significantly to 36% at 1 week, and then partially recovered to 58% at 6 weeks (Figure 25). Control group responses showed no significant change (50% to 45% to 48% at baseline, 1 week and 6 weeks, respectively).

Safety Climate (ICU SAQ) With Individual as the Unit of Analysis

A repeated measures ANOVA was conducted to test the effects of Group (three levels) and Time (three levels) on safety climate scores using individual scaled scores as
Figure 24
Safety Climate Across Groups
ICU SAQ – By Experimental Condition

<table>
<thead>
<tr>
<th></th>
<th>Combined-Experimental</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td>58%</td>
<td>50%</td>
</tr>
<tr>
<td>1 Week</td>
<td>48%</td>
<td>45%</td>
</tr>
<tr>
<td>6 Week</td>
<td>59%</td>
<td>48%</td>
</tr>
</tbody>
</table>

Figure 25
Safety Climate Across Groups
ICU SAQ – Randomized-Experimental and Control Only

<table>
<thead>
<tr>
<th></th>
<th>Randomized-Experimental</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td>67%</td>
<td>50%</td>
</tr>
<tr>
<td>1 Week</td>
<td>36%</td>
<td>45%</td>
</tr>
<tr>
<td>6 Week</td>
<td>58%</td>
<td>48%</td>
</tr>
</tbody>
</table>

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
the unit of analysis. In this set of analyses, the three levels for Group were created by dividing the members of the experimental group into two levels so that the scores for the randomized-experimental and assigned-experimental participants could be compared separately with the scores of the control group members. The ANOVA results for safety climate presented a different picture from that shown by the prior analyses that used the percentage of a group above a safety threshold as the unit of analysis. The results of the repeated measures ANOVA were not statistically significant for Group, $F(2, 46) = 1.91$, $p = 0.160$, or Time $F(1.78, 81.684) = 0.00, p = 0.999$. The interaction effect of Group x Time, $F(3.55, 81.68) = 1.74, p = 0.155$, also was not statistically significant in this analysis.

Similar evaluation methods were used to assess the effects of Group (two levels) and Time (three levels) on safety climate scores using individual scaled scores as the unit of analysis. In this set of analyses, the two levels for Group were created by taking the randomized-experimental participants as one group and the controls as the other. The results of the repeated measures ANOVA were not statistically significant for Group, $F(1, 28) = 0.203, p = 0.656$, but were not statistically significant for Time $F(2, 56) = 1.47, p = 0.238$. The interaction effect of Group x Time, $F(12, 56) = 0.194, p = 0.824$, also was not statistically significant in this analysis.

A third evaluation repeated measures ANOVA assessed the effect of Group (two levels) and Time (three levels) on safety climate scores. In this analysis, the two levels for Group were the combined-experimental group and the control group. The results of the repeated measures ANOVA were not statistically significant for Group, $F(1, 47) = $
01.344, \( p = 0.252 \), or Time \( F(2, 94) = 1.33, p = 0.269 \). The interaction effect of Group \( \times \) Time, \( F(2, 94) = 0.401, p = 0.671 \), also was not statistically significant in this analysis.

Post-hoc Analyses on Safety Climate Questions With Paired \( t \) Test

Safety climate scores decreased significantly at 1 week per the Sexton criteria. The safety climate domain questions were separately analyzed from baseline to 1 week for the randomized-experimental group using a paired \( t \) test. When comparing groups from baseline to 1 week, on average, the decrease in safety climate scores from baseline to 1 week were not statistically significant.

Job Satisfaction (ICU SAQ) With the Group as the Unit of Analysis

Fifty-seven surveys were analyzed at baseline for all participants (randomized-experimental, assigned-experimental, and control) with 54\% reporting good job satisfaction in the intensive care unit. At 1 week 53\% of respondents reported good job satisfaction, and at 6 weeks 58\% of the respondents reported good job satisfaction (Figure 26).

In the combined-experimental group 55\% of the respondents reported good safety climate at baseline followed by a slight improvement at 1 week to 61\%, but this returned essentially to baseline (56\%) at 6 weeks. Control participants exhibited a significant decline in scores from baseline to 1 week (54\% down to 40\%), but this recovered and was slightly above baseline at 6 weeks (61\%). These scores were in or just slightly above the danger zone (Figure 27).
Figure 26
Job Satisfaction Across Groups
ICU SAQ – All Participants

Figure 27
Job Satisfaction Across Groups
ICU SAQ – By Experimental Condition
In the randomized-experimental group there was a very low level of job satisfaction with only 25% reporting positively. This improved at 1 week and again at 6 weeks to 36% and 50%, respectively. Both the randomized-experimental and control groups (discussed above) increased significantly at 6 weeks per Sexton’s criteria, but remained in the danger zone despite improvement (Figure 28).

**Figure 28**

**Job Satisfaction Across Groups**

ICU SAQ – Randomized-Experimental and Control Only

---

**Job Satisfaction (ICU SAQ) With Individual as the Unit of Analysis**

A repeated measures ANOVA was conducted to test the effects of Group (three levels) and Time (three levels) on job satisfaction scores using individual scaled scores as the unit of analysis. In this set of analyses, the three levels for Group were created by dividing the members of the combined-experimental group into two levels so that the
scores for the randomized-experimental and assigned-experimental participants could be compared separately with the scores of the control group members. The ANOVA results for job satisfaction presented a different picture from that shown by the prior analyses that used the percentage of a group above a safety threshold as the unit of analysis. The results of the repeated measures ANOVA were not statistically significant for Group, $F(2, 46) = 1.91, p = 0.160$, or Time $F(1.77, 92) = 0.000, p = 0.999$. The interaction effect of Group x Time, $F(3.55, 92) = 1.74, p = 0.155$, also was not statistically significant in this analysis.

Similar evaluation methods were used to assess the effects of Group (two levels) and Time (three levels) on job satisfaction scores using individual scaled scores as the unit of analysis. In this set of analyses, the two levels for Group were created by taking the randomized-experimental participants as one group and the controls as the other. The results of the repeated measures ANOVA were not statistically significant for Group, $F(1, 28) = 1.75, p = 0.197$, or Time $F(1.50, 42.09) = 0.272, p = 0.700$. The interaction effect of Group x Time, $F(1.5, 42.09) = 2.52, p = 0.106$, also was not statistically significant in this analysis.

A third evaluation repeated measures ANOVA assessed the effect of Group (two levels) and Time (three levels) on job satisfaction scores. In this analysis, the two levels for Group were the combined-experimental group and the control group. The results of the repeated measures ANOVA were not statistically significant for Group, $F(1, 47) = 0.002, p = 0.963$, or Time $F(1.74, 81.9) = 0.403, p = 0.641$. The interaction effect of Group x Time, $F(1.743, 81.9) = 0.665, p = 0.497$, also was not statistically significant in this analysis.
Post-hoc Analyses on Job Satisfaction Questions With Paired t Test

Job satisfaction scores increased significantly at 1 and 6 weeks according to Sexton's criteria. The job satisfaction climate domain questions were separately analyzed from baseline to 6 weeks for the randomized groups using a paired $t$ test. The job satisfaction domain examines how positive the work experience is and includes the following statements:

1. Morale in the ICU is high.
2. Working in this hospital is like being part of a large family.
3. I am proud to work at this hospital.
4. This hospital is a good place to work.
5. I like my job.

When comparing groups from baseline to 6 weeks, on average, the increase in job satisfaction scores from baseline to 6 weeks was statistically significant in the randomized-experimental group. From baseline to 6 weeks when responding to the statement, "Morale in this ICU is high," the randomized participants' mean scores on average increased but were not statistically significant with a Bonferroni correction of $p = 0.01$ ($M = 2.0, SE = 0.275$; $M = 2.58, SE = 0.336$, $t(11) = -2.244$, $p = 0.046$).

Stress Recognition (ICU SAQ) With Group as the Unit of Analysis

Fifty-seven ICU surveys were analyzed at baseline for all participants (randomized-experimental, assigned-experimental, and control) with 37% fully acknowledging effects of stress in the intensive care unit. At 1 week only 22% of respondents fully acknowledged the effects of stress. This increased slightly at 6 weeks.
to 27%, yielding an overall significant decrease in scores regarding stress recognition (Figure 29).

Figure 29
Stress Recognition Across Groups
ICU SAQ – All Participants

In the combined-experimental group, 33% of the respondents fully acknowledged effects of stress at baseline. This decreased to 19% at 1 week and recovered slightly to 22% at 6 weeks. The control group respondents fully acknowledged the effects of stress at 42% initially, with a significant decrease to 25% at 1 week followed by a significant recovery to 35% at 6 weeks. All combined-experimental and control group responses fell into the danger zone (Figure 30).

In the randomized-experimental group, 50% of the respondents fully acknowledged effects of stress at baseline; however, this decreased significantly to 36% at 1 week and even further to 25% at 6 weeks. As stated above, control group
responses were 42% at baseline, 25% at 1 week and 35% at 6 weeks. The responses of all groups remained in the danger zone regarding recognition of the effects of stress (Figure 31).

**Stress Recognition (ICU SAQ) With Individual as the Unit of Analysis**

A repeated measures ANOVA was conducted to test the effects of Group (three levels) and Time (three levels) on stress recognition scores using individual scaled scores as the unit of analysis. In this set of analyses, the three levels for Group were created by dividing the members of the combined-experimental group into two levels (randomized-experimental and assigned-experimental) for separate comparison with the
scores of the control group members. The ANOVA results for stress recognition presented a different picture from that shown by the prior analyses that used the percentage of a group above a safety threshold as the unit of analysis. The results of the repeated measures ANOVA were not statistically significant for Group, $F(2, 46) = 1.12, p = 0.334$, but was statistically significant for Time $F(1.54, 71.08) = 4.19, p = 0.028$, $\eta_p^2 = 0.084$. The interaction effect of Group $\times$ Time, $F(3.09, 71.08) = 2.97, p = 0.036$, $\eta_p^2 = 0.114$, also was statistically significant in this analysis.

Similar evaluation methods were used to assess the effects of Group (two levels) and Time (three levels) on stress recognition scores using individual scaled scores as the unit of analysis. In this set of analyses, the two levels for Group were created by taking the randomized-experimental participants as one group and the controls as the other. The
results of the repeated measures ANOVA were not statistically significant for Group, $F(1, 28) = 0.063, p = 0.804$, or Time $F(1.56, 43.79) = 1.80, p = 0.183$. The interaction effect of Group $\times$ Time, $F(1.56, 43.79) = 0.47, p = 0.012, d = 0.42$, was statistically significant in this analysis.

A third evaluation repeated measures ANOVA assessed the effect of Group (two levels) and Time (three levels) on stress recognition scores. In this analysis, the two levels for Group were the combined-experimental group and the control group. The results of the repeated measures ANOVA were not statistically significant for Group, $F(1, 47) = 1.30, p = 0.261$, or Time $F(1.55, 72.67) = 1.39, p = 0.253$. The interaction effect of Group $\times$ Time, $F(1.55, 72.67) = 5.52, p = 0.010, d = 0.57$, was statistically significant in this analysis.

*Post-hoc Analyses on ICU Stress Recognition Questions With Paired t Test*

Because the repeated measures ANOVA were significant, the climate domain questions were separated and analyzed individually using a paired $t$ test. The stress recognition domain questions were separately analyzed from baseline to 1 week for the randomized-experimental group using a paired $t$ test. When comparing groups from baseline to 1 week, on average, the decrease in stress recognition scores from baseline to 1 week was not statistically significant.

