The Effect of Fatigue on Cognitive and Psychomotor Skills of Surgical Residents

Kanav Kahol[1,2], Mark Smith[1], Stephanie Mayes[1], Mary Deka[1], Vikram Deka[1], John Ferrara[1], and Sethuraman Panchanathan[2]

[1] Phoenix Integrated Surgical Residency Program, SimET Center
Banner Good Samaritan Medical Center, Phoenix Arizona
[2] Arizona State University

Abstract. Surgical residents are exposed to a significant amount of cognitive load during call. While various efforts have been made to quantify the effect of fatigue and sleep deprivation on the psychomotor skills of surgical residents, there is very little investigations into the effect of these factors on cognitive skills. However, this is an important issue in medical curriculum design, as much of the medical errors are procedural in nature and are not psychomotor. In this paper, we present a study that aimed to quantify the effect of fatigue on cognitive skills. We employed hand movement data for developing a proficiency measure of surgical skill. The difference in proficiencies measured through hand movement post call and pre call was determined. The simulation tasks were designed to challenge working memory, attention of the user. The results showed a significant difference in hand movement proficiencies as well as behavioral errors pre and post-call. EEG Data was also gathered during simulation tasks pre and post call through the B-Alert® Bluetooth EEG technology. The B-Alert® software was analyzed to reveal ratings of alertness/drowsiness, engagement, mental workload and distraction. The results showed statistically significant difference in EEG ratings in pre call and post call condition.

1 Introduction

In 1999 the Institute of Medicine published the now famous study “To Err is Human” on the unacceptably high rate of errors in medicine, and that has generated even more demand for changes in the status quo in healthcare. As a result, many of the prevalent approaches in medical field are being questioned. One of the burning questions is the effect of fatigue and sleep deprivation on the skills of residents. Residency programs require medical residents to perform various kinds of medical duties over extended periods of 16-24 hours. These extended hours of operation can significantly influence a doctor’s ability to perform. This factor can be even more critical in surgical procedures that require both cognitive and motor skills. Two decades ago minimally invasive surgery was introduced, and along with it came a more extensive and lengthy learning curve. More complex procedures are being continually developed and introduced, accompanied by more serious risks and complications. These procedures require a

D.D. Schmorrow, L.M. Reeves (Eds.): Augmented Cognition, HCII 2007, LNAI 4565, pp. 304–313, 2007.

significant amount of cognitive as well as psychomotor acuity. It is important to measure the effect of fatigue on these skills. This will provide guidelines on development of residency programs and how to handle fatigue and sleep deprivation.

Many systems have been proposed to study the effect of fatigue, sleepiness, noise and interruptions on surgical proficiency. These systems rely on existing measures of surgical proficiency to measure the effect of operating conditions on skill. Eastridge et al. (Tendulkar et al., 2005) used the MIST VR system for analyzing the effect of sleep deprivation due to calls on the technical dexterity of residents. Thirty-five surgical residents were prospectively evaluated pre-call (rested), on-call (rested), and post-call (acutely sleep deprived). Participants completed questionnaires regarding sleep hours and level of fatigue. Technical skill was assessed using the MIST VR. Speed, errors, and economy of motion were automatically recorded by the MIST VR system. Data were analyzed by pairing Student t-test and analysis of variance. They showed that Call-associated sleep deprivation and fatigue are associated with increased technical errors in the performance of simulated laparoscopic surgical skills. While studies on fatigue, sleep deprivation have been performed, these studies have relied on existing measures of surgical proficiency. Existing measure of surgical proficiencies include simple measure of proficiency such as time, tool movement smoothness etc. These measures are not sufficient to quantify effect of fatigue comprehensively. In the first section of the paper we define a system for measuring hand movement based proficiency. We then describe an experiment with the developed apparatus to quantify the effect of fatigue.

2 Hand Movement Based Proficiency System

Surgical movements such as suturing, knot tying, etc are composed of a few basic movements herein referred to as gestures or basic gestures. These gestures form the building block of most complex surgical procedures. The advantage of deconstructing a complex motion into simplex units lies in facilitating understanding of the complex movements and ease of communication while training. It also allows for detailed evaluation of surgical procedures and pinpointing where the resident is making errors and the type of errors. The overall approach of proficiency analysis relied on recognizing the accuracy of individual basic movements and combining the proficiency measures of individual basic gestures that form a complex movement pattern like suturing.

