Available online at **www.ijmrhs.com**

ISSN No: 2319-5886

International Journal of Medical Research & Health Sciences, 2017, 6(8): 98-104

Test-Retest Reliability of Assessment of Anaesthetist's Performance in Managing Serious Adverse Events on a Screen-Based Computer Simulator

Jyotirmay Kirtania, Shreyasi Ray*

Department of Anaesthesiology, ESI-PGIMSR and ESIC Medical College, Kolkata, India

**Corresponding e-mail: rayshreyasi@rediffmail.com*

ABSTRACT

Serious adverse events (SAE) during anaesthesia procedure are uncommon but potentially fatal. Traditional methods to assess the performance of anaesthetists in SAE management are largely subjective and unreliable. In this study, we assessed the test-retest reliability of the assessment of performance of anaesthetists in managing SAE during anaesthesia procedure by determining if different observations assessing the same clinical scenario on a computer screen-based anaesthesia simulator would have agreement on the rating of the anaesthetist's performance. ***Methods:*** *Three SAE scenarios on a computer screen based simulator were used for assessment. From a pilot study, the minimum number of trials beyond which there is little or no improvement of performance was determined for each SAE scenario. Six fresher anaesthesia postgraduates were enrolled. Each participant undertook ten trials for each SAE scenario and submitted the trial-logs in pdf format to the investigator. SAE detection latency, initial treatment latency period, SAE treatment completion period was collected from the simulator trial-logs. Test-retest reliability (internal consistency) between trials was assessed with Pearson's Product Moment Correlation Coefficient. Scatter plot of improvement of latency period and treatment time versus the number of simulator trials were used to characterize the learning curves. Regression model fitting of the learning curve was used to find the best fit model for the learning curve.* ***Results:*** *There was significant test-retest reliability and internal consistency between the trials for each of the three SAE scenarios.* ***Conclusion:*** *PC based simulated SAE scenarios are effective and reliable for the assessment of performance of fresher anaesthesia postgraduates.*

Keywords: Serious adverse events, Anaesthesia, PC Based Simulation, Inter-rater reliability, Performance assessment

INTRODUCTION

Serious adverse events (SAE) during anaesthesia procedure are uncommon. But SAE during anaesthesia can be rapidly fatal or can cause permanent disability to the patient unless promptly recognized and corrected by the anaesthesia caregiver [1].

Traditional anaesthesia training for SAE management largely involves the cognitive domain, with little or no practice of the higher cognitive domain, psycho-motor domain, and affective domain. This is due to the fact that it is ethically not permissible to expose patients to SAE under anaesthesia for the purpose of training, as also the uncommon occurrence of SAE during anaesthesia. Since SAE under anaesthesia are uncommon, anaesthetists usually "learn" SAE management throughout their career, sometimes risking patients' lives and consequent medico-legal actions. Thus, the traditional process of learning to manage SAE during anaesthesia procedure is no longer tenable in the modern-day anaesthesia practice.

The theory of the learning curve is based on the simple idea that the competence & time required to perform a task decreases as a worker gains experience. High risk industries like aviation and space technology, constantly train and assess their personnel on simulators to improve and ensure the safety of their customers against the occurrence of uncommon but serious adverse incidents involving loss of life or property. Although simulator-based-learning is not new in anaesthesia, the learning curve of fresher anaesthesia-postgraduates in managing serious adverse events during anaesthesia procedure, on a computer simulator was not studied till date.

A learning curve is a collection of data points (x_j, y_j) which describe how the performance (y_j) is related to training sample sizes (x_j), where j=1 to m, m being the total number of instances. These learning curves can typically be divided into three sections: In the first section, the performance increases rapidly with an increase in the size of the training set; the second section is characterized by a turning point where the increase in performance is less rapid and a final section where the classifier has reached its efficiency threshold, i.e., no (or only marginal) improvement in performance is observed with increasing training set size [2].