When reacting to the statement “Fatigue impairs my performance during emergency situations,” the mean scores on average decreased from baseline to 1 week in the randomized-experimental group but were not statistically significant with the Bonferroni correction of $p = 0.0125$ ($M = 3.27, SE = 0.333; M = 2.64, SE = 0.411$,
\( t(10) = 2.30, p = 0.046 \). Mean scores also decreased in the randomized-experimental group and but were not statistical significance from baseline to 6 weeks with the Bonferroni correction when reacting to the statement “I am less effective at work when fatigued” \( (M = 3.92, SE = 0.229; M = 3.20, SE = 0.345, t(11) = 2.46, p = 0.032) \).

**Perception of Management (ICU) With Group as the Unit of Analysis**

Fifty-seven surveys were analyzed at baseline for all participants (randomized-experimental, assigned-experimental, and control) with 19% expressing a positive perception of management. This increased to 22% at 1 week and 29% at 6 weeks, reflecting a significant overall change between baseline and 6 weeks (Figure 32).

**Figure 32**

Perception of Management Across Groups
ICU SAQ – All Participants

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
In the combined-experimental group 21% of the respondents had a positive perception of management at baseline. Following simulation, this increased significantly at 1 week to 32% and was sustained at 6 weeks (31%); however, it did not approach the threshold of the danger zone (60% positive response). The positive control group responses decreased from 17% at baseline to 5% at 1 week. There was significant recovery of perception of management at 6 weeks among control group participants to 26%, but all groups remained in the danger zone despite improvement (Figure 33).

Figure 33
Perception of Management Across Groups
ICU SAQ – By Experimental Condition

In the randomized-experimental group, only 8% of the respondents expressed positive perception of management at baseline. This improved significantly to 27% at 1 week, but decreased back to baseline level at 6 weeks (8%). Despite improvements
within groups, all remained in the danger zone regarding perception of management (Figure 34).

**Figure 34**

Perception of Management Across Groups
ICU SAQ – Randomized-Experimental and Control Only

A repeated measures ANOVA was conducted to test the effects of Group (three levels) and Time (three levels) on perception of management scores using individual scaled scores as the unit of analysis. In this set of analyses, the three levels for Group were created by dividing the members of the combined-experimental group into two levels so that the scores for the randomized-experimental and assigned-experimental participants could be compared separately with the scores of the control group members.
The ANOVA results for perception of managements presented a different picture from that shown by the prior analyses that used the percentage of a group above a safety threshold as the unit of analysis. The results of the repeated measures ANOVA was statistically significant for Group, $F(2, 46) = 3.89, p = 0.027, \eta_p^2 = 0.145$, but not for Time $F(2, 92) = 0.118, p = 0.312$. The interaction effect of Group × Time, $F(4, 92) = 6.11, p = 0.656$, was not statistically significant in this analysis.

Similar evaluation methods were used to assess the effects of Group (two levels) and Time (three levels) on perception of management scores using individual scaled scores as the unit of analysis. In this set of analyses, the two levels for Group were created by taking the randomized-experimental participants as one group and the controls as the other. The results of the repeated measures ANOVA were not statistically significant for Group, $F(1, 28) = 1.07, p = 0.310$, or Time $F(2, 56) = 5.05, p = 0.606$. The interaction effect of Group × Time, $F(2, 56) = 0.405, p = 0.669$, also was not statistically significant in this analysis.

A third evaluation repeated measures ANOVA assessed the effect of Group (two levels) and Time (three levels) on perception of management scores. In this analysis, the two levels for Group were the combined-experimental group and the control group. The results of the repeated measures ANOVA were not statistically significant for Group, $F(1, 47) = 0.680, p = 0.414$, or Time $F(1.78, 83.67) = 0.690, p = 0.488$. The interaction effect of Group × Time, $F(1.78, 83.67) = 0.903, p = 0.399$, also was not statistically significant in this analysis.
Post-hoc Analyses on ICU Perception of Management Questions With Paired t Test

Perception of management scores increased significantly at 1 week and then declined back to baseline in the randomized group at 6 weeks. The combined-experimental group remained significant at 6 weeks per the Sexton protocol. The climate score questions were analyzed in all groups from baseline to 1 week, baseline to 6 weeks, and 1 week to 6 weeks. Perceptions of management statements indicate participant approval of managerial actions and include the following statements:

1. Hospital management does not knowingly compromise the safety of patients.
2. Hospital administration supports my daily efforts.
3. I am provided with adequate, timely information about hospital events.
4. The levels of staffing in this ICU are sufficient to handle the number of patients.

A Bonferroni correction of $p = 0.012$ was applied to the perception of management $t$ tests. From baseline to 6 weeks when reacting to the statement, “Hospital management does not knowingly compromise the safety of patients,” the assigned-experimental participants’ mean scores on average decreased and were statistically significant ($M = 4.3, SE = 0.219; M = 3.5, SE = 0.2.84, t(19) = 3.7, p = 0.001, r = 0.50$). In the assigned-experimental group when reacting to the statement “I am provided with adequate timely information about events in the hospital that might affect my work,” mean scores increased on average, demonstrating a statistically significant change from baseline to 1 week ($M = 3.4, SE = 0.255; M = 3.9, SE = 0.191, t(19) = -2.939, p = 0.008, r = 0.746$).
Working Conditions (ICU SAQ) With Group as the Unit of Analysis

Fifty-seven surveys were analyzed at baseline for all participants (randomized-experimental, assigned-experimental, and control) with 35% of respondents reporting a positive working condition in the intensive care unit. At 1 week 37% of respondents reported a positive working condition, and at 6 weeks 40% of the respondents reported a positive working condition (Figure 35), all below the danger zone threshold.

Figure 35
Working Conditions Across Groups
ICU SAQ – All Participants

In the combined-experimental group 30% of the respondents reported a positive working condition at baseline, which increased to 39% at 1 week, then returned to baseline at 6 weeks (31%). The control group response started somewhat higher at baseline (42%), decreased insignificantly at 1 week (35%), and then returned to baseline (43%) at 6 weeks (Figure 36).
In the randomized-experimental group, 17% of the respondents reported a positive working condition at baseline, which increased significantly to 36% at 1 week but decreased to 25% at 6 weeks. Changes in the control group were not significant as stated above, and despite significant changes in the randomized group, all groups remained in the danger zone (Figure 37).

*Working Conditions (ICU SAQ) With Individual as the Unit of Analysis*

A repeated measures ANOVA was conducted to test the effects of Group (three levels) and Time (three levels) on working conditions scores using individual scaled scores as the unit of analysis. In this set of analyses, the three levels for Group were
created by dividing the members of the combined-experimental group into two levels so that the scores for the randomized-experimental and assigned-experimental participants could be compared separately with the scores of the control group members. The ANOVA results for working conditions present a different picture from that shown by the prior analyses that used the percentage of a group above a safety threshold as the unit of analysis. The results of the repeated measures ANOVA were not statistically significant for Group, $F(2, 46) = 1.56, p = 0.324$, but was statistically significant for Time $F(1.75, 80.26) = 3.5, p = 0.041, \eta^2 = 0.071$. The interaction effect of Group $\times$ Time, $F(3.49, 80.26) = 1.90, p = 0.127$, was not statistically significant in this analysis.

Similar evaluation methods were used to assess the effects of Group (two levels) and Time (three levels) on working conditions scores using individual scaled scores as...
the unit of analysis. In this set of analyses, the two levels for Group were created by
taking the randomized-experimental as one group and the controls as the other. The
results of the repeated measures ANOVA were not statistically significant for Group,
$F(1, 28) = 0.58, p = 0.452,$ or Time $F(2, 56) = 2.90, p = 0.063$. The interaction effect of
Group $\times$ Time, $F(2, 56) = 3.27, p = 0.046, d = 0.77,$ was statistically significant in this
analysis.

A third evaluation repeated measures ANOVA assessed the effect of Group (two
levels) and Time (three levels) on working condition scores. In this analysis, the two
levels for Group were the combined-experimental group and the control group. The
results of the repeated measures ANOVA were not statistically significant for Group,
$F(1, 47) = 0.04, p = 0.84,$ or Time $F(2, 94) = 2.78, p = 0.067$. The interaction effect of
Group $\times$ Time, $F(2, 94) = 1.45, p = 0.24,$ also was not statistically significant in this
analysis.

Post-hoc Analyses on ICU Working Conditions Questions With Paired $t$ Test

Working conditions climate scores increased significantly at 1 week and were
trending back to baseline in the randomized group at 6 weeks. The combined-
experimental group change was approaching significance at 1 week and returned to
baseline at 6 weeks per the Sexton protocol. Repeated measures ANOVA was also
statistically significant, so the climate score questions were analyzed in all groups from
baseline to 1 week, baseline to 6 weeks, and 1 week to 6 weeks. The working conditions
domain examines the perceived quality of work environment and logistical support, and
includes the following statements:
1. All the necessary information for diagnosis and therapeutic decisions is available.

2. This hospital constructively deals with problem physicians and employees.

3. Trainees in my discipline are adequately supervised.

4. This hospital does a good job of training new personnel.

A Bonferroni correction of $p = 0.012$ was applied to the working conditions $t$ tests. From baseline to 1 week, working conditions climate scale scores increased and were statistically significant in the randomized-experimental group ($M = 51.7, SE = 5.72; M = 64.8, SE = 5.0; t(10) = -3.610, p = 0.005, r = 0.80$). When reacting to the statement, "All the necessary information for diagnostic and therapeutic decisions is routinely available to me," the randomized-experimental group increased mean climate scores significantly from baseline to 1 week and was reaching statistical significance ($M = 3.36, SE = 0.338; M = 4.18, SE = 0.2.64, t(10) = -2.764, p = 0.020$). When reacting to the statement, "Trainees in my discipline are adequately supervised," the assigned-experimental group experienced increased mean scores with a change from baseline to 1 week, and was nearing statistical significance ($M = 3.53, SE = 0.208; M = 3.9, SE = 0.215, t(18) = -2.7, p = 0.015$).

Quality of Communication and Collaboration

Participants were asked to use a Likert scale (Very Low, Low, Adequate, High, Very High) to describe the quality of communication and collaboration they have experienced with each type of provider in their clinical area. Paired $t$ test analyses were conducted to evaluate changes in climate scores on each version (OR or ICU) of the
SAQ at baseline to 6 weeks. The Bonferroni correction was applied for running multiple \( t \) tests, which requires the more conservative alpha level of \( p = 0.003 \). In the ICU, statistically significant responses occurred with surgical residents and with unit clerks. All demonstrated increase in means except scores related to the charge nurse (Table 12).

Table 12

*ICU SAQ All Participants – Communication and Collaboration*

<table>
<thead>
<tr>
<th>Communication and Collaboration With Different ICU Team Members</th>
<th>Baseline Mean Score</th>
<th>6-Week Mean Score</th>
<th>( t )</th>
<th>( p ) Value(^6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Medical Attending</td>
<td>3.27</td>
<td>3.79</td>
<td>-2.88</td>
<td>0.006</td>
</tr>
<tr>
<td>Surgical Attending</td>
<td>3.77</td>
<td>4.11</td>
<td>-2.64</td>
<td>0.011</td>
</tr>
<tr>
<td>Medical Resident</td>
<td>3.15</td>
<td>3.66</td>
<td>-2.40</td>
<td>0.021</td>
</tr>
<tr>
<td>Surgical Resident</td>
<td>3.58</td>
<td>4.08</td>
<td>-3.50</td>
<td>0.001*</td>
</tr>
<tr>
<td>Charge Nurse</td>
<td>4.50</td>
<td>4.125</td>
<td>2.92</td>
<td>0.005</td>
</tr>
<tr>
<td>RN</td>
<td>4.32</td>
<td>4.53</td>
<td>-2.19</td>
<td>0.033</td>
</tr>
<tr>
<td>PCT</td>
<td>3.90</td>
<td>4.36</td>
<td>-2.75</td>
<td>0.008</td>
</tr>
<tr>
<td>Unit Clerks</td>
<td>3.96</td>
<td>4.36</td>
<td>-3.35</td>
<td>0.002*</td>
</tr>
</tbody>
</table>

In the operating room, the Bonferroni correction was applied for running multiple \( t \) tests, which requires the more conservative alpha level of \( p = 0.003 \). Statistically significant responses occurred with anesthesia attendings. All revealed an increase in mean scores except with perfusionists (Table 13).

