2020

Assessing Surgical Task Load and Performance: A Comparison of Simulation and Maritime Operation

Spitzer, Holly V.; Hoang, Tuan; Pierce, Eric; Franciose, Reginald J.; Pena, Matthew; Shattuck, Nita L.; Bass, Cameron R.; Blankenship, Jeffrey; Juliano, Michael; Lindemann, Cameron B....

Association of Military Surgeons of the United States

Spitzer, Holly V., et al. "Assessing surgical task load and performance: A comparison of simulation and maritime operation."Military medicine185.Supplement_1 (2020): 599-609.
http://hdl.handle.net/10945/70873

MILITARY MEDICINE, 185, S1:599, 2020

Assessing Surgical Task Load and Performance: A Comparison of Simulation and Maritime Operation

2LT Holly V. Spitzer, MC, USA; CAPT Tuan Hoang, MC, USN†; Eric Pierce, MS‡; Reginald J. Franciose, MD, FACS§; Matthew Pena, MD, FASA‖; Nita L. Shattuck, PhD¶; Cameron R. Bass, PhD**; Jeffrey Blankenship, MS‡; Michael Juliano, MD, FACEP††; ENS Cameron B. Lindemann, MC, USNR*; 2Lt Hayden A. Springer, USAF, MC*; COL Anthony J. LaPorta, MC, USA, (Ret.)†*

ABSTRACT Introduction: This study examined the effects of simulated and actual vessel motion at high seas on task load and surgical performance. Methods: This project was performed in phases. Phase I was a feasibility study. Phase II utilized a motion base simulator to replicate vessel motion. Phase III was conducted aboard the U.S. Naval Ship Brunswick. After performing surgical tasks on a surgical simulation mannequin, participants completed the Surgical Task Load Index (TLX) designed to collect workload data. Simulated surgeries were evaluated by subject matter experts. Results: TLX scores were higher in Phase III than Phase II, particularly at higher sea states. Surgical performance was not significantly different between Phase II (84%) and Phase III (89%). Simulated motions were comparable in both phases. Conclusions: Simulated motion was not associated with a significant difference in surgical performance or deck motion, suggesting that this simulator replicates the conditions experienced during surgery at sea on the U.S. Naval Ship Brunswick. However, Surgical TLX scores were dramatically different between the two phases, suggesting increased workload at sea, which may be the result of time at sea, the stress of travel, or other factors. Surgical performance was not affected by sea state in either phase.

INTRODUCTION

Despite technological advancement and medical innovation, there are still opportunities for improvement in critical care medicine, particularly in austere environments.[1–3] Regardless of the evacuation paradigm, for example the "golden hour" or "platinum thirty", it is clear from previous studies and medical experience that time is of the essence in emergency scenarios.[4] To decrease the time between point of injury and surgical intervention, thereby shortening the time until life-saving or definitive treatment, a possible solution is expanded implementation of surgery at sea. Although aspects of this philosophy have been implemented in the current continuum of care, including the use of large hospital ships, the feasibility of performing emergency surgeries in littoral ships with potential for increased shipboard motion has not been investigated. Effects of motion on skill, speed, focus, and critical thinking are unknown. This study investigated the ability of U.S. Navy medical personnel to perform damage control surgery aboard nontraditional U.S. Navy vessels during high-sea states (SSs). These nontraditional vessels, fast support vessels with modular sea container capabilities, may have higher motion states than larger monohull vessels. This article examines the methods of two experimental phases of this study (II and III) to assess the effects of simulated and real deck motions on surgical performance and individual stress. This will help inform further implementation of surgery at sea trading the potential for more severe motion environment in nontraditional vessels for improvements in time-to-treatment. It will guide future training and the development of procedures for these types of missions.

DISCLOSURES

The Effect of High Deck Accelerations on Surgical Tasks study was performed by the Naval Surface Warfare Center, Panama City Division. Phase I was sponsored by the Office of Naval Research. Phase II was sponsored by the Office of the Chief of Naval Operations N81 Assessments Divisions, Deputy, and Medical Analysis Branch. Phase II was conducted at Naval Surface Warfare Center, Panama City Division (NSWC PCD) Biodynamics Laboratory by the Human Systems Integration Team. The Effect of High Deck Accelerations on Surgical Tasks study, Phase III was sponsored by the Office of the Chief of Naval Operations N81

---

*Rocky Vista University School of Medicine, 8401 S. Chambers Road, Parker, CO 80134

†USN Medical Readiness Division-San Diego, 2450 Craven Street, San Diego, CA 92136

‡Naval Surface Warfare Center, PCD, 110 Vernon Ave, Panama City, FL 32407

§Denver Health Medical Center, 777 Bannock St, Denver CO 80204

‖Naval Hospital Pensacola, 6000 US-98, Pensacola, FL 32512

¶Naval Postgraduate School, 1 University Circle, Monterey, CA 93943

**Duke University, 2301 Erwin Road, Durham, NC 27708

††Fleet Health Services, 7928 14th Street, Norfolk, VA 23505

Presented as a poster at the 2018 Military Health System Research Symposium, August 2018, Kissimmee, FL; abstract # MHSRS-18-1180

The authors have no financial interest that warrants disclosure. Opinions herein are solely those of the author and do not represent those of the U.S. Navy, Department of Defense, or the U.S. Government. Distribution A.

doi:10.1093/milmed/usz297

Assessments Division, Deputy, Medical Analysis Branch and the Navy Advanced Medical Development Program Office. It was planned and executed by NSWC PCD with support from Naval Postgraduate School, Duke University, and subject matter experts (SMEs) from several medical commands and medical schools. It occurred aboard U.S. Naval Ship (USNS) Brunswick 30JAN2017 through 14FEB2017 during a transit between Norfolk, VA and San Diego, CA. The initial report for this study was prepared by the Human Systems Integration Team of the Test Engineering Branch (Code E41) at the Naval Surface Warfare Center Panama City Division. Phase II and Phase III expand upon a previous study sponsored by the Office of Naval Research, Division 342, Warfighter Protection and Applications Division conducted by NSWC PCD. U.S. Navy Experimental Diving Unit Institutional Review Board (IRB) reviewed protocols for all three phases of study. The IRB verified compliance with Title 32 Code of Federal Regulation Part 219 (Protection of Human Subjects), Department of Defense (DoD) Instruction 3216.02 (Protection of Human Subjects and Adherence to Ethical Standards in DoD-Supported Research), Secretary of the Navy Instruction 3900.39E (Protection of Human Subjects), and Naval Sea Systems Command Instruction 3900.12 (Human Research Protection Program). Recruitment of all participating medical personnel followed the informed consent process outlined by each approved protocol. Additionally, research monitors were present during each test event, ensuring the safety of human subjects and well-being.