In order to develop proficiency system, simulation tasks were chosen. The complex surgical procedure that was chosen for analysis involved picking up a linear object with a probe, passing the object through an opening, grasping it with the probe with the other hand, and then passing it back through a vertical opening. The procedure was simulated on a Simuvision® Simulator. Simuvision is a portable laparoscopic video trainer that has a camera and two surgical probes plugged into a laptop. The view of the surgical pad was provided on the laptop and it simulates the work environment in minimally invasive surgery. We chose to de-construct complex movements into 9 basic gestures: (i) In movement (ii) Out movement (iii) Left movement (iv) Right movement (v) Up movement (vi) Down movement (vii) Clockwise Rotation (viii) Counter-clockwise Rotation (ix) Grasping (closing and opening both). These elemental movements were chosen based on data from the literature that suggests that complex gestures in minimally invasive surgery is a combination of these basic movements (Rosen, J.D. Brown, L. Chang, M. Sinanan, & Hannaford, 2006). The complex movement selected spans the entire range of the 9 gestures.

In addition to this complex motion, a simulation was designed for virtual ring transfer task. In the virtual ring transfer task, residents were tasked with grasping a series of a “virtual” rings and placing each on randomly highlighted pegs on a board. The simulation was implemented using the Sensable® haptic joystick. The Sensable haptic joystick allows for generation of 3 degrees of force feedback in response to events in the virtual environment. OpenHL® programming API was used to design the simulation. The simulation allows for measurement of tool tip in the virtual environment. Additionally the sessions could be played back in the software with traces of tool tip movement being shown at various speeds. This allowed for visual analysis of movement.

The basic task involved 10 rings. After the participant places a ring on a highlighted peg, another peg is randomly chosen for participant to the put the ring on. This is repeated till all the 10 rings are correctly placed. The time taken for completing the task is displaced on the bottom middle part of the screen for the participant to follow.

In addition to the basic ring transfer task three variations of the task were designed. In the basic version, the ring is stationary and can be grasped by picking it from a fixed location. To include more manipulations and greater range of motion, a variation of the basic simulation was designed where the ring moved slowly in the environment. This variation required the participants to track the ring movement and pick it while it is in motion. In this variation, the ring moved in the plane of the peg board. In a second variation, the ring was allowed to move anywhere in a 3D environment. In the third variation, the entire peg board moves slowly, creating different orientations in which the surgeons need to accomplish the task. These tasks trained the user to accomplish the tasks in different orientations, a common requirement in actual surgical environments. In addition it also allowed for capture of rotation movements of the wrist. Overall there were 4 different types of ring transfer tasks.

2.1 Participants

Table 1 shows the participant groups included in the initial study.

Table 1. Participants Groups in the initial study

	Senior Surgeons	*1st year residents*	*2nd Year Residents*	*3rd Year Residents*
OBGYN: Male	1	1	4	2
OBGYN: Female	0	5	5	2
General Surgery: Male	2	3	4	2
General Surgery: Female	1	2	4	2
Total	**4**	**11**	**17**	**8**

2.2 Data Capture Methodology

There were 2 different types of capture sessions. The first type of sessions involved basic movements with no specific simulation task. The participants initially filled in a demographic questionnaire which documented their age group, sex, ethnicity and handedness. In addition a laparoscopic experience questionnaire was employed to gauge the overall experience of the participant. Both the questionnaires were filled by the user before the task. In this type of capture sessions, participants performed the 9 basic gestures and the complex surgical procedure using the Simuvision® Simulator. The collection of data was randomized across subjects in a counter-balanced fashion. While performing surgical gestures, the participants were requested to wear CyberTouch® Gloves with a Flock of Bird® Tracker attached to the back of the gloves. The glove required to be calibrated to a users' hand. This was done by the procedure highlighted in the Cyberglove software that requires users to mimic certain poses shown on the computer screen while wearing the gloves. These poses are employed by the software to calibrate the position of joints and measurement points on the glove as suited to a user's hand. The software allowed for saving the calibration and for subsequent sessions the saved calibration for every user was loaded.

CyberTouch datagloves are capable of measuring hand movement (angles at various joints). The gloves use bend sensing technology to measure angles. We measured both wrist and digit kinematics. For wrist kinematics, we recorded flexion/extension and adduction/abduction angles. Digit kinematics consisted of: angles at the metacarpal–phalangeal (mcp), proximal and distal interphalangeal (pip and dip, respectively), joints of the four fingers and the angle of abduction (abd) between adjacent fingers. For the thumb, the mcp, abd, and interphalangeal (ip) angles were measured together with the angle of thumb rotation (rot) about an axis passing through the trapeziometacarpal joint of the thumb and index mcp joint.