\(^{6}\) Statistically significant with Bonferroni correction of \( p = 0.003 \).
Table 13

OR SAQ All Participants – Communication and Collaboration

<table>
<thead>
<tr>
<th>Communication and Collaboration With Different OR Team Members</th>
<th>Baseline Mean Score</th>
<th>6 Week Mean Score</th>
<th>t</th>
<th>p Value(^7)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Surgical Attending</td>
<td>4.02</td>
<td>4.34</td>
<td>-2.98</td>
<td>0.005</td>
</tr>
<tr>
<td>Anesthesia Attending</td>
<td>4.05</td>
<td>4.40</td>
<td>-3.52</td>
<td>0.001(^*)</td>
</tr>
<tr>
<td>Perfusionist</td>
<td>4.03</td>
<td>3.64</td>
<td>2.14</td>
<td>0.040</td>
</tr>
<tr>
<td>Preop Staff</td>
<td>3.33</td>
<td>3.67</td>
<td>-2.03</td>
<td>0.050</td>
</tr>
<tr>
<td>Support Staff</td>
<td>3.33</td>
<td>3.64</td>
<td>-2.23</td>
<td>0.032</td>
</tr>
</tbody>
</table>

Answers to Open-Ended Questions

Participants were asked to give their top three recommendations for improving patient safety in the given unit. This open-ended comments section resulted in 140 suggestions from 65 participants at baseline, 90 suggestions from 40 participants at 1 week, and 82 suggestions from 37 participants at 6 weeks. These suggestions were classified into themes and are reported in Tables 14–16.

\(^7\) Statistically significant with Bonferroni correction of p=0.003
Table 14

**Baseline Recommendations for Improving Safety**

<table>
<thead>
<tr>
<th>Response Category</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Staffing levels / Workload / Appropriate mix of skill levels / Retention of experienced staff</td>
<td>29</td>
</tr>
<tr>
<td>2. Communication (sign-off, provider-nurse, nurse-provider), interdisciplinary communication, Communication of definitive plan and evidence-based information</td>
<td>24</td>
</tr>
<tr>
<td>3. Teamwork, Team attitude, Interdisciplinary cooperation</td>
<td>13</td>
</tr>
<tr>
<td>4. Training / Supervision / Non-punitive request for help</td>
<td>12</td>
</tr>
<tr>
<td>5. Availability of equipment, Supplies, Meds, Work space</td>
<td>6</td>
</tr>
<tr>
<td>6. Management support / Prioritize safety needs / Error and Safety reporting / Promotion of providers (MDs &amp; PAs) to management and leadership roles / Improved morale</td>
<td>8</td>
</tr>
<tr>
<td>7. Implementation of error-reducing technology / Practices, improved reference availability, error recognition, Safety exercises and discussions (simulations, case briefings and debriefings, non-punitive discussion of errors)</td>
<td>29</td>
</tr>
<tr>
<td>8. Clustering of homogenous patient populations</td>
<td>5</td>
</tr>
<tr>
<td>9. Improved physician participation (teaching, multi-disciplinary rounds)</td>
<td>9</td>
</tr>
<tr>
<td>10. Miscellaneous (Clearly identified provider contact information, limit family visitation, statement of refusal to respond / Nothing to suggest / Satisfied)</td>
<td>5</td>
</tr>
</tbody>
</table>

**TOTAL (65 Individuals)** 140 Statements
Table 15  

One Week Recommendations for Improving Safety

<table>
<thead>
<tr>
<th>Response Category</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Staffing levels / Workload / Appropriate mix of skill levels/ Retention of experienced staff</td>
<td>17</td>
</tr>
<tr>
<td>2. Communication (sign-off, provider-nurse, nurse-provider), interdisciplinary communication, Communication of definitive plan and evidence-based information</td>
<td>14</td>
</tr>
<tr>
<td>3. Teamwork, Team attitude, Interdisciplinary cooperation</td>
<td>10</td>
</tr>
<tr>
<td>4. Training / Supervision / Non-punitive request for help</td>
<td>10</td>
</tr>
<tr>
<td>5. Availability of equipment, Supplies, Meds, Work space</td>
<td>11</td>
</tr>
<tr>
<td>6. Management support / Prioritize safety needs / Error and Safety reporting / Promotion of providers (MDs &amp; PAs) to management and leadership roles / Improved morale</td>
<td>8</td>
</tr>
<tr>
<td>7. Implementation of error-reducing technology / Practices, improved reference availability, error recognition, Safety exercises and discussions (simulations, case briefings and debriefings, non-punitive discussion of errors)</td>
<td>16</td>
</tr>
<tr>
<td>8. Clustering of homogenous patient populations</td>
<td>3</td>
</tr>
<tr>
<td>9. Improved physician participation (teaching, multi-disciplinary rounds)</td>
<td>4</td>
</tr>
<tr>
<td>10. Miscellaneous (Clearly identified provider contact information, limit family visitation, statement of refusal to respond / Nothing to suggest / Satisfied)</td>
<td>6</td>
</tr>
</tbody>
</table>

TOTAL (40 Individuals) 99 Statements
Table 16

*Six Week Recommendations for Improving Safety*

<table>
<thead>
<tr>
<th>Response Category</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Staffing levels / Workload / Appropriate Mix of Skill Levels / Retention of Experienced Staff</td>
<td>27</td>
</tr>
<tr>
<td>2. Communication (sign-off, provider-nurse, nurse-provider), Interdisciplinary Communication, Communication of Definitive Plan and Evidence-Based Information</td>
<td>17</td>
</tr>
<tr>
<td>3. Teamwork, Team Attitude, Interdisciplinary Cooperation</td>
<td>2</td>
</tr>
<tr>
<td>4. Training / Supervision / Non-punitive Request for Help</td>
<td>4</td>
</tr>
<tr>
<td>5. Availability of Equipment, Supplies, Meds, Work Space</td>
<td>10</td>
</tr>
<tr>
<td>6. Management Support / Prioritize Safety Needs / Error and Safety Reporting / Promotion of Providers (MDs &amp; PAs) to Management and Leadership Roles / Improved Morale</td>
<td>3</td>
</tr>
<tr>
<td>7. Implementation of Error-Reducing Technology / Practices, Improved Reference Availability, Error Recognition, Safety Exercises and Discussions (simulations, case briefings &amp; debriefings, non-punitive discussion of errors)</td>
<td>8</td>
</tr>
<tr>
<td>8. Clustering of Homogenous Patient Populations</td>
<td>3</td>
</tr>
<tr>
<td>9. Improved Physician Participation (teaching, multi-disciplinary rounds)</td>
<td>3</td>
</tr>
<tr>
<td>10. Miscellaneous (Clearly Identified Provider Contact Information, Limit Family Visitation, Statement of Refusal to Respond / Nothing to Suggest / Satisfied)</td>
<td>5</td>
</tr>
</tbody>
</table>

TOTAL (37 Individuals) 82 Statements
Chapter Summary

As measured by the OR and ICU SAQs, changes in the attitudes of surgical team members were observed following exposure to medical simulation and debriefing exercises. Statistically significant changes were observed in several domains and occurred at various points during the study period. Consistent with the design of team simulation, the greatest positive effect was observed in the teamwork climate domain. An apparent spill-over positive effect was also observed in control group members, which implies the potential for team simulation exercises to have an effect beyond individual participants, extending change into the broader culture of an organization. Additionally, SAQ scores were analyzed to identify domains at risk for adverse effects on patient safety, creating further opportunity for interventions to improve the organization's safety culture. Based on this significance, the null hypotheses were rejected for research questions 1 and 2.

Communication and collaboration items supporting these data were also statistically significant. The ICU revealed statistical significance with residents and unit clerks, while the OR revealed statistical significant rise in anesthesia attending communication and collaboration. Qualitative data collected and collated into themes reflected participants' suggestions for improved safety. The most frequently recommended suggestions were in the areas of staffing levels, workload, and interdisciplinary communication regarding patient hand-offs between disciplines.
CHAPTER V

DISCUSSION

Chapter Preview

This chapter provides a summary of the salient points of the research study and detailed discussion of the results reported in Chapter IV. Implications of these results are also presented along with study limitations. Recommendations for further research at the local and national level are provided as well.

Summary of Research Design and Results

This pre-post partially randomized controlled, quasi-experimental study examined the effects of team medical simulation and debriefing exercises on the attitudes associated with a culture of safety in a hospital. The In-situ® simulation technique focuses on teamwork within a provider’s environment. This technique was presented and described by Hamman et al. (2007) as part of an ongoing research project with the Center of Excellence for Simulation Research at Western Michigan University. That project is designed to evaluate how teams work together during procedures and to characterize behavioral skills of highly effective healthcare team members. An additional objective of the In-situ® project is the development of simulation tools that uniquely address efforts within the healthcare industry to improve patient safety. This technique has not been the norm in healthcare simulation research and there is little reported evidence to support its efficacy.

163
The current study was an extension of the In-situ® project. It focused on the attitudes of two specialty surgical services (cardiac and vascular) in both the operating room (OR) environment and in the Intensive Care Unit/Progressive Care Unit (ICU/PCU) environment to see if the In-situ® simulation technique changed clinicians' attitudes regarding six domains of a culture of safety in either environment. The primary objective of this study was to determine whether attitudes of interdisciplinary team members would change as measured by responses to the OR Safety Attitude Questionnaire (SAQ) and ICU SAQ at baseline (before), 1 week after, and 6 weeks after being exposed to medical team simulation and debriefing. This effect was measured by comparing results for the participants in the experimental simulation approach with survey results gathered at the same time points for control team members who did not take part in the simulation and debriefing exercises.

To date, Marshall et al. (2006), who were working in Australia, are the only researchers who have reported use of the SAQ tool to look at provider attitudes before and after a simulation-based team training program. In contrast with the current study, the Marshall study had a low response rate (56% at baseline and 12% at 3 months), did not include attending surgeons, and showed no statistically significant effects of the simulation on attitudes toward safety. The current research is the first study to look at provider attitudes using the OR and ICU SAQ before and after In-situ® simulations. It had the dual advantages over Marshall's study of: (a) including responses of attending surgeons in the interdisciplinary team (in the quasi-experimental condition, which included assigned-experimental participants); and (b) enjoying a representative response rate (90.6% overall with 96.5% in the OR and 88.7% in the ICU).
Participants were randomly assigned (as much as possible) either to the experimental or control condition, which determined whether or not they completed a simulation exercise and debriefing session. The OR and ICU SAQs (Sexton et al., 2000) were used to measure all participants' perceptions of the hospital's culture of safety at three points—immediately prior to the time when the simulations were scheduled, approximately 1 week after the simulations, and 6 weeks later. Six simulations were conducted, two each in the operating room, the intensive care unit, and a progressive care unit, involving caregivers to cardiac and vascular surgery patients. Because the same ICU questionnaire was used both in the ICU and PCU, the results for those two environments were incorporated in the same analyses.

The study was designed to answer two primary questions. The first asked whether attitudes of cardiac and vascular surgical team members within the operating room (as measured by an OR safety attitudes questionnaire) would change at pre-post-post points (baseline, 1 week, and 6 weeks) after being exposed to medical simulation and debriefing, compared with changes at the same points for members of control group teams. The second question asked if attitudes of cardiac and vascular surgical team members within the intensive care and progressive care units (as measured by an ICU safety attitudes questionnaire) would change after simulation and debriefing, again in comparison with a control group.