FIGURE 1. Simulator used in this study: Moog Series 6DOF500E, Model 170.

TABLE 1. NATO Sea States

Sea State	Wave Height (Feet)
0–1	0–0.3
2	0.3–1.6
3	1.6–4.1
4	4.1–8.2
5	8.2–13.1
6	13.1–19.7
7	19.7–29.5
8	29.5–45.5
>8	>45.5

METHODS

This study was performed in three phases. Phase I was a feasibility study to understand whether further investigation was warranted. This initial phase consisted of general surgeons and Independent Duty Corpsmen performing basic medical procedures under no motion conditions and under emulated SS4 conditions based on accelerations collected aboard Littoral Combat Ship USS Independence (LCS-2). The simulator is a 10 ft. by 10 ft. Moog Series 6DOF500E, Model 170 (Fig. 1) electrically actuated platform. It is a variant of the classic "Stewart Table" composed of six actuators that are operated by an electric motor to produce motion and is similar to flight simulators. It can simulate motions of roll, pitch, yaw, heave, and sway.[5] For Phases I and II, it was programmed to combine these motions to simulate recorded deck motions from various naval vessels collected by the NSWC PCD and NSWC Carderock Division while on voyages at different North Atlantic Treaty Organization (NATO) SSs.[6] NATO SSs are measures of the severity of sea conditions, graded from SS 1–8. Table 1 shows the NATO classification of SS. During Phase I, the Independent Duty Corpsmen completed basic medical procedures, including intravenous catheter insertion, intubation, suturing, cricothyrotomy, needle decompression, and tube thoracostomy. The physicians completed advanced medical procedures including tube thoracostomy, fracture stabilization, fasciotomy, and vascular shunting. The test team, NSWC PCD personnel and trained medical professionals, collected the following outcome measures for each procedure: completion time, accuracy, workload, heart rate, and oxygen consumption. Phase I results supported further investigation.

Phase II added surgical team interaction to the investigation. Six surgical teams, each consisting of a general surgeon, anesthesiologist, surgical technician, and perioperative nurse, performed damage control surgeries under no motion conditions and under emulated deck motion conditions based on acceleration profiles from USS Freedom (LCS-1) at SS3 and USNS Spearhead (EPF 1) at SS4 (Fig. 2). These motions for LCS-1 equated to a mean significant wave height of 2.9 ft and a period range of 5.1 to 15.4 s at 20 knots. The motions for EPF 1 equated to a mean significant wave height of 6.17 ft with a period range of 6.1–16.2 s at 15 knots. These experimental parameters were chosen to investigate the upper limit for conducting damage control surgery while not interfering with surgical outcomes.

For Phase II, the motion platform was equipped with an anesthesia apparatus, portable oxygen monitor, instrument table, military field operating table, vital sign monitor, and two surgical lights. During the simulation, video cameras were positioned to capture activities of team members during

FIGURE 2. USS freedom (LCS-1) and USNS spearhead (EPF 1). The yellow line is indicative of the location used for programming of the simulated deck motions.

procedures such as manual dexterity, movement of surgical tools, and operation of the anesthesiology machine. The Observer XT by Noldus was used to observe and quantify events via real-time video data tagging. SMEs presented the cases to the teams via a teleconference link. They provided vital signs when requested, observed, and evaluated all procedures. These experts included three board-certified general surgeons, one board-certified emergency room physician and one board-certified anesthesiologist.

Following Institutional Review Board (IRB) approval by U.S. Navy Experimental Diving Unit, participants were recruited from multiple active medical commands and were assembled into the six surgical teams. Each surgical team participated in 1 week of experimentation. It became apparent during initial planning with the surgical SMEs, that there are a variety of correct techniques for performing damage control surgeries. Day one familiarized each team with a standard way of performing each procedure to ensure performance outcomes were comparable between groups. Days 2–4 were experimental while the fifth day was used to make up missed procedures. Each team completed eight procedures on the platform during each day, spending 1 day testing at SS0 (stationary platform), 1 day at SS3 and 1 day at SS4. Forty-eight surgeries were performed under each motion condition, 144 total. Order of surgical procedures was selected randomly from within the days' planned procedures. At the conclusion of each procedure, the teams remained in the operating room on the simulator while they completed the Task Load Index (TLX) survey. They could then disembark the platform for a 20-min rest between procedures and were permitted to eat and drink. SME evaluation of surgical performance occurred between simulations. After each testing day, the participants could return to their quarters. The use of caffeine, tobacco, home medications, and other consumed performance enhancers was not regulated in this study. They were permitted access to any food or medication that they would typically have aboard ship.

Phase III was performed shipboard during a 15-day voyage (February 2017) of the USNS Brunswick (Fig. 3) from Joint Base Little Creek-Fort Story, Virginia to Joint Base San Diego, California via the Panama Canal. The USNS Brunswick is the sixth Spearhead-class expeditionary fast transport catamaran and was selected to allow for a class-comparison to phase II SS4 measurements. This vessel is a high-speed catamaran and is suitable for use in littoral or deep-sea environments. The time of year and route of the Brunswick were selected based on historical National Oceanic and Atmospheric Administration's (NOAA) National Data Buoy Center data to provide similar conditions to those in phase II. Surgical teams remained aboard the Brunswick for the entirety of the voyage. The same participants from Phase II performed simulated surgical procedures aboard the Brunswick while experiencing real deck motions. Seaway data was collected throughout the voyage through direct observation and incorporation of forecast data and National Oceanic and Atmospheric Administration seaway buoy data. SSs throughout the voyage ranged from NATO SS1 to SS4 (Fig. 4).

Surgical interventions were performed in an operating room assembled in the personnel bay just anterior to the midline of the vessel (as indicated by the yellow line on Fig. 3). Placement was to take advantage of the portion of the vessel that would experience the largest vertical accelerations, thereby creating a "worst case scenario", adjusting for variability in the placement of surgical modules on other vessels. Also, this area of the ship was directly above the location where acceleration from the comparable USNS Spearhead was collected for purposes of programming the simulator with motion parameters in Phase II.

Phase III used two surgical teams, composed of a general surgeon, Certified Nurse Anesthetist (CRNA), perioperative nurse, corpsman, and two surgical technicians. An auxiliary team of one general surgeon, one CRNA, and one corpsman were on board throughout the voyage in case of illness or injury. The auxiliary team was rotated in to participate as part of the surgical teams during the second week to address team complacency and team adjustment. No dropouts occurred. Participants could consume caffeine, medications, and tobacco, since medical personnel would normally be allowed to do so during deployment.

