Efficacy of high-fidelity simulation debriefing on the performance of practicing anaesthetists in simulated scenarios

P. J. Morgan¹*, J. Tarshis², V. LeBlanc³, D. Cleave-Hogg⁴, S. DeSousa⁵, M. F. Haley¹, J. Herold-McIlroy³ and J. A. Law⁴

¹Department of Anesthesia, Women’s College Hospital, 76 Grenville St., Toronto, ON, Canada M5S 1B2.
²Department of Anesthesia, Sunnybrook Health Sciences Centre, Toronto, ON, Canada.
³Wilson Centre for Research in Education, University of Toronto, ON, Canada.
⁴Department of Anesthesia, Queen Elizabeth II Health Sciences Centre, Dalhousie University, Halifax, Canada

*Corresponding author. E-mail: pam.morgan@utoronto.ca

Background. Research into adverse events in hospitalized patients suggests that a significant number are preventable. The purpose of this randomized, controlled study was to determine if simulation-based debriefing improved performance of practicing anaesthetists managing high-fidelity simulation scenarios.

Methods. The anaesthetists were randomly allocated to Group A: simulation debriefing; Group B: home study; and Group C: no intervention and secondary randomization to one of two scenarios. Six to nine months later, subjects returned to manage the alternate scenario. Facilitators blinded to study group allocation completed the performance checklists (dichotomously scored checklist, DSC) and Global Rating Scale of Performance (GRS). Two non-expert raters were trained, and assessed all videotaped performances.

Results. Interim analysis indicated no difference between Groups B and C which were merged into one group. Seventy-four subjects were recruited, with 58 complete data sets available. There was no significant effect of group on pre-test scores. A significant improvement was seen between pre- and post-tests on the DSC in debriefed subjects (pre-test 66.8%, post-test 70.3%; \( F_{1,57} = 4.18, P = 0.046 \)). Both groups showed significant improvement in the GRS over time (\( F_{1,57} = 5.94, P = 0.018 \)), but no significant difference between the groups.

Conclusions. We found a modest improvement in performance on a DSC in the debriefed group and overall improvement in both control and debriefed groups using a GRS. Whether this improvement translates into clinical practice has yet to be determined.

Br J Anaesth 2009; 103: 531–7

Keywords: education; safety, patient safety; simulation

Accepted for publication: May 22, 2009

Human error has been identified as a significant cause of adverse events in medicine and other high-reliability organizations.¹⁻⁵ Despite increasing concern about the quality of medical care as outlined in the Institute of Medicine reports in 1999 and 2001 in the USA and the Baker-Norton report in Canada in 2004,²⁻⁴ medicine lags significantly behind other high-hazard industries in the implementation of safety systems designed to reduce error. National organizations such as the Agency for Healthcare Research and Quality, the Anesthesia Patient Safety Foundation, the National Patient Safety Foundation, and the Canadian Patient Safety Institute have launched campaigns attempting to address some of these issues.

The specialty of anaesthesia has been at the forefront of research in error recognition and reduction. In 1984, it was estimated that human error is responsible for at least 70% of adverse anaesthesia outcomes.⁵⁻⁶ A subsequent 18 month prospective analysis showed that 82% of witnessed unsafe practices were considered preventable, and 27% could have been fatal if not recognized and corrected.⁷ Similarly, 29% of errors in an intensive care unit were graded as severe or potentially detrimental to patients as...
noted by observers with human engineering experience. A similar process called Crisis Resource Management (CRM), these initiatives have been deemed successful, although strong scientific-based evidence is lacking. A similar process called CRM, these initiatives have been deemed successful, although strong scientific-based evidence is lacking. A similar process called Crisis Resource Management (CRM), these initiatives have been deemed successful, although strong scientific-based evidence is lacking.

The aviation industry has long used high-fidelity simulation training for flight crews. Called Cockpit (or Crew) Resource Management Training (CRM), these initiatives have been deemed successful, although strong scientific-based evidence is lacking. A similar process called CRM has been widely adopted in anaesthesiology utilizing human patient simulators; however, there has been no research examining the role of debriefing alone on performance improvement as opposed to the overall effect of simulation-based training including debriefing and self-reflection.