These questions were analyzed with procedures that could be grouped into two types according to the unit of analysis—first with the group (or entire team) as the unit of analysis, and then with the individual members of the teams as the unit of analysis (allowing further comparison of experimental and control group effects). Because the
broader purpose was to understand how simulation might influence the culture of safety for cardiovascular surgical and immediate care teams in hospital units, the group portion of the analysis involved application of the Sexton et al. (2006) recommended criteria for interpreting SAQ results with the group as the unit of analysis. This form of analysis was used to determine whether team medical simulation and debriefing would: (a) result in a change in attitudes to increase the proportion of positive responses (score of >75) regarding teamwork climate, safety climate, job satisfaction, stress recognition, management, and working conditions among cardiac and vascular surgery team members; or (b) result in at least a short-term sustainable change in attitudes regarding safety (6 weeks).

The other set of analyses, with the individual as the unit of analysis, included ANOVA, GEE, and paired $t$ tests. The main results of the study were analyzed with a repeated measures ANOVA to measure: (a) within-participant repeated-measures effects of change across the three time points (baseline, 1 week, and 6 weeks); (b) between-group effects for those in experimental and control conditions; and (c) most critically, interaction effects of group-by-time to determine whether the simulation treatment could be credited for changes in the experimental participants that contrasted with any changes among those in the control condition. Additional analyses were conducted for different combinations of participants depending on whether they were truly randomized according to the most rigorous criteria. These analyses did not include any participants who had been assigned by necessity (as most surgeons, anesthesiologists, and residents were) or who ended up in the control group by default (because they were not scheduled to work on the day of simulation). This set of analyses was conducted using binary test scores for
individuals (i.e., based on whether they were above or below the recommended 75% cut-off), applying a General Estimated Equation (GEE) model, to analyze the main effects of time and intervention (i.e., group), as well as the interaction effect which was of particular interest in this “true experiment.”

Discussion of Results

Teamwork Climate

_changes in OR teamwork climate after simulation_. In answer to the first research question, significant positive changes were detected in attitudes toward teamwork safety across the three time points, but the main effect for Time was found for both the control and the experimental group participants. This, plus the lack of an interaction effect made it impossible to attribute the effect directly to the experimental intervention of simulation because everyone changed. As discussed subsequently, however, there may be a number of understandable reasons for that finding.

More specifically, in the operating room there was a main effect of Time on the GEE in the randomized-experimental versus randomized-controls groups with an average increase in scores. This main effect of Time was also seen in operating room in the teamwork climate domain in those individuals who had low baseline scores as measured by the repeated measures ANOVA and in the teamwork climate domain when assessing all baseline scores as measured by the repeated measures ANOVA. These results indicate that average scores increased over time regardless of whether the participant was in the intervention or control group; this was also seen when using the Sexton criteria when using group as the unit of analysis.
One possible explanation for the increase in teamwork climate scores among both the intervention and control groups in the OR is that members of the control group were affected directly by the simulation activities as well as by spillover effects by changes in the culture of the OR. This possible interpretation is supported by an incident that occurred in the operating room during the simulation exercise. In spite of prior instructions, some control members of the cardiac surgery team opened a “pass-through” door located between the operating room and outside hallway. This OR feature allows supplies to be passed into the room through a smaller opening than using the regular door and limits foot traffic into the room during surgery. It also allowed members of the control group to watch the simulation surreptitiously and experience it vicariously. One control group nurse went so far as to call into the room on a portable phone to remind the circulating nurse that she had not done a “time-out” (i.e., a pause prior to the start of a case when the surgeon identifies the patient and states the planned procedure) and told her not to forget it. When asked about the incident, one of the control group nurses said, “There was so much excitement in the air and we couldn’t stand not being a part of it.” She went on to further state, “It was killing us not to be able to jump in and help out our colleagues like we would on any other emergency.” It is likely that this “spillover” effect of the simulation contributed to the increased teamwork climate among both the experimental and control groups. It was also the case that the control group had climate scores that started lower than the randomized-experimental group (50% vs. 64%), so the group had more room to improve, making it easier for them to show a statistically significant improvement.
The “spillover” effect in the OR highlights an important problem with research design in the operating room, which is the difficulty in trying to perform a randomized controlled trial (RCT). The RCT is the gold standard in research, but it is very difficult to keep “pure” in an environment so reliant on close team interaction.

Another possible explanation to consider in the improvement of the control group scores for the OR is that the control participants subsequently worked with the assigned-experimental surgeons as much as the randomized-experimental nurses did. That could explain the benefit the control group indirectly received from the simulation project. Control participants in the OR had many hours of daily exposure to the surgeons who comprised the majority of the assigned-experimental group and who drove the statistically significant increase in teamwork climate scores in the OR. The OR assigned-experimental group was unique in that they started with high teamwork climate scores and ended high, unlike the control group who started in the danger zone and came out of the danger zone as indicated by their 6-week climate scores, where 68% of control participants reported a good teamwork climate score (J. Sexton, personal communication, September 6, 2007).

Subjectively, the In-situ® simulation and debriefing exercise brought a lot of excitement to the OR and was talked about extensively by many team members throughout the project. This is an important consideration given the cost of simulation and debriefing. If the assigned participants (who were mostly surgeons) “acted” differently after involvement and thereby affected the entire team, it would appear that at least in smaller teams, not all members would need the costly intervention to reap benefit.
Changes in ICU/PCU teamwork climate after simulation. In answer to the second research question regarding changes in the ICU following simulation, the results support a conclusion that participants in the randomized-experimental group in the ICU showed improvement in attitudes toward teamwork climate versus the randomized-control group after exposure to simulation and debriefing. This was confirmed when analyzing the individual as the unit of analysis in the randomized-experimental group using a generalized estimating equation (GEE), which revealed a slope of 0.265 \( (p = 0.03) \). This indicated that the intervention group had a significant change in teamwork climate scores that were greater than or equal to 75 over time, versus no significant changes in the teamwork climate scores that were greater than or equal to 75 over time in the control group.

When analyzing only results for those individuals who had low baseline scores, repeated measures ANOVA revealed a main effect for Time when comparing the randomized-experimental and control groups, but no interaction effect was found. This would indicate that average scores changed over time regardless of whether the participant was in the intervention or control group. When analyzing all teamwork climate scores (not just those who started low), the repeated measures ANOVA showed a significant effect for group differences, which suggests that there is a difference in average scores among the members of the randomized-experimental, assigned-experimental, and control groups.

The spillover phenomenon observed in the OR simulation was not replicated in the ICU. Some factors possibly contributing to the lack of spillover in the ICU/PCU environments include the larger team size (three times larger than the OR), scheduling
structure (separate day/night shifts), and function (more independent and less integrated with surgeons and other staff). Controls might not have been scheduled to work with the randomized-experimental group during the study period, so they were less likely to “catch” the positive response from the experimental group. This appears to be the case given that the randomized participants taking the ICU SAQ increased significantly over the control participants at 6 weeks but the group as a whole decreased slightly (51%–49%) and remained in the danger zone (<60%) for teamwork climate. This aspect of the study, which meets the standards of a “true RCT experiment,” provides the strongest positive evidence for concluding that simulation can have a significant influence on teamwork climate in the intensive care and progressive care phases following cardiovascular surgery.

Further interpretation of both OR and ICU/PCU Teamwork Climate results and conclusions. When analyzing those participants who did not report a good teamwork climate in the OR at baseline (i.e., they had baseline SAQ teamwork climate scores of <75%), analyses using repeated measures ANOVA and paired t tests were statistically significant. This would suggest that simulation was more powerful in people who thought teamwork was not “good” at baseline, and did not change those who thought “things are just fine around here” (J. Sexton, personal communication, September 6, 2007). Those who reported good teamwork at baseline (52%) may have made it more difficult to detect the change among those who did not.

When breaking down the OR teamwork climate score, the questions that showed significant positive changes were “The physicians and nurses here work as a well coordinated team” and “Nurse input about patient care is well received in the OR.” This
stresses the importance of the communication and collaboration between physicians and the rest of the interdisciplinary team. Members of interdisciplinary teams can benefit from acknowledging that they are trained differently and may need to communicate differently with other team members. Sexton (2004), for example, recommended that nurses should “talk more like physicians (concisely) and physicians listen more like nurses (attentively)” (p. 50). Physicians and others trained in the medical model (e.g., PAs) are “trained to expect and communicate in bullet points whereas nurses are trained to expect and communicate the story of the patient” (Sexton, 2004, p. 50). This difference between the two groups can be problematic as reported by Sexton (2004):

Nurses are trained to give the narrative of the patient in a detailed and contextualized account. When communicating to physicians (who must then make decisions about the delivery of care), there are oftentimes many opportunities for the physicians to tune out parts of the story, as they wait for the more clinically relevant input upon which to base their decision-making. The result is that clinically relevant information can be lost in the exchange, due to the delivery style of nurses and the listening style of the physicians. (p. 50)

Sexton (2004) discussed other ways for interdisciplinary teams to improve teamwork and stated that “teamwork is about ‘we’ not ‘me’ and encouraging familiarity enhances predictability, which is associated with better team performance.” Sexton reported further that:

Crewmember familiarity may manifest itself as a function of crewmembers referring to themselves in the first person plural. The first person plural (e.g., we, our, us) is frequently expressed in the form of “let’s,” e.g., “let’s get out the landing checklist.” (p. 20)

Enhancement of crewmember familiarity is an important factor in aviation safety. Researchers have found that fatigued crewmembers who had been a team in the past outperformed well-rested crewmembers who had never flown together (Foushee, Lauber, Beatge, & Acomb, 1986). The OR environment traditionally fosters a greater focus on
teamwork. Its level of value to team members is reflected in the positive changes in all SAQ scores in this study.

Improving teamwork requires communication, collaboration and education of interdisciplinary teams. In this study, In-situ® simulation improved teamwork climate scores in a short time period in a statistically significant fashion as measured by the Sexton criteria for all participants in the OR, regardless of whether they participated directly in the simulation, and for experimental participants only in the ICU/PCU, with no significant changes being observed for those in the control group. Because attitudes take time to develop, it would be beneficial to administer the SAQ again in 6 months or 1 year to evaluate the continued progress of the scores.

In conclusion, the In-situ® simulations as performed by the Michigan Center of Excellence in Simulation Research in this study appear to have had a positive effect on teamwork attitudes, as measured by the OR and ICU SAQs, and therefore a positive impact on patient safety. In the case of the OR, a spillover effect seems to have led to positive changes in the teamwork culture of the entire surgical team; whereas in the case of the ICU/PCU, the effect of the experimental simulation could be detected for the experimental group members but not those in the control group who were not exposed.

Safety Climate

The safety climate as measured by the OR and ICU SAQs is defined as the perceptions of a strong and proactive organizational commitment to safety (Sexton et al., 2006). In this study, changes in the safety climate over time were not statistically significant in the OR or ICU as measured by the GEE. The results of the repeated
measures ANOVA also showed no significant effects for Time or Group and had no interaction effect in the ICU. These analyses would indicate that simulation and debriefings had no effect on attitudes of safety climate as measured by the OR/ICU SAQs.

In the OR, when comparing the randomized-experimental versus the control group as analyzed by a repeated measures ANOVA, there actually was a significant decrease in climate scores in both groups indicated by a main effect for the variable of Time. This suggests a change in scores over time regardless of whether the participant was in the intervention or control group, but provides no information on the intervention effect.