The roles of each participant in the scenarios were fixed and defined. The surgeon was at the patient's side throughout the scenario and was responsible for medical treatment and assessment. The CRNA was at the patient's head, responsible for anesthesia, airway, and other assessments. The nurse had limited mobility and responsibility for managing overall workflow and circulation. Tech 1 and Tech 2 were mobile and responsible for supporting tasks. The Corpsman was mobile and responsible for supporting tasks. SMEs provided instructions, vital signs, and scenarios.

Resuscitative procedures were added to the methods during phase III to better evaluate medical interventions that would have to be performed in managing patients with these

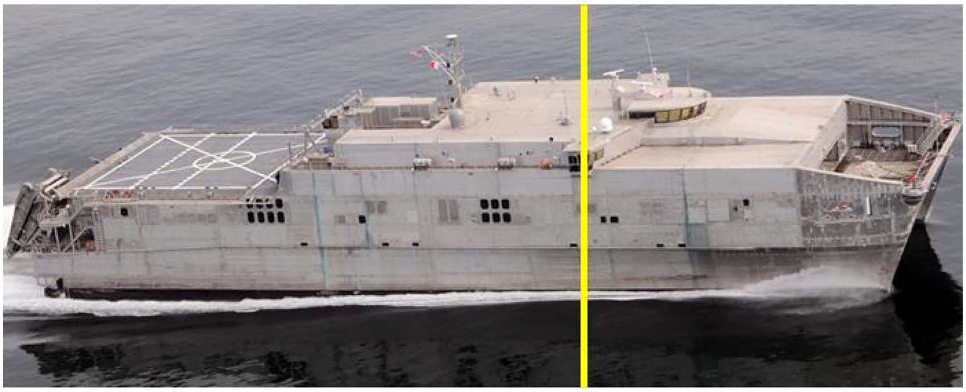
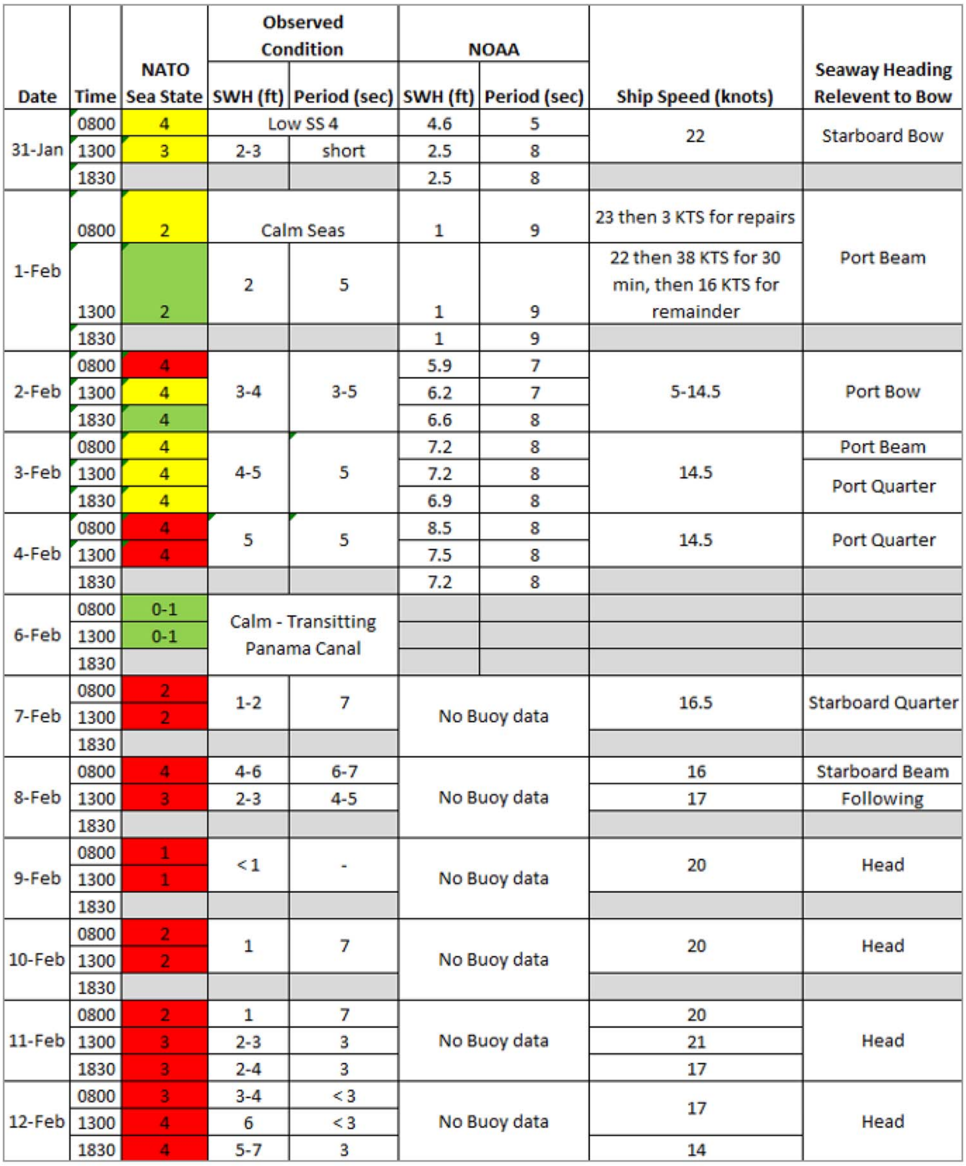


FIGURE 3. USNS Brunswick (T-EPF-6). The yellow line is indicative of the location where the operating room was assembled and motions were recorded.