We also measured palm arch, minimum and maximum values being associated with the thumb and little finger either touching or being maximally separated. Flexion and abduction were defined as positive; the mcp and pip angles were defined as 0° when the finger are straight and in the plane of the palm. At the thumb, positive values of thumb rotation denoted internal rotation. The spatial resolution of the CyberGlove is ~0.1°. Fingertip and hand 3-D position and orientation are measured by magnetic trackers (Flock of Birds). The sampling rate of the CyberTouch and magnetic trackers was 120 Hz each. The gloves were open at the fingertips and hence they did not interfere with tactile perception.

In addition to the 3D data, video data was gathered for documentation. Three video cameras were employed. The first video camera stream was obtained from the camera on the Simuvision® simulator. This camera captured the tool movements. Two cameras focused on the two hands for capturing the movements. These video streams were synchronized with the 3D data streams. Each surgeon performed the entire set of gestures. Twenty trials for each basic gesture were performed. Ten trials for each complex gesture were captured as it contained enough repetitions of basic gestures for analysis.

In the second type of data capture, the ring transfer task was performed. In general the ring transfer task covered all the 9 basic gestures and the task served as an appropriate procedure to evaluate proficiency of users in accomplishing the task. The software designed for ring transfer recorded time taken to complete simulations and hand movements from the Cybergloves® and Flock of birds.

The following was the data capture methodology for the ring transfer tasks. The participants initially filled in a demographic questionnaire which documented their age group, sex, ethnicity and handedness. Then the participants were requested to fill a questionnaire that assessed their fatigue level. The questionnaire designed by Behrenz et al. (Behrenz & Monga, 1999) was used for this purpose. This questionnaire allowed for gauging the beginning stress levels and fatigue levels of the residents when doing the task. In addition a laparoscopic experience questionnaire was employed to gauge the overall experience of the participant. All the questionnaires were filled by the user before the task. The user was then requested to wear the Cyberglove®. In the first capture session for each user, the Cyberglove had to be calibrated to a user's hand. Cyberglove was calibrated and the calibration was saved for each user as per the procedure documented above. For subsequent sections, the saved calibration was loaded and employed for measurements.

After wearing the gloves, the participants were randomly assigned a ring transfer task to perform. Data was collected randomized across subjects in a counter balanced fashion. As mentioned above, there were 4 different types of task. In each capture session, the participants performed 2 repetitions of each of the 4 ring transfer tasks. The repetitions allowed for compensation of lack of familiarization with the simulations within a session. For 4 weeks, subjects were requested to perform the tasks before their call and then return and perform the tasks after call. Hence overall, there were overall 8 capture sessions for each user with 4 capture sessions being pre-call and 4 capture sessions being post-call. This methodology enabled gathering data on the effect of fatigue and sleep loss on surgical skills. Additionally, capturing 4 sessions enabled analysis of learning on surgical skill development.

2.3 Data Capture Methodology

The basic movement capture sessions on the Simuvision simulator were recorded through three video cameras. The senior surgeons involved in the experiments viewed the three video streams. The complex movement was divided into basic gestures by senior surgeons. The senior surgeons viewed all the data capture sessions. The corresponding basic gestures in the 3D data stream were identified (the streams were synchronized in capture). The senior surgeons then rated each individual gesture for their subjective measure of proficiency. Each basic movement was rated on a scale of 0 through 10, 0 being the least proficient and 10 being the most proficient. Fractional values were not permitted. The identity of the participant as well as time of capture was not revealed to the senior surgeon preventing any bias in the ratings. Each senior surgeon rated all the performed data sessions except the sessions performed by the senior surgeon themselves. Both the hand movements were rated for proficiency separately as each gesture in the approach was not bimanual in nature. In analysis, left hand and right hand were treated equally as the investigators deemed ambidexterity to be a desirable feature of surgical training. Surgeons were requested to rate proficiency based on visual appearance of the basic gesture as well as the behavioral effect it would produce if conducted in a real surgical environment. For example, a very fast in movement may be detrimental and should get low proficiency rating. The ring transfer data capture sessions were segmented into individual gestures and rated for their proficiency by viewing playback of the virtual simulation in the software. The hand movements were also played back synchronously through visualization software. The same rating system and scale was employed as detailed above.