Several interpretations for decreases in scores regarding safety climate may be offered. Many factors affect how a participant answers survey questions on any given day. There were many confounding variables that might have affected this domain at the time of the study. Just prior to and during the months of the study there was a large turnover in experienced nursing staff that might have had a negative effect on the outcome related to the factor of team familiarity contributing to a good safety climate. The staff was also introduced to a new unit configuration in preparation to move to a new hospital tower during the study time frame. The staff vocally objected to the plan and claimed that it is an “unsafe” configuration. It would appear that the low-scoring responses as measured by the SAQs correspond to the timing of these external factors that were occurring coincidentally during the time of the study.
Job Satisfaction

There was no interaction effect as measured by the GEE or repeated measures ANOVA on the OR or ICU SAQs in the domain of Job Satisfaction. There were also no effects of Time or Group on either of the SAQs in this domain. Paired t tests at the item level were also not significant on either the OR or ICU SAQs. In the ICU there was a significant increase in percentage of participants who reported good job satisfaction from baseline to 6 weeks according to the Sexton criteria in the randomized-experimental group. This contrasted, however, with a finding in the OR that there was a significant decrease in percentage of participants who reported good job satisfaction from baseline to 6 weeks according to the Sexton criteria in the randomized-experimental group.

Low morale is important as a safety factor because it can result in cognitive failures, which are a product of active failures and can lead to medical error (Vincent, Taylor-Adams, & Stanhope, 1998). The randomized-experimental group in the ICU significantly increased job satisfaction climate scores from baseline to 6 weeks according to the Sexton criteria (25%-50%). When participants in the ICU responded to the statement “Morale in this ICU is high” the mean score increased from baseline to 6 weeks and was nearing statistical significance. However, the percentage of ICU group members (both combined-experimental and control) started in the danger zone and remained there at 6 weeks (54%-58%). The percentage of members of the OR group (combined-experimental and control) started much higher (75%) and remained in the safety zone at 6 weeks (69%), which may have influenced the contrasting results for the two environments.
The randomized-experimental group in the ICU showed significant changes in climate scores when using the group as the unit of analysis according to the Sexton criteria but this might be because they started out so low and had more room to improve. The ICU as a group had a score of 54% at baseline so the randomized-experimental group’s 6-week score may be biased based on a phenomenon known as regression to the mean or regression threat. Theoretically, In-situ® simulation could indirectly have an affect on job satisfaction attitudes, but the likelihood seems small that a single intervention could change those attitudes within 6 weeks. One possible explanation for the higher scores in the OR group also involved a physical relocation of the operating rooms to an entirely new construction. This OR move, however, took place several months prior to the simulation study, possibly resulting in higher scores from individuals who were now comfortable with their new and higher technology unit.

**Stress Recognition**

In the domain of stress recognition, there was no interaction effect as measured by the GEE or repeated measures ANOVA on the OR SAQs. There were also no effects of Time or Group on the OR SAQ as measured by the repeated measures ANOVA. None of the paired t tests at the item level were significant on the OR SAQ.

In the ICU there was no interaction effect as measured by the GEE but there was an interaction effect of Group x Time as measured by the repeated measures ANOVA when comparing the following groups: (a) randomized-experimental, assigned-experimental, and control groups; and (b) combined-experimental and control group. In this interaction, the randomized-experimental and assigned-experimental climate scores
decreased while the control group scores increased slightly from 1 week to 6 weeks. The repeated measures ANOVA was also significant for Time when comparing the randomized-experimental, assigned-experimental, and control groups. There was also a decrease in percentage of participants who fully acknowledged the effect of stress from baseline to 6 weeks according to the Sexton criteria in the combined-experimental and randomized-experimental groups in the ICU. The control group did not change significantly from baseline to 6 weeks according to the Sexton criteria.

This somewhat surprising result is difficult to interpret. Humans make more errors when stressed: therefore, believing that one is able to function effectively when stressed is a defense mechanism that is inconsistent with safety (Helmreich & Merritt, 2001). Interestingly, in this study the increased teamwork and probable familiarity with each other may have led the ICU/PCU team members (regardless of group) to report less stress recognition later in the study. Thirty-seven percent of the participants answering the OR and ICU SAQ fully acknowledged effects of stress at baseline. The OR personnel did not change significantly throughout the study period, whereas the ICU participants significantly decreased at 6 weeks per the Sexton criteria (27%). This decrease occurred in all three groups (combined-experimental, randomized-experimental, and control). This decline also was statistically significant in the repeated measures ANOVA and was nearing statistical significance in the paired t tests. All stress recognition climate scores also were well within the danger zone according to the Sexton criteria. Both questions in the post hoc analyses that were nearing significance had to do with lessened recognition of the effects of fatigue on performance.
This domain requires further research. Prior studies have shown that physicians have less stress recognition than pilots, but this can be explained by education and the restriction of pilots' work hours (Helmreich & Merritt, 2001). Resident physicians have in recent years had a restriction on work hours but this limitation does not apply after they finish residency. Likewise, nurses and other members of the interdisciplinary teams do not have restrictions on their work hours. That can lead to unsafe extension of hours for many reasons, not the least of which is mandated overtime in the occurrence of staffing shortages.

Both teams in this study (cardiac and vascular surgery) involve high risk surgical procedures. Helmreich and Merritt (2001) explain the unrealistic attitudes toward stress recognition among OR personnel as follows:

. . . medical staff who work in the operating room often face the reality that patients die or suffer permanent damage during surgery. Too much thinking upon the consequences of professional acts could have a chilling effect on the physicians' willingness to undertake difficult procedures. Denial of the effects of fatigue is a special case that relates to organizational practices. (p. 37)

The debriefing portion of the In-situ® simulations could be a format for increasing awareness about stress recognition, but as long as clinicians are allowed (and encouraged or even mandated) to work long hours without restriction, educational intervention probably will not have a lasting effect. A local, state, and national emphasis on increasing stress recognition among medical professionals could enhance patient safety, but restructuring of workload and hours for the sake of enhanced safety will likely only be achieved through mandated intervention, as was the case in the aviation industry and other high risk (now government-regulated) industries.
Despite this stated need for improvement among surgical personnel, and also to account for the low recognition scores regarding the effects of stress and fatigue, it is important to be reminded that research has, in fact, shown that fatigued crewmembers who had been a team in the past outperformed well-rested crewmembers who had never flown together before (Foushee et al., 1986). This same experience among surgical and other team-oriented caregivers may tend to blunt their recognition of the potential for increased error in the presence of fatigue, as by routine, they expect their reduced performance in such settings to be enhanced or “covered” by their familiar team partners.

Perception of Management

In this study, scores regarding the perception of management showed no statistically significant changes in the OR or ICU based on simulation as measured by the GEE. This analysis would indicate that simulation and debriefings had no effect on attitudes of perception of management as measured by the OR/ICU SAQs.

On the other hand, the repeated measures ANOVAs were not significant for Time or Group in the OR but did have an interaction effect of Group X Time. The repeated measures ANOVAs were not significant for Time but were significant for Group; there was no interaction effect of Group X Time on the ICU SAQ when comparing the randomized-experimental, assigned-experimental, and control groups. This would suggest that the simulation and debriefing did have an effect on the attitudes of participants regarding perception of management as measured by the OR SAQ but not on the ICU SAQ.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
Vincent et al. (1998) stated that “in medicine, latent failures would be primarily the responsibility of management and of senior clinicians at those times when they are making decisions effecting the organization and their unit” (i.e., heavy workloads, inadequate knowledge or experience, inadequate supervision, inadequate maintenance of equipment). Study participants taking the OR SAQ had higher baseline scores than those taking the ICU SAQ (52% vs. 19%), but both were in the danger zone per Sexton’s scoring criteria. The OR scores were not statistically different at 6 weeks while the ICU scores increased and were statistically significant, with 29% reporting a positive perception of management; however, this change may be the result of a regression threat. When responding to the statement “hospital management does not knowingly compromise the safety of patients,” mean scores decreased in both areas but only achieved statistical significance in the ICU. The low scores would appear to indicate that respondents feel that management is responsible for potential latent failures.

The comments section asked participants to give recommendations for improving safety. The top theme involved staffing levels, workload, and retention of experienced staff, all of which come under the purview of “management,” which lends explanation to the low perception of management scores. Other comments made by study participants on the surveys that fell into the management category related to limited availability of equipment, supplies, work space, management support, and training and supervision of personnel.

Given the strong commitment to safety professed by the study hospital, this is one area that appears to need intervention. It is difficult to delineate by the statements,
however, whether some of these perceptions are a result of anxiety among personnel anticipating the upcoming move into a new hospital tower.

**Working Conditions**

The GEE showed no significant effects for Time and no interaction effect on the OR SAQ. The GEE showed no interaction effect but was significant for Time on the ICU SAQ. This represents a change in climate scores over time regardless of whether the participants of the ICU were in the intervention or control group. However, these analyses would indicate that simulation and debriefings had no effect on attitudes of working conditions as measured by the OR or ICU SAQs.

The working conditions domain measures the perceived quality of the work environment and logistical support. Climate scores related to working conditions started and ended in the danger zone for both the OR and ICU according to the Sexton criteria. The randomized-experimental group in the ICU again started very low (17%) and had a statistically significant gain (36%) at 1 week, which could be a bias based on regression threat. Responses in the comments section indicate the basis for some of the low working conditions climate scores. On the baseline questionnaire free response area there were many comments about the training and supervision of new employees. Participants felt this area needed improvement and that concentration in this area would improve safety.

Participants also indicated by their comments a need for additional resources at their disposal, which ties in to the working conditions climate measure of having “all the necessary information for diagnosis and therapeutic decisions.” Several individuals also stated a desire for education (from physicians) regarding evidence based practices. This
appears to be another area that the participating institution, particularly physicians, could focus on to improve safety.

Quality of Communication and Collaboration

During the course of the study, almost all participants’ individual mean scores in the communication and collaboration section of the SAQ increased (with the exception of ICU charge nurses and perfusionists in the OR). Scores with statistically significant changes included communication and collaboration with the surgical residents, the unit clerks in the ICU, and with anesthesia attendings in the OR. Improvement in communication and collaboration is directly related to the enhanced teamwork focus at the heart of team simulation training.

Qualitative analysis of responses in the comments section of the questionnaire suggested that interdisciplinary communication comments were the second most common remarks made by all participants. Hand-offs or sign-out from providers to nurses and nurses to providers were a concern of many participants. The number of hand-offs between providers at shift change have increased since 2003 when the Accreditation Council for Graduate Medical Education set limits for resident duty hours (Arora & Johnson, 2006). Standardization of this form of communication is currently a Joint Commission on Accreditation Healthcare Organizations (JCAHO) National Patient Safety Goal. JCAHO recommends development of standardized hand-off protocols by mapping the current hand-offs, creating standardized check lists, implementing the check lists, and then monitoring the protocols. The participating institution has just formed a
task force to review, formulate, and implement standardized hand-offs, which should help address this important patient safety concern.

Broader Implications of Results

The Pew Health Professions issued a specific edict for improving interdisciplinary teamwork in 1998. The commission cited a lack of research as an impediment to capitalizing on the unique contributions of interdisciplinary team members to cost and quality of healthcare (O’Neil & Pew Health Professions Commission, 1998). The interdisciplinary teamwork focus of the simulation exercises employed in this study addressed this gap by demonstrating that the use of team simulation and debriefing exercises can improve team members’ perceptions of the climate of interdisciplinary teamwork.

The positive effect observed in the control group in the operating room via “spillover” or “contamination” would suggest that not every member of the team need be directly involved in the simulations to gain benefit. This has tremendous cost implications given that simulation and debriefing is a very costly endeavor. A rotation of individuals for periodic exposure to simulation and debriefing might carry continuous benefit to those who do not participate in each repeat exercise, resulting in the need for less frequent and smaller scale simulation efforts.

Combining SAQ administration with other existing safety measures could help identify which areas within a hospital might benefit from simulation and debriefing intervention. For example, the anonymous error reporting system at the study hospital (PEERS) could be queried to identify units reporting multiple errors in a given time.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
frame. If administration of the SAQ in such a unit demonstrated low teamwork climate scores, it could be postulated that the unit would benefit from a simulation and debriefing intervention. Outcomes following the intervention could be measured both by repeat administration of the SAQ and by monitoring the subsequent number of PEERS reports.