In this study, we assessed the inter-rater reliability of the assessment of performance of anaesthetists in managing SAE during anaesthesia procedure by determining if different observers assessing the same clinical scenario in a computer screen-based anaesthesia simulator would agree on their rating of the anaesthetist's performance. This study also aims to characterize the learning curve of fresher anaesthesia-postgraduates in managing serious adverse events during anaesthesia procedure on a simulator.

MATERIALS AND METHODS

The study was conducted between June and July 2016, after seeking approval from the Institute Ethical Committee. The study design was prospective observational-non-randomized allocation, same participant act as control and test, with repetitive measures.

A survey among four anaesthesia faculty members was carried out to prepare a validated list of the minimum number (N_s) of SAE scenarios for the simulator training. It was found that N_s=3 (three). These three SAE scenarios were: (i) Anaphylaxis, (ii) ST-T changes, and (iii) Obstructed airway. Using this validated list of the minimum number (N_s=3) of SAE scenarios, a pilot study was conducted on two senior anaesthetists (>6 years post MD experience), to determine the number of trials beyond which there is little or no improvement of performance (time to detect SAE, initiate treatment, and complete treatment for each scenario). Using this pilot study data, the minimum number of trials (N_t) to be tested for each SAE scenario, was determined to be, N_t=10. Thus, the sample size (number of simulator trials or, simulator trial logs) to detect differences between performances on repeated test trials was at least = ($N_s \times N_t$, or 3×10) = 30, for each participant. Six (6) fresher anaesthesia-postgraduates were enrolled in the study. Thus, the sample size (the number of simulator trials) allocated to each of the six participants, m = ($N_s \times N_t$) = 30, and the total sample size (total number of simulator trials) was 6×30=180.

Stratified Random Sampling was done for the participants

Only fresher postgraduate senior residents in Anaesthesiology were approached for participation and willing participants were enrolled. The participants were contacted from the data obtained in the public domain database of Institutes which run three-year MD Anaesthesiology Curriculum approved by the Medical Council of India.

Consecutive Sampling was done for the simulator data

All the completed simulator trial logs (satisfying the inclusion criteria) were included in the study till the required sample size (training set size) is achieved.

Inclusion criteria for simulator logs

Completed simulator logs for the validated SAE scenarios.

Inclusion criteria for participants

Fresher postgraduate senior residents in Anaesthesiology with less than 1-year post MD experience.

Senior residents in Anaesthesiology, having passed MD in the first attempt.

Participants with basic knowledge of Windows based PC systems.

Exclusion criteria for simulator logs

Incomplete simulator logs.

Simulator logs from non-validated SAE scenarios.

Exclusion criteria for participants

Postgraduate senior residents with more than 1-year post MD experience.

Non-MD senior residents (e.g. with DA, DNB).

Senior residents having passed MD on more than one attempt.

Senior residents with additional training in anaesthesiology (e.g. DA, DNB, Junior Residency) prior to MD.

Participants without basic knowledge of Windows based PC systems.

Fresher postgraduate in Anaesthesiology (senior residents with less than 1 year of post MD experience) were approached by email invitation for participation in the study. Among the nine contacted candidates fulfilling the inclusion criteria, six willing participants were enrolled in the study after obtaining their expressed consent for participation. The Anesoft anaesthesia Simulator 6™ (Windows™ PC based simulator) was used for the study.

In the learner sensitization phase of the study, the study participants were instructed to use the simulator by the principal investigator (PI), and also by self-directed learning by participants using the "Consultant" module of the simulator. During each practice trial, one SAE scenario (from the validated list) was manually selected by the participant. Each participant undertook 'm=10' number of training sessions for each of the three SAE scenarios, over a period of six weeks. At the end of each simulator trial, the participant saved the pdf log of the session with a unique file-name and emailed it to the PI. The participant also viewed the simulator log to get appropriate feedback regarding his/her performance in that simulator session.

The following data was collected for each participant:

• Demographic data: (Data tool: Data from participant-enrolment-form)

• Age

• Sex

• Date of passing MD Anaesthesiology

• Prior training on anaesthesia simulator (Yes/No).