The next step in data analysis was to find kinematic features for each gesture that depicted high correlation with the subjective measures of proficiency. For each gesture, a set of kinematic features was determined that showed a significant correlation ($p<0.05$) with the subjective proficiency ratings of gestures performed by all the groups. The combination of kinematic features that revealed statistically significant correlation with subjective proficiency measures of all samples of gestures (each gesture class is presented separately) performed by senior surgeons is shown in Table 2. For example, Table 2 shows that the thumb MCP angle value, wrist flexion and the wrist pitch had statistically high correlation with subjective measures. Similarly the thumb MCP angle, x and y component of the wrist, and the wrist flexion and wrist pitch together showed high correlation with rotation gesture. These measures showed high correlation with subjective ratings of the senior surgeon group. These kinematic features were employed as the basis to develop a model that could predict subjective ratings based on the set of kinematic features. Hence for example as shown in table 2, the thumb MCP angle, wrist flexion and wrist pitch were employed for developing predictive model for Grasping subjective proficiency ratings while for in gesture x,y,z component of wrist velocity was employed for development of the model.

Multiple regression analysis was employed to determine a linear model that predicted subjective proficiency ratings for a gesture based on kinematic features identified for the gesture as shown in Table 2. The subjective proficiency measures and the associated feature values were divided into a training set and testing set. The ratings were normalized to fit between a range of (0 – 1). The training set employed 70% of all the data capture sessions and the remaining 30% of the data capture sessions were testing set. Parameters were determined for a regression model for the confidence interval of 95% based on the training set between normalized proficiency ratings and predicted ratings. These parameters are shown in Table 2. The test data feature values were fed to the linear model determined by the regression analysis to determine the predicted values of subjective measures. The predicted values were multiplied by 10 to fit in a range of 0 through 10 and then rounded to determine the predicted subjective measure. Correlation coefficient between predicted measures and actual subjective measures was determined. In all the gestures, the correlation between predicted measures and actual subjective measure was above 0.89.

The developed regression models showed high construct validity and predictive validity. In addition, the developed computational measures of proficiency were sensitive to fluctuations in proficiency due to fatigue. This was established by the following methodology. Surgical proficiency of gestures as assigned by the senior surgeons was correlated with normalized fatigue ratings. The questionnaire included tasks in which participants had to rate their present level of fatigue and the maximum and minimum fatigue level during the past 24 hours on a scale of 0 through 10. The normalized rating was obtained taking the weighted average of the three ratings in the questionnaire with the current fatigue level given twice the weight of the minimum and maximum fatigue ratings. The normalized fatigue ratings showed statistically significant correlations ($p<0.1$) with predicted proficiency ratings and the subjective proficiency ratings obtained through senior surgeons. The developed proficiency measures served as a robust computational method to automatically judge proficiency of the performed movements by surgeons. They are based on de-constructing a complex motion sequence into simple gestures and analyzing the gesture for their proficiency through hand movements. This approach offers an easy to understand rating which surgical residents can employ to gauge their skill levels and improve their skills.

Table 2. Set of kinematic features that showed high correlation with subjective proficiency measures for each gesture. The regression coefficients for the linear model are also shown.

Movement	Thumb MCP Angle	Wrist Component Velocity			Aggregate Palm Angle	Wrist Adduction	Wrist Flexion	Wrist Pitch
		X	Y	Z				
Grasp	0.9	-	-	-	-	-	-0.9	-0.8
Rotation(CW)	0.9	0.3	0.2	-	-	-	-0.67	0.23
Rotation(CCW)	0.8	0.4	0.23	-	-	-	-0.23	-0.78
In	-	0.2	0.67	-0.11	-	-	-	-
Out	-	-0.2	-0.56	0.12	-	-	-	-
Up	-	0.8	0.1	0.1	0.1	-0.23	-0.34	-0.12
Down	-	0.21	0.56	0.11	0.23	0.67	-0.23	-0.17
Left	-	0.45	0.56	0.11	-	-0.23	0.36	0.79
Right	-	-0.45	0.56	-0.11	-	-0.45	0.23	0.67

This simple system showed that it is possible to predict deterioration of psychomotor skill. We then further developed a system to predict cognitive skill deterioration. In order to do so, we developed some more variations of the ring transfer tasks.