Date	Time	NATO Sea State	Observed Condition		NOAA		Ship Speed (knots)	Seaway Heading Relevent to Bow
			SWH (ft)	Period (sec)	SWH (ft)	Period (sec)		
31-Jan	0800	4	Low SS 4		4.6	5	22	Starboard Bow
	1300	3	2-3	short	2.5	8		
	1830				2.5	8		
1-Feb	0800	2	Calm Seas		1	9	23 then 3 KTS for repairs	Port Beam
	1300	2	2	5	1	9	22 then 38 KTS for 30 min, then 16 KTS for remainder	
	1830				1	9		
2-Feb	0800	4	3-4	3-5	5.9	7	5-14.5	Port Bow
	1300	4			6.2	7		
	1830	4			6.6	8		
3-Feb	0800	4	4-5	5	7.2	8	14.5	Port Beam
	1300	4			7.2	8		Port Quarter
	1830	4			6.9	8		
4-Feb	0800	4	5	5	8.5	8	14.5	Port Quarter
	1300	4			7.5	8		
	1830				7.2	8		
6-Feb	0800	0-1	Calm - Transitting Panama Canal					
	1300	0-1						
	1830							
7-Feb	0800	2	1-2	7	No Buoy data		16.5	Starboard Quarter
	1300	2						
	1830							
8-Feb	0800	4	4-6	6-7	No Buoy data		16	Starboard Beam
	1300	3	2-3	4-5			17	Following
	1830							
9-Feb	0800	1	< 1	-	No Buoy data		20	Head
	1300	1						
	1830							
10-Feb	0800	2	1	7	No Buoy data		20	Head
	1300	2						
	1830							
11-Feb	0800	2	1	7	No Buoy data		20	Head
	1300	3	2-3	3			21	
	1830	3	2-4	3			17	
12-Feb	0800	3	3-4	< 3	No Buoy data		17	Head
	1300	4	6	< 3				
	1830	4	5-7	3			14	

FIGURE 4. Buoy and sea data.

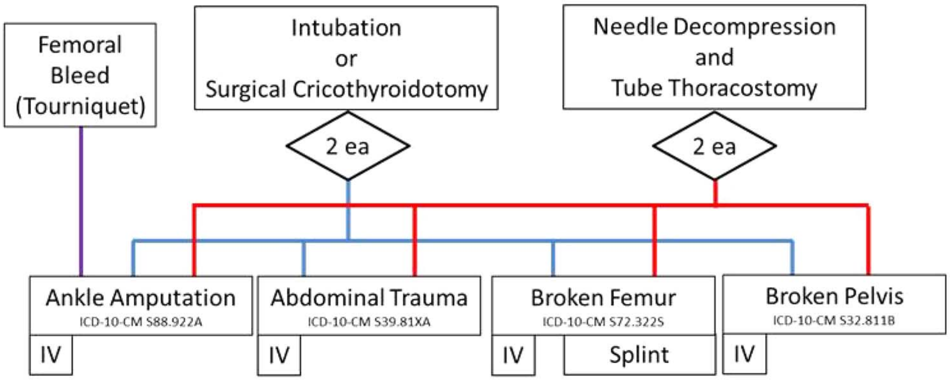


FIGURE 5. Patient flow diagram used to control potential confounders of the order of procedures by randomizing order.

injuries. Resuscitative procedures included tourniquet application, intubation, surgical cricothyrotomy, needle decompression, tube thoracostomy, intravenous catheter insertion, and splinting. The flow from resuscitative procedure to surgical intervention was determined by the patient flow diagram (Fig. 5). An across subjects counterbalancing design was used to control for order effects (the possibility that the position of treatment in the order of treatment matters) and sequence effects (the possibility that a treatment will be affected by the treatment proceeding it). So, although randomly presented, all four damage control surgical procedures and all seven resuscitative procedures were repeated with each block of four surrogate patients. By day, the teams were separated, so one team performed only resuscitative procedures while the other team performed only surgical procedures on a given day, for direct comparison with phase II. Only the damage control surgical procedures that were common between Phase II and Phase III are compared in this article.

In Phase II and III, surgical teams performed simulated surgical procedures on realistic surgical models call the Cut Suit (Strategic Operations, San Diego, California) (Fig. 6). The Cut Suit is an anatomically correct surgical trainer that has been utilized as a trainer for military personnel and has been validated in numerous studies.[7,8,9] The Cut Suit includes bones, perfused organs, muscle, and skin. For the abdominal trauma and unstable pelvic simulation, the Cut Suit was placed on a surgical mannequin, providing a realistic simulated operative experience. A Rescue Randy (Simulaids, Saugerties, New York) was attached to Cut Suit lower extremities for femur fixations and amputations. Pressurized simulated blood was circulated through the Cut Suit with a battery-powered pump.

The surgical procedures for Phases II and III were selected from the Patient Condition Occurrence Frequency (PCOF) tool that has been accredited for use by the DoD medical planners, material developers, clinicians and logisticians to manage combat, humanitarian assistance, and disaster relief PCOF distribution tables to develop clinically-based planning estimates across the range of military operations.[10] The PCOF table lists International Classification of Diseases codes and their associated probability distribution for each patient type (eg, wounded in action, disease, and nonbattle injury). The four surgical procedures were selected for the highest frequency of occurrence in theater. Expectations for surgical intervention for each injury were clearly described (Table 2).

FIGURE 6. The Cut Suit.

Surgical TLX scores measured cognitive workload during medical interventions. This TLX is based on the National Aeronautics and Space Administration TLX and obtains an overall workload score based on six categories: mental demands, physical demands, temporal demands, task complexity, situational stress, and distractions. This survey has previously been validated as an accurate measure of cognitive workload during the completion of surgical tasks.[11] Electroencephalographic monitoring was also utilized in phase II to measure cognitive workload, but this method was eliminated in phase III because of logistical difficulty.