• Basic knowledge of windows operating system and email (Yes/No).

Simulator Data (Data tool: Time & event log of each simulator session generated by the simulator and emailed by the participant to the principal investigator).

The following primary data from the simulator were collected for each practice session from assessment of the simulator session log:

a) Practice session start time.

b) SAE start time.

c) SAE detection time.

d) Initial treatment start time.

e) Treatment end time.

f) All the recommended treatment steps completed or not.

The following derived data was calculated from the simulator log for each simulator session for analysis:

a) The response time of the study participant to detect SAE ("SAE detection latency"). It was defined as the interval between the occurrence of the SAE and it detection by the study participant during each practice session.

b) The response time of the study participant for initial action to treat the SAE ("initial treatment latency"). It was defined as the interval between the occurrence of the SAE and its initial corrective action taken by the study participant during each practice session.

c) Whether the participant correctly performed all the steps of the standard treatment for the SAE (yes/no).

d) The time taken by the study participant to perform all the corrective measures for the SAE (if correctly completed at all). "SAE treatment completion time" was defined as the interval between the start of the initial step of treatment and the completion of the last step of treatment performed by the participant.

e) Learning curve of participants (scatter plot) for each scenario was constructed based on improvement over successive simulator trials, using the "SAE detection latency", "initial treatment latency" and "SAE treatment completion time".

Statistical methods

Test-retest reliability (internal consistency) between trials was assessed with Pearson's product moment correlation coefficient alpha. $P<0.05$ was considered statistically significant. Scatter plot of number of practice sessions versus participant performance (improvement of latency period and treatment time) was used to characterize the learning curve for each SAE scenario. Regression analysis was used for model fitting (to characterize the best fitting distribution model) of the scatter plots of the learning curves.

Statistical Analysis was done exclusively with "LibreOffice Calc" Version 5.1.4.2 (2016), and "R" version 3.3.1 (2016), with package "Rcmdr" Version 2.2.5 (2016), all of which are free and open-source software.

RESULTS

The descriptive statistics of demographic profile and professional experience of the experts (pilot study participants) and the fresher postgraduates (study participants) are described in Table 1. All the experts and fresher postgraduate participants completed the stipulated number of simulator trials for each SAE scenario. Every expert and participant completed all the recommended steps for treatment of the SAEs in each of the simulator trials.

Table 1 Demographic characteristics of pilot study participants (experts) and study participants (fresher postgraduates)

Variables	Experts	Fresher postgraduates
Number of participants	2	6
Male: Female	01:01	03:03
Median age in years	53.5	30
(Min., Max.)	(52, 55)	(29, 32)
Median clinical experience (post MD) in years	23.5	0.67
(Min., Max)	(22, 25)	(0.67, 0.92)

Test-retest reliability (internal consistency) of performance data (SAE detection latency, Initial treatment latency, and SAE treatment completion time) between trials was assessed with Pearson's product moment correlation coefficient alpha. Closer the alpha value towards unity (1.0), greater the test-retest reliability of the performance data. Test-retest reliability measures of performance of experts in the pilot study were: Alpha reliability=0.7933; Standardized alpha=0.6752. Test-retest reliability measures of performance of fresher postgraduates in the study were: Alpha reliability=0.7957; Standardized alpha=0.6681. So, these simulator trials were reliable assessment tools to assess the learning (improvement of performance) of fresher postgraduates while they manage simulated SAE during anaesthesia procedures.