3 Cognitive Skill Deterioration

For psychomotor skill evaluation, residents were tasked with grasping a series of a “virtual” rings and placing each on randomly highlighted pegs on a board. This task was modified to evaluate attention and memory. The attention task primed the user by highlighting a peg for 1 second with the user tasked to place the ring on the peg. The software does not permit placement of the ring on an incorrect peg. The memory task consisted of 3 sessions in which 4 pegs are highlighted in a sequence. The user is tasked with remembering the sequence and place the ring in that sequence on the pegs. An error is recorded every time the user did not correctly identify the peg for both tasks. Time taken to complete the task was measured in all the three tasks.

In addition to the cited above, we also included EEG based analysis. Research conducted over the past 40 years has established that electroencephalogram (EEG) reliably and accurately reflects subtle shifts in alertness, attention and workload that can be identified and quantified on a millisecond time-frame. We tested the hypothesis that EEG based analysis would predict effect of fatigue on surgical residents.

Advanced Brain Monitoring (ABM) Inc. implemented an integrated hardware and software solution for acquisition and real-time analysis of the EEG and demonstrated feasibility of operational monitoring of EEG indices of alertness, attention, task engagement and mental workload. The system includes an easily-applied wireless EEG system designed to be used in a mobile environment. A novel analytical approach was developed that employs linear and quadratic discriminant function analyses (DFA) to identify and quantify cognitive state changes using model-selected variables that may include combinations of the power in each of the 1-Hz bins from 1-40 Hz, ratios of power bins, event-related power and/or wavelet transform calculations. This unique modeling technique allows simultaneous selection of multiple EEG characteristics across brain regions and spectral frequencies of the EEG, providing a highly sensitive and specific method for monitoring neural signatures of cognition in both real-time and off-line analysis.

Participants wore the wireless EEG sensor headset configured to include the following bi-polar sensor sites: F3-F4, C3-C4, Cz-PO, F3-Cz, Fz-C3, Fz-PO while conducting the ring transfer task. The wireless sensor headset combines battery-powered hardware with a sensor placement system to provide a lightweight, easy-to-apply method to acquire and analyze six channels of high-quality EEG. The headset requires no scalp preparation and provides a comfortable and secure sensor-scalp interface for 12 to 24 hours of continuous use. The headset was designed with fixed sensor locations for three sizes (e.g., small, medium and large). Sensor placement was determined using a database of over 225 subjects so that each sensor is no more than one centimeter from the Residentational 10 – 20 system coordinates. The headset has been evaluated in over 800 healthy participants (fully-rested and sleep-deprived) and more than 200 patients with obstructive sleep apnea. Bi-polar recordings were selected in order to reduce the potential for movement artifacts that can be caused by linked mastoid or ear references during applications that require ambulatory recordings. The seven sensor site locations were selected to optimize the cognitive state classifications while ensuring the sensor headset could be easily applied in less than 10 minutes. Amplification, digitization and radio frequency (RF) transmission of the signals are accomplished with miniaturized electronics in a portable unit worn on the head. The combination of amplification and digitization of the EEG close to the sensors and wireless transmission of the data facilitates the acquisition of high quality signals even in high electromagnetic interference environments. Data are sampled at 256 samples/second with a bandpass from 0.5 Hz and 65Hz (at 3dB attenuation) obtained digitally with Sigma-Delta A/D converters. When utilized in the bi-directional mode, the firmware allows the host computer to initiate impedance monitoring of the sensors, select the transmission channel (so two or more headsets can be used in the same room), and monitor battery power of the headset.

Quantification of the EEG in real-time, referred to as the B-Alert™ system, is achieved using signal analysis techniques to identify and decontaminate fast and slow eye blinks, and identify and reject data points contaminated with excessive muscle activity, amplifier saturation, and/or excursions due to movement artifacts. Decontaminated EEG is then segmented into overlapping 256 data-point windows called overlays. An epoch consists of three consecutive overlays. Fast-Fourier transform is applied to each overlay of the decontaminated EEG signal multiplied by the Kaiser window ($\alpha = 6.0$) to compute the power spectral densities (PSD). The PSD values are adjusted to take into account zero values inserted for artifact contaminated data points.