TABLE 2. Expectations for Surgical Interventions

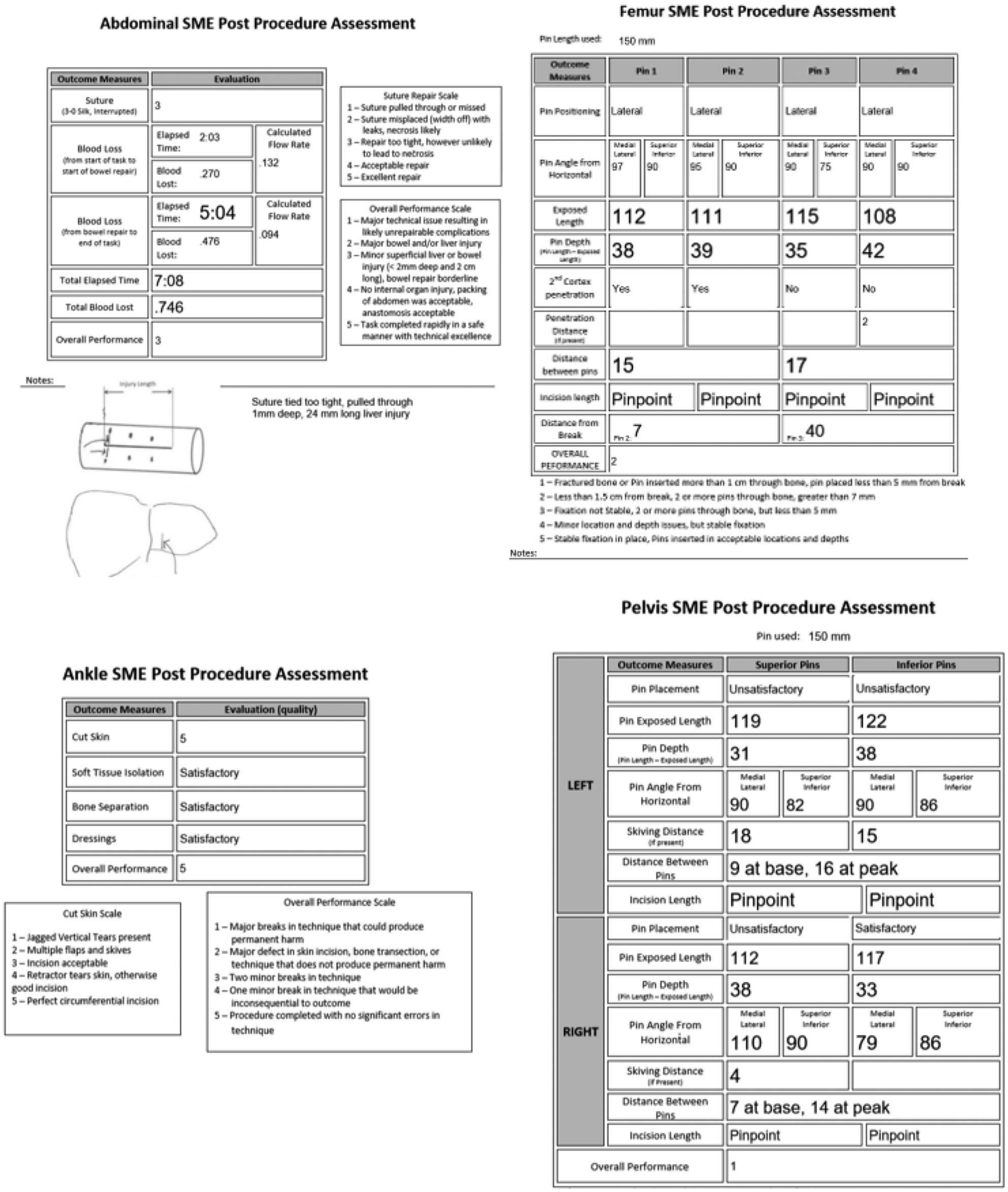
ICD Code	Injury	Surgical Interventions
ICD-10-CM S39.81XA	Other specified injuries of the abdomen	Surgical preparation Confirm anesthesia is ready Incision into the abdomen Packing Start injury repair End injury repair Begin closure End surgery
ICD-10-CM S32.811B	Multiple fractures of pelvis with unstable disruption of the pelvic ring	Request X-Ray Confirm anesthesia is ready Surgical preparations Insert the first pin Insert the second pin Insert the third pin Insert the fourth pin Stabilizer placement Rod placement Vascular re-evaluation Dressing placement End surgery
ICD-10-CM S72.322S	Displaced transverse fracture of shaft of the femur	Request X-Ray Confirm Anesthesia is ready Vascular control Surgical preparation Insert the first pin Insert the second pin Insert the third pin Insert the fourth pin Stabilizer placement Rod placement Vascular re-evaluation Dressing placement End surgery
ICD-10-CM S88.922A	Partial traumatic amputation of the lower leg	Request X-Ray Confirm anesthesia is ready Surgical preparation Amputation begins Skin cut Vascular control Bone separation Cut first bone Cut second bone Re-evaluation of vascular control Dressings End surgery

Therefore, electroencephalographic results are not discussed in this article.

SME evaluations were used as measures of surgical performance. Evaluations were administered using standardized grading sheets to assess the procedures performed. Criteria assessed included the effectiveness of techniques, efficiency of performance, and precision of interventions. SME scoring included ranking of performance for subjective measures and numerical grading of objective measures like screw depth and lengths between anastomotic sutures (Fig. 7). Final scores for each task were aggregated into a five-point Likert scale score. Net scores of 4 or 5 were considered satisfactory.

Statistical Analyses

The initial analysis examined whether the team combinations in Phase III could be treated as a homogenous population. Since each of the teams was not exposed to every SS condition, but all were exposed to SS4, this assessment focused on SS4 conditions. This was the largest data set, (46% of patients were treated in SS4), and SS4 was expected to have significant



Abdominal SME Post Procedure Assessment

Outcome Measures	Evaluation	
Suture (3-0 Silk, Interrupted)	3	
Blood Loss (from start of task to start of bowel repair)	Elapsed Time: 2:03	Calculated Flow Rate .132
	Blood Lost: .270	
Blood Loss (from bowel repair to end of task)	Elapsed Time: 5:04	Calculated Flow Rate .094
	Blood Lost: .476	
Total Elapsed Time	7:08	
Total Blood Lost	.746	
Overall Performance	3	

Suture Repair Scale
1 – Suture pulled through or missed
2 – Suture misplaced (width off) with leaks, necrosis likely
3 – Repair too tight, however unlikely to lead to necrosis
4 – Acceptable repair
5 – Excellent repair

Overall Performance Scale
1 – Major technical issue resulting in likely unrepairable complications
2 – Major bowel and/or liver injury
3 – Minor superficial liver or bowel injury (< 2mm deep and 2 cm long), bowel repair borderline
4 – No internal organ injury, packing of abdomen was acceptable, anastomosis acceptable
5 – Task completed rapidly in a safe manner with technical excellence

Notes:

Suture tied too tight, pulled through 1mm deep, 24 mm long liver injury

Femur SME Post Procedure Assessment

Pin Length used: 150 mm

Outcome Measures	Pin 1	Pin 2	Pin 3	Pin 4
Pin Positioning	Lateral	Lateral	Lateral	Lateral
Pin Angle from Horizontal	Medial Lateral 97 Superior Inferior 90	Medial Lateral 95 Superior Inferior 90	Medial Lateral 90 Superior Inferior 75	Medial Lateral 90 Superior Inferior 90
Exposed Length	112	111	115	108
Pin Depth (Pin Length – Exposed Length)	38	39	35	42
2nd Cortex penetration	Yes	Yes	No	No
Penetration Distance (if present)				2
Distance between pins	15		17	
Incision length	Pinpoint	Pinpoint	Pinpoint	Pinpoint
Distance from Break	Pin 2: 7		Pin 3: 40	
OVERALL PEFORMANCE	2			