Although there is lack of evidence in the literature with respect to the role of debriefing or ‘feedback’ in affecting performance, Issenberg and colleagues, in a systematic review of the literature to determine the features and uses of high-fidelity medical simulation that lead to the most effective learning, identified the education feedback to be the most important item in the process of learning.

The primary purpose of this study was to determine whether high-fidelity simulation debriefing and feedback involving reflection and discussion led by an experienced facilitator improved performance of practicing anaesthetists in managing simulated clinical scenarios.

Methods

This prospective, randomized, controlled study was conducted at the Canadian Simulation Centre at Sunnybrook Health Sciences Centre, University of Toronto, Toronto, Canada, and the Atlantic Health Training and Simulation Centre, Queen Elizabeth II Health Sciences Centre, Dalhousie University, Halifax, Canada.

After Institutional Ethics Board approval, a brochure outlining the nature of the study was sent to all anaesthetists in Ontario and four eastern Canadian provinces. Also, brochures were sent to anaesthesia departments and the study was advertised in the national anaesthesia journal. Participants were eligible if they were certified anaesthetists with FRCPC (Fellow of the Royal College of Physicians and Surgeons of Canada) or equivalent status and would be in a stable practice for the study time period. Participants were excluded if they had Advanced Cardiac Life Support (ACLS) training or Anaesthesia Crisis Resource Management (ACRM) training within 6 months of participation or were ACRM instructors. Participants agreed not to take part in an ACLS/ACRM training program during the study period. Participants received continuing education credits and honoraria.

Two clinical scenarios developed in a previous study were used and each included two critical events, one of which required the use of an ACLS protocol. Situations were introduced into the scenario enhancing the likelihood of human error. All sessions were recorded and labelled using unique identifiers which excluded participant identification, session date, and pre- or post-test labels.

The participants were enrolled on a first-come, first-served basis, and were randomly assigned using computer-generated random number lists to one of three groups; Group A: high-fidelity simulation debriefing and discussion led by an experienced facilitator; Group B: Home Study program; or Group C: control, no educational intervention. After initial randomization, participants were further randomized to receive either Scenario 1 or Scenario 2 as their first (pre-test) scenario.

Two performance assessment tools were used to evaluate the anaesthetists’ management of the case and included a scenario-specific assessment tool developed in a previous study which consisted of a 104-item (Scenario 1) and a 99-item (Scenario 2) dichotomously scored checklist (DSC) (Supplementary Appendix A). A Global Rating Scale of Performance (GRS) allowed for assessors’ overall rating of performance (Supplementary Appendix B).

Scripted actors filled the roles of circulating nurse, scrub nurse, surgeon, and respiratory therapist. An additional anaesthetist was available, if help was sought during the scenario. All participating faculty facilitators and actors were provided information pertaining to the purpose and conduct of the study. Actors would assist the anaesthetist but did not provide information nor give assistance without a specific request. Faculty facilitators responded to specific questions about the patient as requested by the participant. Once the participant had no further questions, the faculty facilitator left the room. If a second anaesthetist was summoned, he/she would assist the participant as specifically requested.

All faculty facilitating the simulation sessions were academic anaesthetists familiar with simulation and ACRM. Facilitators completed all performance checklists during the course of the simulation scenario while watching the performance of the subject through a one-way window and were blinded as to whether the participant was performing the pre- or post-test scenario and to the participant’s group assignment. If the facilitator had problems viewing the participant performing a certain item, they were able to review the videotape after the session to ensure completion of the checklists.

Educational intervention and debriefing strategy

Group A received an individual debriefing session after the completion of the pre-test scenario with only the subject and facilitator present. This debriefing consisted of two parts. Part 1 was a standardized Powerpoint presentation discussing the nature and categories of medical
errors including: (a) knowledge-based errors; (b) rule-based errors; (c) skill-based errors; (d) technical errors; and (e) latent errors. The slide presentation included a discussion of what errors had been built into the scenario the subject had just completed. The second part consisted of reviewing sections of the videotape by the subject and the facilitator. The facilitator would allow the subject to watch the videotape, but focused on certain time periods where technical or non-technical skills issues were most evident. The subject was then asked to comment on the videotape segment and the facilitator allowed the subject to fully discuss the issues. The facilitator would then attempt to associate the subjects’ comments with the categories of errors described in the previous slide presentation, explore the subjects’ opinions, and provide feedback on methods of mitigating error. Anaesthesia Crisis Resource Management, designed for the debriefing of anaesthetists following the management of simulated scenarios, has been described in more detail by both Howard and colleagues and Gaba and colleagues.