Wavelet analyses are applied to detect excessive muscle activity (EMG) and to identify and decontaminate eye blinks. Once the artifacts are identified in the time-domain data, the EEG signal is decomposed using a wavelets transformation. Thresholds are developed for application to the wavelet power in the 64 – 128 Hz bin to identify epochs that should be rejected for EMG. The wavelets eye blink identification routine uses a two-step discriminant function analysis. The DFA classifies each data point as a control, eye blink or theta activity. Multiple data points that are classified as eye blinks are then linked and the eye blink detection region is established. Decontamination of eye blinks is accomplished by computing mean wavelet coefficients for the 0-2, 2-4 and 4-8 Hz bins from nearby non-contaminated regions and replacing the contaminated data points. The EEG signal is then reconstructed from the wavelets bins ranging from 0.5 to 64 Hz. Zero values are inserted into the reconstructed EEG signal at zero crossing before and after spikes, excursions and saturations. EEG absolute and relative power spectral density (PSD) variables for each 1-second epoch using a 50% overlapping window are then computed. The PSD values are scaled to accommodate the insertion of zero values as replacements for the artifact.

A single 30-minute baseline EEG test session is required for each participant to adjust the software to accommodate individual differences in the EEG. The output of the B-Alert software includes EEG metrics (values ranging from 0.1-1.0) for alertness/drowsiness, engagement, mental workload and distraction calculated for each 1-second epoch of EEG using quadratic and linear discriminant function analyses of model-selected EEG variables derived from power spectral analysis of the 1-Hz bins from 1-40Hz. These metrics have proven utility in tracking both phasic and tonic changes in cognitive states, in predicting errors that result from either fatigue or overload and in identifying the transition from novice to expert during skill acquisition. We employed these measures for analysis of fatigue on sleep deprivation.

Before and after taking in-hospital call, 32 obstetrical and surgical residents performed a predefined order of ring transfer tasks while wearing datagloves to measure hand movements. They also answered survey's assessing fatigue levels and, activities during call. Acceleration of the hand wrist was measured as smoothness of hand movement. This was calculated through Cyberglove system. Errors in attentional and memory task and time lags in each trial were evaluated. Data was collected over a period of 1 month with 4 pre call and post call trials for each resident. Average number of errors per trial per person and time lag in the pre-call and post call condition was used to perform t-test. A statistically significant decrement ($p<0.01$) in the smoothness of hand movements was observed post-call (Fig 1(b))and in errors in attentional and memory task (Fig (1(c) and 1(d)). Average time lags in post call conditions are also shown in Fig (1b,c,d).

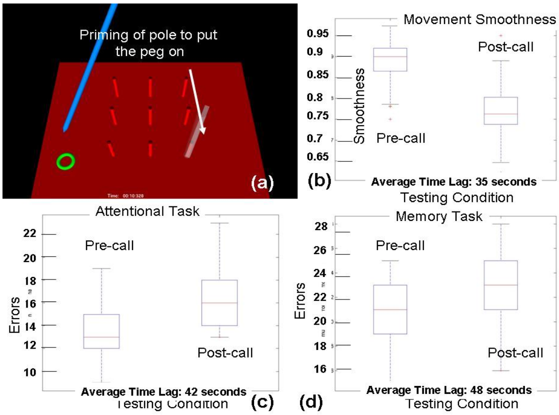

Fig. 1.

In addition, the EEG signals based ratings showed a statistically significant difference between post call and pre call ratings. All the measured metrics were averaged across a task. It was determined that averaged alertness/drowsiness, engagement, mental workload and distraction ratings across tasks were significantly different across call. Further the ratings showed high correlation with subjective fatigue ratings (0.91).

Call-associated fatigue is associated with increased error rates in psychomotor and cognitive skills. Cognitive skill as measured through the task and EEG analysis showed a significant decline. These results need to be accounted for in the design of resident curriculum.

References

1. Behrenz, K., Monga, M.: Fatigue in Pregnancy: A comparative study. American Journal of Perinatology 16(4), 185–188 (1999)
2. Rosen, J., Brown, J.D., Chang, L., Sinanan, M., Hannaford, B.: Generalized Approach for Modeling Minimally Invasive Surgery as a Stochastic Process Using a Discrete Markov Model. IEEE Transactions on Biomedical Engineering 53, 399–413 (2006)
3. Tendulkar, A., Victorino, G., Chong, T., Bullard, M., Liu, T., Harken, A.: Quantification of Surgical Resident Stress "On Call". Journal of the American College of Surgeons 201(4), 560 (2005)