1 – Fractured bone or Pin inserted more than 1 cm through bone, pin placed less than 5 mm from break
2 – Less than 1.5 cm from break, 2 or more pins through bone, greater than 7 mm
3 – Fixation not Stable, 2 or more pins through bone, but less than 5 mm
4 – Minor location and depth issues, but stable fixation
5 – Stable fixation in place, Pins inserted in acceptable locations and depths

Notes:

Ankle SME Post Procedure Assessment

Outcome Measures	Evaluation (quality)
Cut Skin	5
Soft Tissue Isolation	Satisfactory
Bone Separation	Satisfactory
Dressings	Satisfactory
Overall Performance	5

Cut Skin Scale
1 – Jagged Vertical Tears present
2 – Multiple flaps and skives
3 – Incision acceptable
4 – Retractor tears skin, otherwise good incision
5 – Perfect circumferential incision

Overall Performance Scale
1 – Major breaks in technique that could produce permanent harm
2 – Major defect in skin incision, bone transection, or technique that does not produce permanent harm
3 – Two minor breaks in technique
4 – One minor break in technique that would be inconsequential to outcome
5 – Procedure completed with no significant errors in technique

Pelvis SME Post Procedure Assessment

Pin used: 150 mm

	Outcome Measures	Superior Pins	Inferior Pins
LEFT	Pin Placement	Unsatisfactory	Unsatisfactory
	Pin Exposed Length	119	122
	Pin Depth (Pin Length – Exposed Length)	31	38
	Pin Angle From Horizontal	Medial Lateral 90 Superior Inferior 82	Medial Lateral 90 Superior Inferior 86
	Skiving Distance (if present)	18	15
	Distance Between Pins	9 at base, 16 at peak	
	Incision Length	Pinpoint	Pinpoint
RIGHT	Pin Placement	Unsatisfactory	Satisfactory
	Pin Exposed Length	112	117
	Pin Depth (Pin Length – Exposed Length)	38	33
	Pin Angle From Horizontal	Medial Lateral 110 Superior Inferior 90	Medial Lateral 79 Superior Inferior 86
	Skiving Distance (If Present)	4	
	Distance Between Pins	7 at base, 14 at peak	
	Incision Length	Pinpoint	Pinpoint
Overall Performance		1	

FIGURE 7. Sample SME grading forms.

impact on performance. Independent sample *t*-tests confirmed homogeneity among teams. Following assessment of homogeneity, to assess the between- and within-group effects on perceived workload, a within-subjects general linear model was used to assess the effects of SS, procedure type, and number of simulations completed on the perceived workload of participants. Results were deemed significant if $\alpha = 0.05$, $P < 0.05$. If statistical significance was found, post hoc analyses were completed using independent samples *t*-tests to determine the effect.

TABLE 3. Comparison of Motions Experienced in Phase II and Phase III

		Stewart Table	USNS Brunswick
Sea States Experienced		SS0, SS3, SS4	SS1, SS2, SS3, SS4
Motion period		0.1 Hz to 0.25 Hz	0.1 Hz to 0.25 Hz
Periodicity		5-16 seconds	3.5-10 seconds
Pitch	Average	2 degrees	3.6 degrees
	Peak	1.006	8.8
Roll	Average	7 degrees	10 degrees
	Peak	1.13	15.2

To assess factors that influence performance patient outcome across SS, an ordinal logistic model was created with factors of surgeon, procedure type, SS, number of patients treated, and first-degree interaction terms. All analyses were performed using Minitab 17 (Minitab, LLC).

RESULTS

Phase II and Phase III Comparison

Motion Analysis

The Stewart Table replicated SS0, SS3, and SS4 during phase II. Monitoring of the motion base simulator revealed a typical motion period, regardless of SS, of 0.1–0.25 Hz with a periodicity of 5–16 s. Heave (vertical translational) accelerations were larger in amplitude than pitch, roll, surge, and sway accelerations.

The USNS Brunswick experienced SS1–SS4. Shipboard monitoring revealed an average roll motion of 10 degrees, with a maximum roll of 15.2 degrees. Peak pitch was 8.8 degrees. The average motion period (wave peak to wave peak) ranged from 0.1 Hz to 0.25 Hz with a periodicity of 3.5–10 s (Table 3). The USNS Brunswick experienced substantial rolling motions throughout Phase III of the experiment, likely because of its wide stance, as a catamaran. Additionally, pitch accelerations were substantial aboard the USNS Brunswick, while heave, surge, and sway accelerations were relatively small.

Measures of motion, including motion period and periodicity, were comparable across phase II and phase III in this comparison, as well. This shows that simulation with a motion base simulator approximates realistic deck motions and may be effective in replicating conditions at SS4. The small differences between the motions of the simulator and the Brunswick did not contribute to a significant difference in surgical performance.

Surgical Performance

In both phase II and phase III, participants were able to perform the majority of surgical procedures to receive a satisfactory score (4/5 or 5/5 on the SME grading form), despite the motion environments they experienced. Phase II included 144 surgical procedures. A total of 120 of these procedures were satisfactory, producing a success rate of 83.3%. At SS0, 41/48 procedures were satisfactorily completed (85.4%). At SS3, 38/48 procedures were satisfactorily completed (79.2%). At SS4, 41/48 procedures were satisfactorily completed (85.4%). Regardless of SS, the satisfactory completion rates by procedure type were: Abdominal-28/36 (77.8%), Ankle-34/36 (94.4%), Pelvis-30/36 (83.3%), Femur-27/36 (75%). Surgical performance was analyzed using a repeated-measures analysis of variance to determine, which variables impacted surgical performance, including SS, surgical team, and surgical procedure. SS did not significantly affect performance. Surgical procedures were the strongest predictor of performance ($\alpha = 0.05$, $P < 0.05$), with decreased performance being correlated with femur and pelvic procedures.

Correlation between the surgical team and procedure was also significant ($\alpha = 0.05$, $P < 0.05$), demonstrating significant differences in skill between teams.

Phase III included 112 surgical procedures. Of these, 100 procedures were satisfactory, the overall success rate of 89.2%. A single procedure received a score of 1. This procedure was the first procedure performed in phase III and was completed by a surgeon who was unfamiliar with it. The remaining 11 unsatisfactory procedures were graded 2 or 3. The ordinal regression model for factors including surgeon, procedure type and SS showed that the only significant predictor of score was surgeon/team performing the procedure ($\alpha = 0.05$, $P < 0.05$). This may reflect experience in the procedures or their adaptability to onboard conditions. However, interpreting this effect is also complicated by the surgeons' assignment to teams. Because the team rosters were shuffled between the first and second week, the performance of the three individual surgeons and the four unique surgical teams cannot be uniquely disentangled.