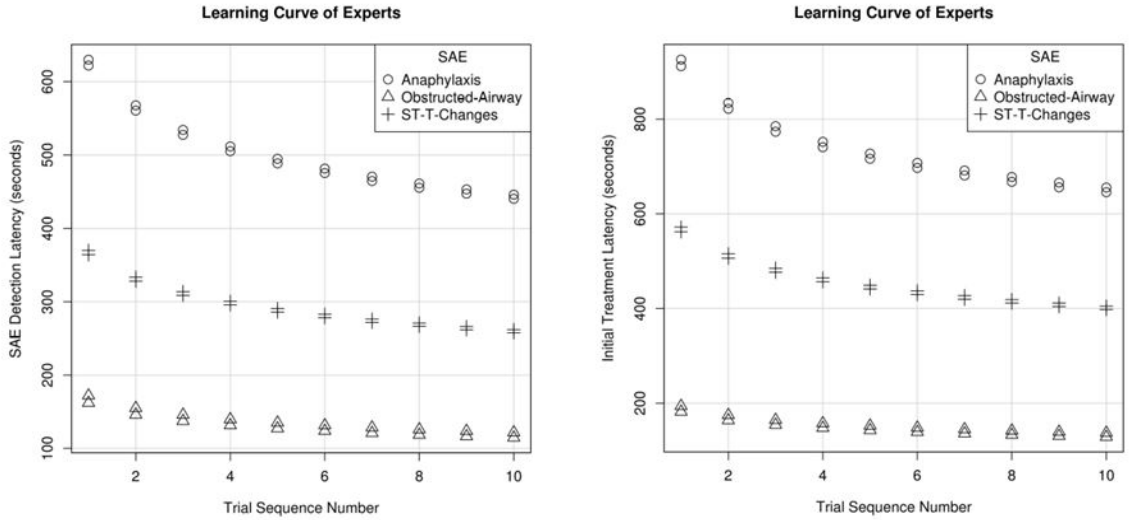

Figure 1 Scatter plot showing learning curve of experts

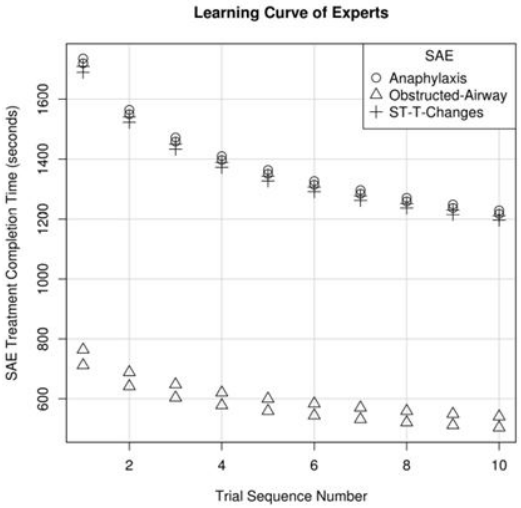

Figure 1(a) Scatter plot showing learning curve of experts

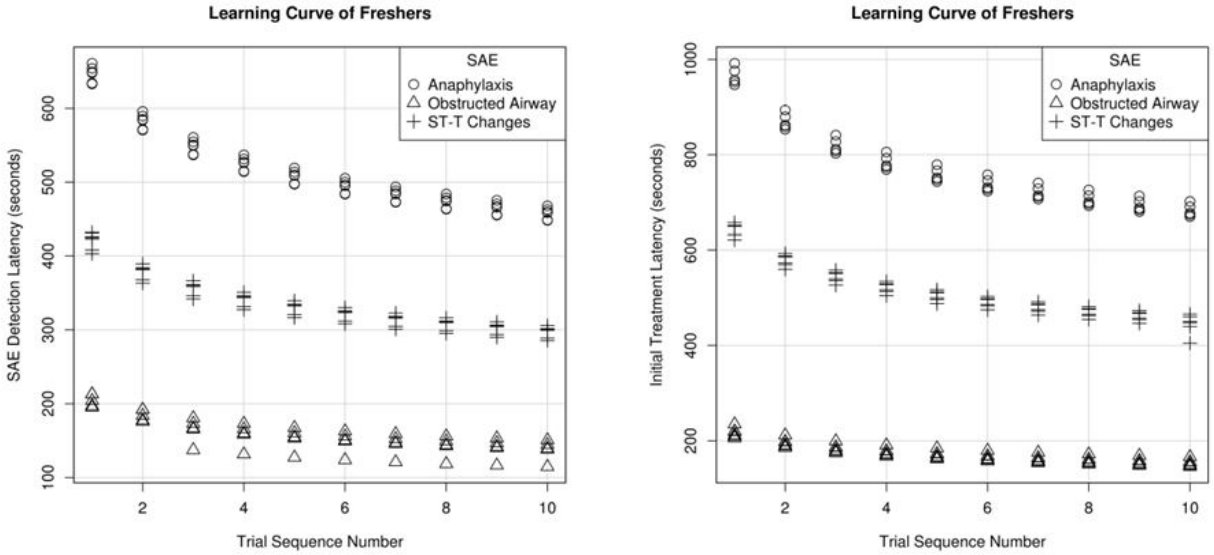

Figure 2 Scatter plot showing learning curve of fresher postgraduates

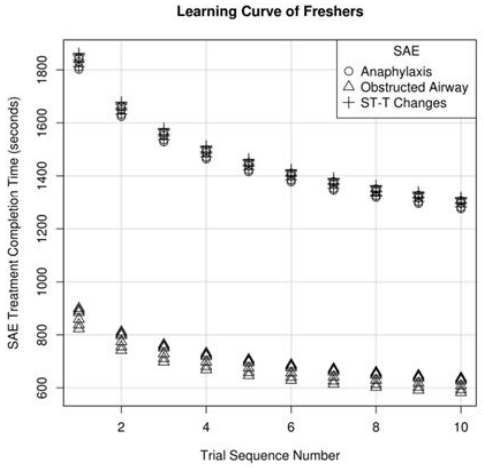

Figure 2(a) Scatter plot showing learning curve of fresher postgraduates

Scatter plots of various performance measurements (time) of experts, versus trial numbers for each SAE scenarios are depicted in Figures 1 and 1(a). It is seen that the learning curve (scatter plot) of experts for each SAE scenario reached a plateau within ten successive trials on the simulator. Scatter plots of various performance measurement (time) of fresher postgraduates, versus trial numbers for each SAE scenarios are depicted in Figures 2 and 2(a). It is seen that the learning curve (scatter plot) of fresher postgraduates for each SAE scenario also reached a plateau within ten successive trials on the simulator.

Table 2 Overall learning curve (performance improvement) characteristics of experts during the pilot study for the three SAE scenarios

Type of scatter plot	Regression Model	Standard error of the regression (S)*	Slope (α)	Intercept (B)
SAE detection latency versus trial number	Linear	154.7	-11	371
	Logarithmic	154.2	-48.5	383.7
	Power	0.551*	-0.149	337.4
Initial treatment latency versus trial number	Linear	245.5	-15.9	535.2