Group B received a Home Study Program consisting of peer-reviewed articles outlining the causes of human error in medicine. Group C received no educational intervention and was dismissed after completion of their pre-test scenario. After the post-test session, participants in Groups B and C were offered debriefing, as given to the intervention group, following their pre-test scenario.

Upon arrival to the simulation centre, written, informed consent was obtained and participants signed confidentiality forms requesting that they not discuss the session events. In addition, they signed consent forms allowing session videotaping. Participants then changed into operating theatre greens and were given as much time as needed to familiarize themselves with the simulated operating theatre, anaesthetic gas machine, and drug cart. Participants were supplied with the history, physical examination, and laboratory findings of the simulated patient and the nature of the surgical procedure. The facilitator provided further information upon request. The facilitator or simulator operator provided the voice of the patient and interacted with the participant in real time. Participants were informed that the facilitator would be giving an anaesthetic in the next operating theatre.

Recording of the simulation scenario began from the moment the participant entered the operating theatre. A MedSim high-fidelity patient simulator (MedSim Advanced Medical Simulations Ltd, Fort Lauderdale, FL, USA) was used at the Toronto site and a METI (Medical Education Technologies, Inc., Sarasota, FL, USA) high-fidelity simulator at the Halifax site. One of three experienced simulator operators conducted all sessions. Six to nine months after completion of the pre-test scenario, participants returned and managed the scenario not encountered during the initial session.

Facilitators rated the live performances of all participants using the previously described assessment tools and they were not informed as to whether they were facilitating a first or second session. However, to ensure non-bias of facilitators, two non-expert raters (medical student and international medical graduate with a foreign anaesthesia fellowship) were trained to use the tools and once their ratings of videotaped performances achieved >90% correlation with two expert raters, they rated all participants’ performances using the videotapes beginning at the induction of anaesthesia to the completion of the scenario. Non-expert raters, who rated all performances on videotapes, were blinded as to participant group assignment and as to whether they were viewing the participants’ pre- or post-tests.

**Statistics**

Statistical analyses were performed using SPSS Version 13.0 (SPSS Inc., Chicago, IL, USA). Interim data analysis (mixed design analysis of variance [ANOVA], with time (pre, post) and group (B and C) as predictor variables) revealed no difference between Groups B and C (home study group and control group), so these groups were merged into a single group (control) for comparison with the intervention Group A in order to increase the power of our analyses between the individuals who received the debriefing and those who did not.

To ensure that the scenarios were equivalent at baseline, we conducted independent sample t-tests on the scores of the DSC and the GRS for Scenarios 1 and 2 at baseline. Averaged scores on the DSC and GRS were submitted to separate mixed design analyses of variance with time (pre, post-test) as the repeated measures and the group (control, debrief) as the between-subject variable. In the presence of any significant interactions, main effects were tested using t-tests.

**Results**

Seventy-four participants were recruited, with 58 complete data sets available (Fig. 1). There were no statistically significant differences in participant characteristics (Table 1).

The scores on the assessment scales did not differ between Scenarios 1 and 2 at baseline (DSC: Scenario 1=67.9, Scenario 2=67.3, P=0.82; GRS: Scenario 1=67.8, Scenario 2=71.5, P=0.40). In addition, the pre-test scores did not differ between the two study sites (DSC: Site 1=67.2, Site 2=68.6, P=0.58; GRS: Site 1=69.0, Site 2=71.1, P=0.68) nor was there a difference in the mean number of days between the pre- and the post-test sessions between the groups (Table 2).