Surgical performance in both phases was analyzed by the ordinal regression model to determine which variables impacted surgical performance, including SS, surgical team, and surgical procedure. The factors of surgical procedure, SS, and their interaction poorly predicted performance scores; however, in both phases, the femur fracture procedures tended to be scored lower than the other procedure types. This may be indicative of less experience among the surgeons, or a more difficult grading rubric for femur fracture compared to other procedures.

Surgical performance between phase II and phase III was not significantly different, suggesting that the simulator did not create conditions that were significantly different enough to impact surgical performance for trained professionals. In phase III, the only significant predictor of surgical performance was the surgeon/surgical team performing a given procedure, suggesting disparate skill levels between teams or the inability of one team to adapt to the conditions of surgery at sea. Phase II data did not demonstrate this finding, which could be the result of the testing schedule for phase II that included 1 day each of SS0, SS3, and SS4 conditions. The short exposure to motion may have allowed the individuals to adapt and recover more uniformly. Comparatively, the 2-week long voyage of the USNS Brunswick exposed all participants to chronic deck motion, which may have adversely affected individual participants more significantly than others, thereby causing poorer individual performance. An alternative explanation of these findings would be that the individuals involved in the study simply had disparate skill levels from one another.

Although not significant, the data trends for phase II and III demonstrated poorer surgical performance for femur fixation compared to other procedures. In the regression model, it was the only procedure that predicted a lower score. This trend should be examined more closely as it may be the result of a lack of experience with this procedure among the participants or the result of a more difficult grading rubric of this procedure.

Workload

Overall in Phase II, TLX scores for surgeons were consistently higher than the other roles in all states. Increased TLX scores were associated with higher SSs ($\alpha = 0.05$, $P < 0.001$) but varied according to role within the team. When comparing SS3 and SS4 mean TLX scores to the SS0 scores, TLX scores were significantly higher for surgeons and lower for surgical techs in the motion conditions ($\alpha = 0.05$, $P = 0.002$). One team had significantly higher error rates than the others with a total of 36 errors (5 at SS0, 12 at SS3, and 19 at SS4). The mean number of errors per procedure for this team was 3 compared to other teams (0.67). The single team predominance of errors caused a bivariate correlation to reveal a positive correlation between workload and surgical errors ($\alpha = 0.05$, $P = 0.025$). This was not representative of the whole data trend. A second bivariate correlation was performed, excluding that team's data. The resulting trend demonstrated no significant correlation between workload and errors, which is representative of the actual data trend. To further examine the effect of workload, TLX scores were evaluated with a nonparametric ordinal logistic regression on performance scores. TLX was not a significant predictor of performance.

In phase III, the within-subjects general linear model for workload showed a slight trend toward increased TLX scores with the increasing SS. The effect was small, with an increase of approximately 2 points per SS, but was statistically significant ($\alpha = 0.05$, $P < 0.001$). However, the effects of SS were only observed at conditions up to SS4. Conditions above SS4 may have a stronger effect on workload, so extrapolation of this trend beyond SS4 is ill-advised without further data.

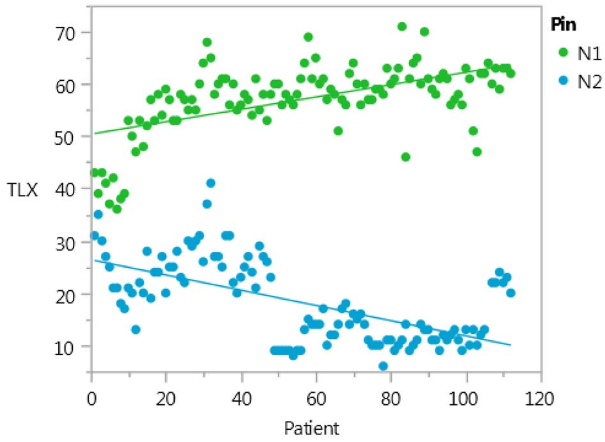


FIGURE 8. Diverging TLX trends between two nurses in phase III.

Within this model, the relationship between the TLX score and number of patients treated varied widely across participants. For some participants, TLX decreased with repetition (eg, nurse "N2", Fig. 8); this may reflect repetition of scenarios or acclimation to the environment. In contrast, other participants had increasing or nearly constant TLX scores over time (eg nurse "N1", Fig. 8), which may be indicative of fatigue or decreased motivation. However, interpretation of these results is complicated by reorganization of teams. Differences in dynamic and division of labor between teams could also affect perceived workload for individuals and these differences cannot be teased apart from experience.

There were significant differences between TLX scores for participants in different roles ($P < 0.0001$). On average, surgeons, nurses, and CRNAs reported higher TLX scores than the mean, while surgical technicians and corpsmen reported lower TLX scores.

Phase III SS4 results are compared with phase II simulator results in Figure 9. Though the phase III TLX scores for Surgeon, CRNA, and Technician roles are consistently larger than the phase II SS4 results, phase III results are within the range for those seen within phase II. The Nurse in phase III had consistently and statistically significantly greater scores ($\alpha = 0.05$, $P < 0.0001$) than simulator scores. Since the nurse has responsibility over personnel in the OR, increased task load may be attributable to the addition of two personnel (tech 2 and corpsman). This result, if confirmed with additional data, emphasizes the need to maintain teams to only those necessary for effective task flow.

Individual surgical procedures were not associated with any changes in perceived workload in either phase.

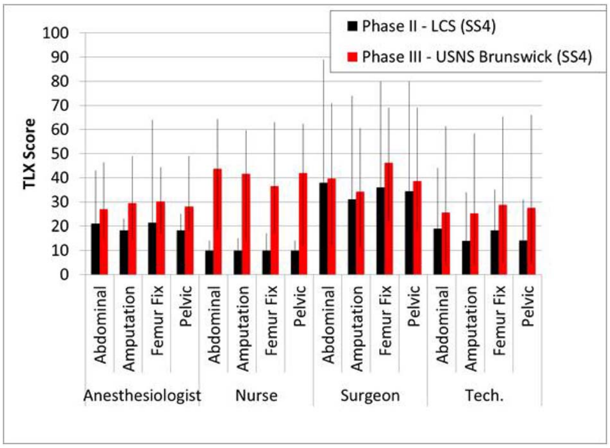


FIGURE 9. TLX scores by role, procedure, and phase at SS4.