Several institution-specific implications of this study can be suggested. They include the need for team-building exercises to move teamwork climate scores out of the danger zone for the ICU respondents, intervention regarding perception of management by improving staffing levels and training of new employees, increasing visibility of management's commitments to patient safety, application of stress recognition training for all hospital employees and physicians, encouragement of physicians to engage team members in discussions regarding outcomes measures in evidence based practice, and follow-through on the recently established plan to develop standardized hand-offs.

Beyond the measurements of the domains, there were also some conditions identified by participants during the debriefing sessions that the study hospital has already acknowledged as requiring attention. These related primarily to communication deficiencies between departments within the institution, such as the ability to directly communicate with radiographers to help identify the true nature of “STAT” orders such that workload can be appropriately prioritized based on patient condition, education regarding terminology differences between caregivers and blood bank technologists, and improved availability of certain supplies for use during emergencies.

Discussions during debriefing sessions also provided an opportunity for attending physicians to acknowledge the importance of their role in encouraging improved team function. Seemingly elementary concepts such as team members’ name recognition and
the encouragement of input from ancillary personnel emerged as areas that participating physicians recognized as important to improving team dynamics. A related theme emerged throughout discussions regarding the importance of team dynamics on patient safety. The surgeons acknowledged the importance of team briefing before cases to engage members of the team regardless of their roles and to “break the ice” to improve communication during procedures.

Much like the aviation counterpart as the pilot, physicians must remain central to the focus of interventions. “Change the physician, change the team” was a recurring theme in this study. Modern medicine has attempted in recent years to de-emphasize the physician, but studies such as this support retaining the physician as a central player with a leadership role, as well as a team member role in all patient safety endeavors, particularly relating to improving team dynamics.

Study Limitations

The randomized controlled trial is the gold standard for research but the complexity of this study made it difficult to accomplish. As a result, this study had limitations in several areas including randomization, study design, sample size, and the use of the survey instrument.

Randomization

Marshall (2007) stressed how important it is to include key members, such as attending physicians in a simulation and debriefing in order to be successful. The number of attending physicians in cardiovascular surgery at the participating institution was not amenable to true randomization, but it was felt that surgeon participation was vital in
determining whether simulation could affect attitudes of surgical team members. This resulted in the surgeons being assigned to the experimental group based on availability.

Following approval of the overall concept by hospital administrators, preliminary meetings were held with multiple nursing and ancillary personnel managers who were involved with the coordination of the simulation and debriefing exercises. Despite approval from these key personnel, when it came time to randomizing participants, there was a staffing shortfall necessitating individuals to be assigned to simulation as opposed to allowing true randomization. Situations necessitating assignment versus randomization were varied: some providers needed advanced notification for scheduling purposes (surgeons, anesthesiologists); some categories of nurses or technicians included too few consented participants in the job classification (advanced practice nurses, unit clerks); some managers were concerned about staffing on a particular day (surgical intensive care unit nurses on one simulation day; respiratory therapists and blood bank personnel on all days); and two surgical intensive care unit nurses declined to increase the staffing pool to allow randomization but agreed to come in as simulation participants on a day off. Many of these participants needed to know if they were going to be involved in simulation and when they were expected to be in the hospital. These factors led to assignment of 23 individuals based on availability in order to fill critical roles.

For the randomization process, schedules were given to the Quality Institute at the participating institution and those who were available on a given day were randomized to the experimental or control condition. Those consented participants not available on the day of a simulation (i.e., vacation or not scheduled to work) defaulted to the control group. These participants did not have equal chance to be randomized since they were not
available but were kept in the study to ensure adequate sample sizes. Many of the analyses included nonrandomized participants and, thus, should be interpreted with a grain of caution. Only the GEE analysis results reported in Chapter IV were based on the responses of fully randomized experimental and randomized control subjects. Thus, they represent a true experiment, whereas all other results represent “quasi-experimental” results.

Even if no randomization complications had been experienced that made it necessary to assign some participants to the experimental condition or for some participants to default to the control group based on lack of availability, there would have been problems with the experimental design because of locating it within a single organization. Marshall (2007) reported that it is inappropriate to designate a control group within the same organization to measure safety attitudes because the culture of the organization is being measured as opposed to the individual. However, safety culture comprises individual attitudes (climate scores) all of which contribute to an organizational culture, which the researcher in this study regarded as important to evaluating the effectiveness of the simulation intervention. Marshall also indicated that a single center research study is a limitation. The contamination of data in the OR by members of the control group through the pass-through window provided a good example of this. It showed that having a control group in a single center is difficult to accomplish and can be a limitation in the same unit. However, it did result in some interesting findings in both units that may prove to be beneficial regarding the positive “generalization” influences of spill-over “contamination.”
Study Design

Twenty participants (surgeons and PAs) had dual roles in the OR and ICU and could have been randomized to either clinical area. As such, those individuals were given both OR and ICU SAQs to complete at all stages of the study. They were instructed to carefully consider their roles and relationships in the specified unit (OR vs. ICU) as they completed each questionnaire. Participants indicated that they would have no trouble assuming alternate "mindsets" for responding to surveys pertaining to what they regarded as very different environments (OR vs. ICU). Though there is no way to completely rule out that responses to one of the questionnaires influenced responses to the other, the questionnaires themselves are designed to direct responses to the given unit in question. Based on this, Sexton (personal communication, August 31, 2007) recommended retention of the dual response group in the data analysis and agreed with the preservation of each individual's experimental status in evaluating the responses (i.e., simulation participants in the OR were counted as "experimental" in analysis of their responses to both the OR SAQ and the ICU SAQ).

The desire to include as many allied health professional team members as possible led to a modification in the original design of the study. The radiographers and medical technologists were an integral part of the sampling frame, but they spend little time if any in the OR and work sporadically for brief encounters in the ICU and PCU. As such, their responses on the ICU SAQ could not provide a true representation of the culture of these units. This flaw was brought to the researcher's attention by Sexton as he was reviewing the data as part of their initial analysis within the Johns Hopkins Quality and Safety Research Group. Once it was determined that these individuals ($n = 14$) should be
eliminated from analysis, the sample size was reduced to 83 participants. Those 14 individuals could have been retained in the analysis had they been given a different SAQ. Unknown to the research team during the study design phase, an alternate SAQ, the Teamwork and Safety Climate SAQ, is available for application to any unit within an organization. It is designed to direct individuals’ responses to their home unit or department as their point of reference. Future research could add the other version of the SAQ.

Sample Size

Simulation fidelity relies on inclusion of as many team members as are normally involved in a given clinical scenario. Study design oversight resulted in a smaller number of participants in the analysis than originally intended and a reduction in sample size occurred when nonrandomized participants were eliminated from some of the analyses, specifically the GEE.

A reduction in number of participants was also a result of the exclusion of the surveys completed by the medical technologists and radiographers, as mentioned above. These individuals were vital to preserving simulation fidelity, and their full participation in the debriefing session resulted in valuable discussions in terms of identifying process changes needed within the hospital, but their survey results were not included in the analyses.

Other deficiencies in sample size were created simply by the lack of willingness of some individuals to participate. In particular, the anesthesiologists did not have equal numbers of control group participants despite an ample pool of potential participants.
simply through lack of volunteers for randomization. Scheduling availability was
expressed as a major concern, along with the statement from one anesthesiologist that “If
we don’t do a case, we don’t get paid.” This type of response is reportedly not unique to
this study hospital, as noted by Hamman (personal communication, March 20, 2007),
whose team routinely struggles to engage anesthesiologists at various hospitals working
with the Michigan Center of Excellence in Simulation Technology. This response might
be attributed to the large private practice affiliation of the anesthesia group; however, this
response was not encountered from the smaller private practice groups of surgeons, who
likewise do not generate income unless they are performing fee-for-service procedures.

Nursing managers and staff seemed to have less familiarity with the concept of
randomization and the need for “control” participants. They repeatedly expressed great
concern about the need for increased staffing to cover responsibilities of those who might
be randomized to participate in simulations despite a number of nurses volunteering to
work on off-shifts to cover the simulation days. Budget concerns were also expressed by
charge nurses as a deterrent for assigning additional staff on simulation days. A budget
for the required up-staffing was offered to the participating unit managers to help
alleviate this issue.

Though nurse managers agreed to not give initial patient assignments to nurses
who were randomized, the morning of one simulation found the randomized nurses
already working with patient assignments for the day. This resulted in a need to have
their assignments “covered” when the simulation scenario began, which typically
generates a certain degree of angst among conscientious nurses. These individuals had
expressed this potential as a specific deterrent to their agreement to be involved in the project, but they had been reassured by their manager that this would not be the case.

Survey Instrument Tool

With over 1,300 institutions in 13 countries using Sexton’s SAQ, it has become the industry standard for evaluating the safety climate in an organization. Currently, the ICU SAQ is being used in the research institution for a safety program known as the Keystone ICU Patient Safety Initiative. This program has been limited within the study hospital to the intensive care unit and has not involved the operating room or progressive care unit. Many participants in the ICU have filled out the same questionnaire multiple times during the past 2 years, which could result in a history threat to internal validity, given that the Keystone project could have caused the change within this group.

Sexton (via personal communication at various points during and after the study) offered additional input regarding scores obtained from the surveys in this study on the basis of his application of the surveys to over 1,300 hospitals. He and his team of researchers at the Johns Hopkins Institute of Quality and Safety have identified certain “culture nuggets” they have found to have consistent impact on survey results:

1. Early post-intervention scores sometimes worsen as individuals express their disappointment with “how it is” in comparison to “how it could be.”
2. Change in geographic location, unit mergers and changes in unit management each negatively impact teamwork and safety climate scores.
3. Manager changes (even “bad” to “good”) negatively impact climate scores for up to 2 years.
4. Introduction of new technology to a unit can negatively impact climate scores for up to a year.

5. Low safety climate units rely more on agency nurses.

Having these "culture nuggets" available during interpretation of scores within the study institution may improve the study hospital's ability to plan action items on the basis of this project.

Recommendation for Future Studies

Hospitals who participate in simulation and debriefing exercises might benefit from using the SAQ at the institution on an annual basis, benchmarking their climate scores against other institutions and against themselves, as suggested by Sexton et al. (2006). This is currently being done at several institutions around the country. Further administration of the survey throughout the institution should be undertaken to see whether climate scores in other areas of the hospital are consistent with the OR and ICU.

Several deficiencies and opportunities for intervention already have been identified in this study; appropriate interventions should be undertaken given that attitudes predict performance and outcomes, and climate can be targeted and improved. Further simulation exercises would seem beneficial in improving the organization's safety culture; the results of this study indicate that not all employees need to be directly involved in order to yield improvement in the safety climate. Further research should be designed to investigate this recommendation specifically.

Other services and areas within the hospital may be able to replicate the outcomes of this project, and the parent organization will be able to encourage its application in
other hospitals within its system. Because the hospital participates in state and national quality improvement projects for cardiac surgery, results of this study could be applied to other similar programs. The data gathered during the simulation and debriefing intervention and the safety culture assessments will also complement research currently be done by the Michigan Center of Excellence in Simulation Research at Western Michigan University.

Further research to determine how In-situ® simulations affect attitudes toward safety would also be beneficial. This could be accomplished using the SAQ and benchmarking results with current data available and eventually with other institutions using In-situ® simulations to improve safety. Randomized studies would not necessarily be required, but rather a pre-post data collection within the units utilizing the simulation techniques. Sexton recommends waiting 6 months to 1 year after an intervention to measure attitudes since it takes time for attitudes to develop (J. Sexton, personal communication, January 20, 2007), but this study has shown that changes in attitudes can be seen as early as 6 weeks post intervention. Replication studies would be valuable to see if this finding can be repeated.