	Logarithmic	244.9	-70	553.7
	Power	0.675*	-0.149	461.2
SAE treatment completion time versus trial number	Linear	380.4	-39.5	1331.3
	Logarithmic	377.9	-174.3	1377.3
	Power	0.405*	-0.15	1294.3

*Lesser the value of the standard error of the regression (S) of the data, better the data fits with that regression model. In this case the best fit model is the Power model, $T = B * n^{-\alpha}$

Table 3 Overall learning curve (performance improvement) characteristics of fresher postgraduates during the study for the three SAE scenarios

Type of scatter plot	Regression Model	Standard error of the regression (S)*	Slope (α)	Intercept (B)
SAE detection latency versus trial number	Linear	149.8	-12.2	405.2
	Logarithmic	149.2	-54.1	419.6
	Power	0.499*	-0.157	380.4
Initial treatment latency versus trial number	Linear	248.5	-17.5	565.2
	Logarithmic	247.8	-76.8	602.6
	Power	0.633*	-0.151	512.5
SAE treatment completion time versus trial number	Linear	374.2	-43.1	1450.2
	Logarithmic	371.2	-190.1	1500.4
	Power	0.357*	-0.15	1429.4

*Lesser the value of the standard error of the regression (S) of the data, better the data fits with that regression model. In this case the best fit model is the Power model, $T = B * n^{-\alpha}$

Regression analysis (Tables 2 and 3) shows that all the learning curves (scatter plots of performance measurement versus trial number) approximately fit the "power-law distribution model", rather than the linear or the logarithmic models.