TLX scores were significantly higher for all participants in phase III compared to phase II, suggesting increased cognitive load between the two phases. Given the comparable deck motions for the simulator and the Brunswick, it is unlikely that the simulated environment alone accounts for the difference. It is possible that the psychological impact of the simulated state compared to the actual voyage could have explained these results, as motion sickness, fear of the ocean, and other mental stressors could have confounded this finding. Another explanation for this difference is the increased team size from four personnel in phase II to six personnel in phase III. It is also possible that the cumulative time spent in motion (2 weeks at sea vs interrupted events over 3 days) could have confounded this data as well. These factors should be examined more closely in future work to help identify ways to improve the cognitive load during surgery at sea.

CONCLUSION

The SSs tested in this study ranged from SS1 to SS4. Both real and simulated motion constituted dramatic motion environment conditions, typical assumed frigate and destroyer conditions assumed to produce motion induced interruptions are about 5-degree rolls.[12] Despite the motions encountered, there was strong evidence that teams could perform damage control surgery without significant effects of SS on performance. This should be further examined for potential application to real operational scenarios as this study found no statistical evidence that increased SSs up to SS4 interfered with task completion.

This study provides evidence for the validity of the motion base simulator in replicating realistic deck motions. In both phases, TLX scores increased with SS, suggesting workload should be considered in the operational application of surgery at sea, particularly when high seas are encountered, as this may affect the health and abilities of personnel. In addition, the effects of motion on TLX suggest that training in a motion environment may be important for future medical operations at sea as an inoculation to anticipated operational stressors. Such training could use this simulator or training voyages aboard operational vessels.

The variance of TLX scores between roles suggests that limiting team size to necessary personnel is an important consideration in the application of surgery at sea. This should be studied further to determine if team size impacts cognitive workload for individuals and to further guide recommendations on team size.

Although there is no head-to-head comparison study with land-based treatment facilities or large deck platforms, this study provides evidence for surgical intervention aboard ships similar to the USNS Brunswick in the SSs studied. Future research in this field should focus on training for these missions, the effects of deck motion above SS4, and other damage control procedures. Additionally, studies should be expanded to include patient movement and transport.

ACKNOWLEDGMENTS

Bureau of Medicine USN; OPNAV N-81 Assessments Division, Medical Analysis Branch; Navy Advanced Medical Development; Naval Surface Warfare Center, PCD.

FUNDING

Phase I of this study was sponsored by the Office of Naval Research. Phase II was sponsored by the Office of the Chief of Naval Operations (OPNAV) N-81 Assessments Division, Medical Analysis Branch (N813). Phase III was sponsored by the OPNAV N-81 (N813) and Navy Advanced Medical Development (AMD).

REFERENCES

1. Remick KN, Dickerson JA, Nessen SC, Rush RM, Beilman GJ: Transforming US Army Trauma Care: an evidence-based review of the trauma literature. US Army Med Dep J 2010; 4–21. Available at. http://cdm15290.contentdm.oclc.org/cdm/ref/collection/p15290coll3/id/1148; accessed June 24, 2019.
2. Nealieigh MD, Kucera WB, Bradley MJ, et al: Surgery at sea: exploring the training gap for isolated military surgeons. J Surg Educ 2019; 76(4): 1139–45. Available at https://www.researchgate.net/publication/332161654_Surgery_at_Sea_Exploring_the_Training_Gap_for_Isolated_Military_Surgeons; accessed June 24, 2019.
3. Pierce E, Rapada R, Herder PA, et al: Surgery at sea: the effect of simulated High Sea states on surgical performance. J Spec Oper Med 2018; 18(2): 64–70. Available at. https://www.semanticscholar.org/paper/Surgery-at-Sea%3A-The-Effect-of-Simulated-High-Sea-on-Pierce-Rapada/242920de643b3e7f3a91bc73e5632ac44dd60ed8; accessed June 24, 2019.
4. Alarhayem AQ, Myers JG, Dent D, et al: Time is the enemy: mortality in trauma patients with hemorrhage from torso injuries occurs long before the “golden hour”: Southwestern surgical congress, Am J Surg 2016; 212(6):1101–5. Available at https://www.sciencedirect.com/science/article/abs/pii/S0002961016305542; accessed June 24, 2019.
5. Nieuwenhuizen FM, van Paassen MM, Stroosma O, Mulder M, Bulthoff HH: Cross-platform validation for a model of a low cost Stewart platform. J Model Simul Identif Control 2013; 1: 1–23. Available at https://pdfs.semanticscholar.org/71b5/f9797a704be2294c13b22e71cf0f46b1b982.pdf; accessed June 24, 2019.

6. Haddon S: Sea state and swell metService. Sea and Swell August 3, 2015. Available at. https://blog.metservice.com/sea-state-and-swell; accessed June 24, 2019.
7. Hunt B, Wall V, LaPorta AJ, et al: New methods of early surgical training using the human worn partial task surgical simulator in scenario based stress immersion training. J Med Simul 2012; 4(3). Available at. https://issuu.com/jsomwinter2012/docs/jsom_winter_2012; accessed June 24, 2019.
8. LaPorta AJ, Hutchinson C, Robinson D, et al: The use of hyper-realistic surgical simulation. I/ITSEC Proceedings Interservice/Industry Training, Simulation, and Education Conference, Orlando, FA 2015. Available at https://authorzilla.com/9XAgo/2015iitsecprogramguidelores.html; accessed June 24, 2019.
9. Lateef F: Training for emergency department thoracotomy. Global J Emerg Med 2018; 1(2): 1–5. Available at. http://www.remedypublications.com/global-journal-of-emergency-medicine/articles/pdfs_folder/gjem-v1-id1006.pdf; accessed June 24, 2019.
10. Jackson C, Brown C, Pang G: Patient Condition Occurrence Frequency (PCOF) Tool Version 1.0.0.0 User Guide and Technical Report. Defense Technical Information Center 2013. Available at https://apps.dtic.mil/dtic/tr/fulltext/u2/a619244.pdf; accessed June 24, 2019.
11. Wilson MR, Poolton JM, Malhotra N, Ngo K, Bright E, Masters RSW: Development and validation of a surgical workload measure: the surgery task load index (SURG-TLX). World J Surg 2011; 35(9): 1961–9. Available at. https://www-ncbi-nlm-nih-gov.proxy.rvu.edu/pmc/articles/PMC3152702/; accessed June 24, 2019.
12. Graham R: Motion induced interruptions as ship operability criteria. Naval Eng J 1990; 102: 65–71. Available at. https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1559-3584.1990.tb02556.x; accessed June 24, 2019.