A randomized controlled trial (RCT) is the gold standard but is difficult in a single institution as demonstrated in this study. An RCT could be accomplished in two different hospitals by having one hospital serve as the experimental group and one serve as the control hospital. However, it would be important for these two hospitals to be in different regions because nurses often work in more than one hospital; if the same nurse were to be involved in both studies there would be a potential “spillover” or
contamination of data. Two separate institutions would also introduce other confounding variables that would require multiple randomized sites to control.

Conclusions

Statistically significant changes were observed in several domains and occurred at various points during the study period. Specifically, the following conclusions can be drawn from the analyses related to the two primary research questions.

Analysis of results of pre-post-post (baseline, 1 week, and 6 week) surveys of attitudes of cardiac and vascular surgical team members (as measured by an OR safety attitudes questionnaire) showed significant differences in scores for the experimental group and control group members as follows:

- In the teamwork climate domain, the GEE in the randomized-experimental versus randomized-controls groups revealed an interaction effect for Time by group with an average increase in scores for the experimental group but not the control group in the ICU condition. The interaction term was not statistically significant in the OR for teamwork climate.
- In the teamwork climate domain in those individuals who had low baseline scores, repeated measures ANOVA revealed a main effect for Time in the following groups: (a) randomized-experimental, assigned-experimental, and control; (b) randomized-experimental and control; and (c) combined-experimental and control.
- In the teamwork climate domain when assessing all baseline scores, repeated measures ANOVA revealed a main effect for Time in the following groups:
(a) randomized-experimental, assigned-experimental, and control; and (b) combined-experimental and control.

- In the teamwork climate domain there was a significant increase in percentages of participants who thought there was a good teamwork climate from baseline to 6 weeks according to the Sexton criteria in the following groups: (a) combined-experimental, and (b) control groups.

- In the safety climate domain when assessing all baseline scores, repeated measures ANOVA revealed a main effect for Time in the randomized-experimental and control group, with the scores going down rather than up.

- In the stress recognition domain there was a significant increase in the percentage of participants who fully acknowledged the effect of stress from baseline to 6 weeks according to the Sexton criteria in the randomized-experimental group.

- In the perception of management domain when assessing all baseline scores, repeated measures ANOVA revealed an interaction effect of Group x Time
when comparing the randomized-experimental, assigned-experimental, and control groups.

- In the working conditions domain there was a significant decrease in the percentage of participants reporting a positive working condition from baseline to 6 weeks according to the Sexton criteria in the following groups: (a) combined-experimental; (b) randomized-experimental, and (c) control groups.

No significant group (experimental, control) by time (pre-post-post) interaction effects were found for the GEE in the other five domains (safety climate, job satisfaction, stress recognition, perception of management, and working conditions). No significant group (experimental, control) by time (pre-post-post) interaction effects were found for the repeated measures domain in the following domains: (a) job satisfaction, (b) stress recognition, and (c) working conditions. In the operating room no significant changes occurred in mean scores on paired t tests at the item level from baseline to 6 weeks in any of the six domains.

Analysis of results of pre-post-post (baseline, 1 week, and 6 week) surveys of attitudes of cardiac and vascular surgical team members (as measured by an ICU safety attitudes questionnaire) showed significant differences in change scores for the experimental group and control group as follows:

- In the teamwork climate domain the GEE in the randomized-experimental versus randomized-controls groups revealed an interaction effect with an average increase in scores among the randomized-experimental group versus the control group over time.
In the teamwork climate domain in those individuals who had a low baseline score repeated measures ANOVA revealed a main effect for Time in the following groups: (a) randomized-experimental, assigned-experimental, and control; and (b) randomized-experimental and control.

In the teamwork climate domain when assessing all baseline scores, repeated measures ANOVA were statistically significant for Group when comparing the randomized-experimental, assigned-experimental, and control groups.

When conducting a paired t test on the teamwork climate domain at the item level, the randomized-experimental group scores increased from baseline to 6 weeks when responding to the statement, “the physicians and nurses here work as a well coordinated team.”

In the teamwork climate domain there was a significant increase in the percentage of participants who reported a good teamwork climate from baseline to 6 weeks according to the Sexton criteria in the randomized-experimental group.

In the teamwork climate domain there was a significant decrease in the percentage of participants who reported a good teamwork climate from baseline to 6 weeks according to the Sexton criteria in the control group.

In the job satisfaction domain there was a significant increase in the percentage of participants who reported good job satisfaction from baseline to 6 weeks according to the Sexton criteria in the randomized-experimental group.
• In the stress recognition domain when assessing all baseline scores, repeated measures ANOVA were statistically significant for Time when comparing the randomized-experimental, assigned-experimental, and control groups.

• In the stress recognition domain when assessing all baseline scores, repeated measures ANOVA were statistically significant for interaction effect of Group x Time when comparing the following groups: (a) randomized-experimental, assigned-experimental, and control groups; and (b) combined-experimental and control group.

• In the stress recognition domain there was a significant decrease in the percentage of participants who fully acknowledged the effect of stress from baseline to 6 weeks according to the Sexton criteria in the combined-experimental and randomized-experimental groups.

• When conducting a paired $t$ test on the perception of management domain at the item level, the assigned-experimental group scores decreased from baseline to 6 weeks when responding to the statement, “hospital management does not knowingly compromise the safety of patients.”

• In the perception of management domain when assessing all baseline scores, repeated measures ANOVA were statistically significant for Group when comparing the randomized-experimental, assigned-experimental, and control groups.

• In the perception of management domain there was a significant increase in the percentage of participants who reported a positive perception of
management from baseline to 6 weeks according to the Sexton criteria in the combined-experimental group.

- In the working condition domain, the GEE in the randomized-experimental versus randomized-controls groups revealed a main effect of Time with an average increase in scores. The interaction term was not statistically significant in the ICU for working conditions.

- In the working conditions domain when assessing all baseline scores, repeated measures ANOVA were statistically significant for Time when comparing the randomized-experimental, assigned-experimental, and control groups.

- In the working conditions domain when accessing all baseline scores repeated measures ANOVA were statistically significant for an interaction effect of Group × Time when comparing the randomized-experimental and control group.

No significant group (experimental, control) by time (pre-post-post) interaction effects were found for the GEE in the four domains (safety climate, job satisfaction, stress recognition, and perception of management). No significant group (experimental, control) by time (pre-post-post) interaction effects were found using repeated measures ANOVAs in the following domains: (a) safety climate, and (b) job satisfaction. In the ICU no significant changes occurred in mean scores on paired t tests at the item level from baseline to 6 weeks in the following domains: (a) safety climate, (b) job satisfaction, (c) stress recognition, and (d) perception of management.
Beyond the analyses using individual scores as the unit of analysis, conclusions were drawn about experimental treatment effects using the group as the unit of analysis applying Sexton’s criteria for group change. These results were as follows:

- In the teamwork climate domain there was a significant increase in the percentage of all participants who reported a good teamwork climate from baseline to 6 weeks according to the Sexton criteria on the OR SAQ.
- In the safety climate domain there was a significant decrease in the percentage of all participants who reported a good safety climate from baseline to 6 weeks according to the Sexton criteria on the OR SAQ.
- In the stress recognition domain there was a significant decrease in the percentage of all participants who fully acknowledged the effect of stress from baseline to 6 weeks according to the Sexton criteria on the ICU SAQ.
- In the perception of management domain there was a significant increase in the percentage of all participants who reported a positive perception of management from baseline to 6 weeks according to the Sexton criteria on the ICU SAQ.

When using the group as the unit of analysis and applying Sexton’s criteria for group change there were no significant changes from baseline to 6 weeks in the job satisfaction domain and working conditions domain.

Additionally, SAQ scores were analyzed to identify domains at risk according to Sexton’s criteria for adverse effects on patient safety (<60% of participants reporting a good climate domain at 6 weeks), creating further opportunity for interventions to
improve the organization’s safety culture. According to the Sexton criteria, interventions should be focused for improving the following domains and clinical areas:

- Teamwork climate domain in the ICU
- Safety climate domain in the OR and ICU
- Job satisfaction domain in the ICU
- Stress recognition domain in the OR and ICU
- Perception of management domain in the OR and ICU
- Working conditions domain in the OR and ICU

Consistent with the design of team simulation exercises, the greatest positive effect was observed in the teamwork climate domain. An apparent spillover positive effect was also observed in OR control group members, which implies the potential for team simulation exercises to have an effect beyond individual participants, extending change into the broader culture of an organization.
REFERENCES


Hamman, W., Rutherford, W., Seiler, B., Beaubien, J., Rubinfeld, I., Lammers, R., et al. (2007, March). *Moving simulation to new levels of realism within healthcare*
organizations. Paper presented at the Safety Across High-Consequence Industries Conference, St. Louis, MO.


Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.


Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.


Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.


Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
Appendix A

Human Subjects Institutional Review Board
Letters of Approval
Date: December 19, 2006

To: Nickola Nelson, Principal Investigator
    William Hammon, Co-Principal Investigator
    Diane Jones, Student Investigator for dissertation

From: Amy Naugle, Ph.D., Chair

Re: HSIRB Project Number: 06-05-04

This letter will serve as confirmation that your research project entitled "Creating a Culture of Safety: The Influence of Medical Simulation" has been approved under the full category of review by the Human Subjects Institutional Review Board. The conditions and duration of this approval are specified in the Policies of Western Michigan University. You may now begin to implement the research as described in the application.

Please note that you may only conduct this research exactly in the form it was approved. You must seek specific board approval for any changes in this project. You must also seek reapproval if the project extends beyond the termination date noted below. In addition if there are any unanticipated adverse reactions or unanticipated events associated with the conduct of this research, you should immediately suspend the project and contact the Chair of the HSIRB for consultation.

The Board wishes you success in the pursuit of your research goals.

Approval Termination: May 17, 2007
September 13, 2006

Diane Jones, MS, PA-C
5325 Elliott Drive Suite 102
Ypsilanti, MI 48197

Dear Ms Jones:

At its meeting on September 5, 2006, the SJMHS Institutional Review Board reviewed the research protocol listed below:

**HSR-06-0685** Creating a culture of safety: The influence of medical simulation.

**IRB Action:** The IRB reviewed response letter dated August 7, 2006, letter from Legal, letter from HIPAA Privacy Officer, letter from David Getty, Respiratory Care and Cardiopulmonary Services, Revised Study Protocol for Dissertation research and revised consent and approved this study for 12 months. (This study was initially reviewed at the June 6, 2006 full-convened meeting.) Comments from the June 6, 2006 meeting have been addressed.

A copy of the approved, revised consent form, with the date of IRB approval, is enclosed. Please review the consent form, and if you have any questions, feel free to contact me at (734) 712-3283.