DISCUSSION

The slope (steepness) of the learning curve for experts and fresher postgraduate are similar (Figures 1 and 2). The intercepts (initial performance measures) are different (Tables 2 and 3) for experts and fresher postgraduates. However, this study was not designed to detect quantitative differences in the performance measure of experts and fresher postgraduates.

In regression analysis, lesser the value of the standard error of the regression (S) of the data in a regression model,

better the data fits with that distribution model. The coefficient of determination (R^2) is not a valid indicator of model fitting, particularly when the data follow nonlinear model, as in our study [3]. So R^2 was not used for the model fitting. In our study, the learning curve approximately fits with the power-law equation, $T = B * n^{-\alpha}$, where 'T' is the performance measurement (time), and 'n' is the number of trials. The constant 'B' represent the performance measurement at the beginning of the trial (i.e. n=1). The exponent (power) 'α' represent the 'slope' or 'steepness' of the learning curve (scatter plot). Unlike an exponential curve, whose linear slope is a constant multiple of 'T', the slope of this "power-law" distribution curve varies with the number (n) of the trials. In this case, at higher values of 'n', the curve flattens out (plateau of the learning curve).

Traditional methods (written examination, oral case presentation, and direct observation of procedural skills) to assess the performance of anaesthetists in SAE management are largely subjective and non-reproducible. The ability to measure actual performance, vigilance, interpretation of data in real time, and formulation and implementation of a management plan, is not readily demonstrable by these traditional methods. Our study is novel as it has it has proven that a computer screen based simulator can assess the performance of anaesthetists with a valid test-retest reliability.

The long-term impact of this simulator training on patient safety during anaesthesia care cannot be assessed by performance of anaesthetists on simulators. Hays, et al. conducted a meta-analysis of flight simulation research to identify important characteristics associated with the effectiveness of simulator training of military aircraft pilots. The major finding was that the use of simulators combined with aircraft training consistently produced improvements in training compared to aircraft training only [4]. However, in the field of anaesthesia, no study has been undertaken which investigated the efficacy of simulator training to improve the management of real life SAE scenarios. Again, ethical issues and the rarity of SAE under anaesthesia may prevent such studies to be undertaken. If widespread simulator training and assessment of SAE management are introduced for anaesthetists, the impact can be judged by studying the trend (over years or decades) in closed claim registries and medico-legal suits in anaesthesia practice.

CONCLUSION

The performance data (learning curve) of fresher anaesthesia postgraduates managing computer simulated SAE scenarios approximately follow a "power-law distribution" curve. For each SAE scenario, a performance plateau was observed within ten practice sessions. This simulator trial method has valid test-retest reliability for each of the three SAE scenarios. Hence this method is valid and reliable for performance assessment of anaesthetists managing simulated SAE under anaesthesia on a PC based simulator. This method of performance assessment should be incorporated into postgraduate medical curriculum to replace traditional methods of assessment like written examination, oral case presentation, and direct observation of procedural skills. For a given SAE scenario during anaesthesia procedure, ten simulation trials are adequate and reliable for the learning and assessment of learning. Long term retention of simulator acquired skills and efficacy of simulator training in improving patient safety need to be assessed by further studies.

REFERENCES

[1] Jenkins, Kath, and A.B. Baker. "Consent and anaesthetic risk." *Anaesthesia* 58.10 (2003): 962-984.

[2] Hopper, A.N., M.H. Jamison, and W.G. Lewis. "Learning curves in surgical practice." *Postgraduate Medical Journal* 83.986 (2007): 777-779.

[3] Spiess, Andrej-Nikolai, and Natalie Neumeyer. "An evaluation of R^2 as an inadequate measure for nonlinear models in pharmacological and biochemical research: A Monte Carlo approach." *BMC Pharmacology* 10.1 (2010): 6.

[4] Hays, Robert T., et al. "Flight simulator training effectiveness: A meta-analysis." *Pharmacology* 4.2(1992): 63-74.