Federal regulations require that the IRB review each research project at least annually. In approximately one year, or upon completion of your research project, a report will need to be submitted to the IRB that addresses the following points:

1. The current status of the investigation (completed or continuing);
2. A list of subjects participating in the project since its beginning, (using initials, code number or other means to maintain anonymity), and the date of entry for each subject;
3. A description of the experience of the subjects, including adverse reactions, complications, benefits and/or withdrawals from the study;
4. A summary of the research results thus far;
5. A current assessment of the risks and benefits based on study results, including any new information that has come to light since the IRB's last review;
6. A copy of the current consent form, if applicable.
Appendix B

Results Summary Tables
Results Summary

<table>
<thead>
<tr>
<th>TEAMWORK CLIMATE</th>
<th>OR</th>
<th>ICU</th>
<th>OR</th>
<th>ICU</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sexton Criteria</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All Participants</td>
<td>Sig.</td>
<td>AB: All 52 / 60 / 67 (n=46)</td>
<td>ns</td>
<td>AB: All 51 / 53 / 49 (n=37)</td>
</tr>
<tr>
<td>Combined-Experimental / Control</td>
<td>Sig.</td>
<td>Combined-Exp. 54 / 60 / 65 (n=36)</td>
<td>ns</td>
<td>Combined-Exp. 56 / 58 / 59 (n=33)</td>
</tr>
<tr>
<td>Randomized-Experimental / Control</td>
<td>ns</td>
<td>Randomized-Exp. 64 / 65 / 64 (n=31)</td>
<td>Sig.</td>
<td>Randomized-Exp. 55 / 57 / 50 (n=32)</td>
</tr>
</tbody>
</table>

Repeated Measures ANOVA

| Randomized-Experimental / Assigned-Experimental / Control | Sig. | Time F(2,52)=2.94, p=0.058 |
| Randomized-Experimental / Control | ns | ns |
| Combined-Experimental / Control | Sig. | Time F(2,49)=3.7, p=0.029 |

Low Baseline Scores: Repeated Measures ANOVA

| Randomized-Experimental / Assigned-Experimental / Control | Sig. | Time F(2,49)=6.4, p=0.004 |
| Randomized-Experimental / Control | ns | ns |
| Combined-Experimental / Control | Sig. | Time F(2,45)=5.1, p=0.010 |

Paired t-test on Individual Climate Items

| "The physicians and nurses here work as a well coordinated team." | ns | ns |
| "Nurse input about patient care is well received in the OR." | ns | ns |
| "It is easy for personnel in this ICU to ask questions when there is something they do not understand." | ns | ns |
## Results Summary—Continued

### SAFETY CLIMATE

<table>
<thead>
<tr>
<th>Sexton Criteria</th>
<th>OR</th>
<th>OR</th>
<th>ICU</th>
<th>ICU</th>
</tr>
</thead>
<tbody>
<tr>
<td>All Participants</td>
<td>Sig.</td>
<td>All 61 / 57 (n=46)</td>
<td>ns</td>
<td>All 54 / 57 (n=47)</td>
</tr>
<tr>
<td>Combined-Experimental / Control</td>
<td>Sig.</td>
<td>Combined-Exp. 58 / 54 (n=26)</td>
<td>ns</td>
<td>Combined-Exp. 58 / 54 (n=13)</td>
</tr>
<tr>
<td>Randomized-Experimental / Control</td>
<td>Sig.</td>
<td>Randomized-Exp. 48 / 53 (n=11)</td>
<td>ns</td>
<td>Randomized-Exp. 48 / 55 (n=12)</td>
</tr>
</tbody>
</table>

**Repeated Measures ANOVA**

- Randomized-Experimental / Assigned-Experimental / Control: ns
- Randomized-Experimental / Control: Sig. Time $P=0.27$
- Combined-Experimental / Control: ns

**General Estimating Equation (GEE)**

- Randomized-Experimental / Randomized-Control: Time $P=0.08$

**Paired $t$-test on Individual Climate Items**

### JOB SATISFACTION

<table>
<thead>
<tr>
<th>Sexton Criteria</th>
<th>OR</th>
<th>OR</th>
<th>ICU</th>
<th>ICU</th>
</tr>
</thead>
<tbody>
<tr>
<td>All Participants</td>
<td>ns</td>
<td>All 76 / 69 (n=46)</td>
<td>ns</td>
<td>All 54 / 57 (n=47)</td>
</tr>
<tr>
<td>Combined-Experimental / Control</td>
<td>ns</td>
<td>Combined-Exp. 77 / 69 (n=26)</td>
<td>ns</td>
<td>Combined-Exp. 55 / 61 (n=13)</td>
</tr>
<tr>
<td>Randomized-Experimental / Control</td>
<td>Sig.</td>
<td>Randomized-Exp. 82 / 64 (n=11)</td>
<td>Sig.</td>
<td>Randomized-Exp. 25 / 36 (n=12)</td>
</tr>
</tbody>
</table>

**Repeated Measures ANOVA**

- Randomized-Experimental / Assigned-Experimental / Control: ns
- Randomized-Experimental / Control: ns
- Combined-Experimental / Control: ns

**General Estimating Equation (GEE)**

- Randomized-Experimental / Randomized-Control: Time $P=0.08$

**Paired $t$-test on Individual Climate Items**

*“Moral in this ICU is high.”*

<table>
<thead>
<tr>
<th>ns</th>
<th>Randomized-Experimental</th>
<th>Baseline to 6 Weeks</th>
</tr>
</thead>
<tbody>
<tr>
<td>7.50</td>
<td>$M=2.0$, $SD=0.275$</td>
<td>$t(11)=2.244$, $P=0.046$</td>
</tr>
<tr>
<td>STRESS RECOGNITION</td>
<td>OR</td>
<td>OR</td>
</tr>
<tr>
<td>---------------------</td>
<td>----</td>
<td>----</td>
</tr>
<tr>
<td>Sexton Criteria</td>
<td></td>
<td></td>
</tr>
<tr>
<td>All Participants</td>
<td>ns</td>
<td>All 37 / 40 / 36 (n=46)</td>
</tr>
<tr>
<td>Combined-Experimental / Control</td>
<td>ns</td>
<td>Combined-Exp. 31 / 36 / 31 (n=26)</td>
</tr>
<tr>
<td>Randomized-Experimental / Control</td>
<td>Sig.</td>
<td>Randomized-Exp. 38 / 36 / 45 (n=13)</td>
</tr>
<tr>
<td>Repeated Measures ANOVA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Randomized-Experimental / Assigned-Experimental / Control</td>
<td>ns</td>
<td></td>
</tr>
<tr>
<td>Randomized-Experimental / Control</td>
<td>ns</td>
<td></td>
</tr>
<tr>
<td>Combined-Experimental / Control</td>
<td>ns</td>
<td></td>
</tr>
<tr>
<td>General Estimating Equation (GEE)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Randomized-Experimental / Randomized-Control</td>
<td>ns</td>
<td>Time $p=0.59$</td>
</tr>
<tr>
<td>Paired t-test on Individual Climate Items</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&quot;Fatigue impairs my performance during emergency situations.&quot;</td>
<td>ns</td>
<td>$t(10)=2.30, p=0.046$</td>
</tr>
<tr>
<td>&quot;I am less effective at work when fatigued.&quot;</td>
<td>ns</td>
<td>$t(10)=2.30, p=0.046$</td>
</tr>
</tbody>
</table>
Results Summary—Continued

<table>
<thead>
<tr>
<th>PERCEPTION OF MANAGEMENT</th>
<th>OR</th>
<th>OR</th>
<th>ICU</th>
<th>ICU</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sexton Criteria</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All Participants</td>
<td>ns</td>
<td>ns</td>
<td>All 19 / 22 / 29 (n=57)</td>
<td></td>
</tr>
<tr>
<td>Combined-Experimental / Control</td>
<td>ns</td>
<td>ns</td>
<td>Combined-Exp. 21 / 32 / 31 (n=33)</td>
<td></td>
</tr>
<tr>
<td>Randomized-Experimental / Control</td>
<td>ns</td>
<td>ns</td>
<td>Randomized-Exp. 8 / 27 / 8 (n=12)</td>
<td></td>
</tr>
<tr>
<td>Repeated Measures ANOVA</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Randomized-Experimental / Assigned-Experimental / Control</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group x Time, F(4,82)=0.46, p=0.612* RE - M=67.6(11.5), M=62(15.2) AE - M=70.5(18.3), M=72(19.3) C - M=71(14.5), M=72(16.3)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time x Intervention x Group, F(2,46)=3.89, p=0.027* RE - M=66.4(18.4), M=61.2(22.6) AE - M=54.3(14.7), M=55.8(20.2)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>General Estimating Equation</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Randomized-Experimental / Randomized-Control</td>
<td>ns</td>
<td>ns</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time p=0.29</td>
<td>Intervention x Time p=0.04</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Paired t-test on Individual Climate Items</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overall Mean Scores Baseline to 1 Week</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Assigned-Experimental</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M=70.5, SE=4.38; M=79.01, SE=3.86</td>
<td>t(13)=3.18, p=0.007*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overall Mean Scores One Week to 6 Weeks</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Assigned-Experimental</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M=79, SE=0.96; M=71.9, SE=5.15</td>
<td>t(13)=2.30, p=0.012*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&quot;Hospital administration supports my daily efforts.&quot;</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Assigned-Experimental Baseline to 1 Week</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M=3.9, SE=0.00; M=4.14, SE=0.25</td>
<td>t(13)=0.9, p=0.36</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&quot;If I am provided with adequate, timely information about events in this hospital that affect my work.&quot;</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Assigned-Experimental One Week to 6 weeks</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M=3.73, SE=0.33; M=3.64, SE=0.27</td>
<td>t(13)=1.47, p=0.16</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&quot;The levels of staffing in our ORs are sufficient to handle the number of patients.&quot;</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Assigned-Experimental One week to 6 Weeks</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M=4.29, SE=0.12; M=3.86, SE=0.25</td>
<td>t(13)=2.12, p=0.04</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&quot;Hospital management does not knowingly compromise the safety of patients.&quot;</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Assigned-Experimental Baseline to 6 Weeks</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M=4.3, SE=0.19; M=3.5, SE=0.24</td>
<td>t(13)=2.3, p=0.03</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
Results Summary—Continued

<table>
<thead>
<tr>
<th>WORKING CONDITIONS</th>
<th>OR</th>
<th>OR</th>
<th>ICU</th>
<th>ICU</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sexton Criteria</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All Participants</td>
<td>Sig.</td>
<td>Sig.</td>
<td>ns</td>
<td>All 35 / 37 / 40 (n=57)</td>
</tr>
<tr>
<td>Combined-Experimental / Control</td>
<td>ns</td>
<td>Combined-Exp. 30 / 39 / 31 (n=33)</td>
<td>ns</td>
<td>Control 42 / 35 / 43 (n=24)</td>
</tr>
<tr>
<td>Randomized-Experimental / Control</td>
<td>Sig.</td>
<td>Sig.</td>
<td>ns</td>
<td>Randomized-Exp. 17 / 36 / 25 (n=12)</td>
</tr>
</tbody>
</table>

| Combined-Experimental / Control | ns | Control 50 / 45 / 37 (n=20) |

| **Repeated Measures ANOVA** |    |    |     |     |
| Randomized-Experimental / Assigned-Experimental / Control | ns | ns | ns  |

| Randomized-Experimental / Control | ns | ns | ns  |

| Combined-Experimental / Control | ns | ns | ns  |

| **General Estimating Equation** |    |    |     |     |
| Randomized-Experimental / Assigned-Experimental / Control | Time F(2. 56)=3.27, p=0.046* |

| Randomized-Experimental / Assigned-Experimental / Control | Time p=0.03* |

| Combined-Experimental / Control | ns | ns | ns  |

| **Paired t-test on Individual Climate Items** |    |    |     |     |
| Overall Mean Scores Baseline to 1 Week | Sig.↑ | Randomized-Experimental M=51.7, SE=5.2, t(15)=3.610, p=0.005 |

| Assigned-Experimental M=3.79, SE=0.306, t(15)=2.857, p=0.013 |

| Randomized-Experimental Baseline to 1 Week M=3.36, SE=0.338, t(19)=2.704, p=0.020 |

| Assigned-Experimental Baseline to 1 Week M=3.53, SE=0.208, t(15)=2.7, p=0.015 |

| Randomized-Experimental M=3.43, SE=0.173, t(15)=2.857, p=0.013 |

| Assigned-Experimental M=3.9, SE=0.215, t(15)=2.7, p=0.015 |

---

**All the necessary information is available before the start of a procedure.**

**Trainees in my discipline are adequately supervised.**

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.