The inter-rater reliability coefficients for the scales were acceptable to high (r=0.71 GRS, r=0.91 DSC). As such, the scores for the three assessors (facilitator and two non-expert raters) were averaged together, and used in subsequent analyses.
Assessed for eligibility, $n=74$

Excluded due to ineligibility, $n=3$ [due to family practice anaesthetist ($n=1$); due to recent ACLS ($n=1$); due to not practicing in OR ($n=1$)]

Group A*
- Allocated to intervention, $n=29$
- Did not receive allocated intervention, $n=2$ [due to simulator malfunction ($n=1$); due to session cancellation by participant and inability to reschedule another session ($n=1$)]

Baseline Scenario 1
$n=14$

- Lost to follow-up, $n=1$ (due to refusal to return)
- Discontinued intervention, $n=1$ (due to more than 9 months between baseline and post-test)

Group B§
- Allocated to intervention, $n=21$
- Did not receive allocated intervention, $n=1$ [due to inability to schedule faculty facilitator for participant’s scheduled session ($n=1$)]

Baseline Scenario 1
$n=10$

- Lost to follow-up, $n=0$
- Discontinued intervention, $n=0$

Group C¥
- Allocated to intervention, $n=21$
- Did not receive allocated intervention, $n=2$ [due to refusal upon arrival at scheduled session ($n=1$); due to session cancellation by participant and inability to reschedule another session ($n=1$)]

Baseline Scenario 1
$n=13$

- Lost to follow-up, $n=4$ (due to refusal to return)
- Discontinued intervention, $n=1$ (due to more than 9 months between baseline and post-test)

Group A*
- Analysed, $n=24$
- Excluded from analysis, $n=3$ [due to simulator malfunction ($n=1$); due to missing post-test data ($n=2$)]

Group B/C£
- Analysed, $n=34$
- Excluded from analysis, $n=5$ [due to missing post-test data ($n=5$)]

Fig 1 Study design and conduct. *Received debriefing after baseline simulation session. §Received literature after baseline simulation session. ¥Received no intervention after baseline simulation session. £Groups B and C merged: received no debriefing.
The mean scores for the assessments are given in Table 2. The difference in DSC and GRS scores between the groups is presented in Table 3. The ANOVA revealed a significant improvement from pre- to post-test session [main effect of time: $F_{(1,57)}=5.94$, $P=0.018$; see Table 2]. However, participants in the debriefing group did not perform better overall than those in the control group ($P=0.52$), but a significant improvement in scores for the debriefed group ($P=0.03$). All participants improved on the GRS from pre- to post-test session [main effect of time: $F_{(1,57)}=5.94$, $P=0.018$; see Table 2]. However, participants in the debriefing group did not perform better overall than those in the control group ($P=0.57$). The two groups improved equally from pre- to post-test, as indicated by a non-significant time by group interaction ($P=0.61$).

### Discussion

The Institute of Medicine report has suggested that simulation be used to improve patient safety and in 2006, the Agency for Healthcare Research and Quality (AHRQ) awarded more than $5 million to support research under its ‘Improving Patient Safety Through Simulation Research portfolio’ (http://www.ahrq.gov/qual/simulproj.htm).

The strength of this study is that it is the first randomized, controlled, blinded study in which participants were tested on two different but statistically comparable, standardized scenarios.

With respect to the primary outcome of the study, we demonstrated a significant improvement in the scores on the DSC in the debriefing group when compared with the control. This 3.5% improvement is a modest improvement and whether this improvement translates to a significant reduction in clinical adverse events is unknown. Extrapolating from the Australian Incident Monitoring study in which human error accounted for 1660 of 2000 anaesthesia incidents, this degree of performance improvement would represent a decrease of 58 incidents. The results do suggest, however, that exposure to simulation-based practice appears to have a beneficial effect, with debriefing tending to shift the performance in a more positive direction. Whether this ‘single shot’ training intervention is adequate to improve performance in performance is questionable. A confounding factor is the fact that subjects may become more familiar with the environment, hence the improvement in scores.

A second important observation from this study was that the effect of the educational intervention was still evident 6–9 months later, a fact that is important in the decision as to how frequently this educational technology should be reinforced. Previous studies have shown that debriefing interventions can lead to improvements in performance when the post-intervention test is administered on the same day as the intervention, 1 month after the intervention, and 3 months after the intervention. However, for the intervention to be considered successful and clinically meaningful, it is important to demonstrate that improved performance lasts for a significant amount of time.

Although there is evidence in the literature suggesting that GRS may be more sensitive and reliable summative measures of performance, the results of our study demonstrated that the checklist was more sensitive to changes between the groups than the GRS. The GRS is a summative, broad scale, and both groups improved, likely representing an individual’s independent learning between the two sessions or the increased comfort, and hence performance, with the simulated environment. The debriefing group received a focused debriefing, with only two or three learning points discussed to avoid cognitive overload during the debrief session. The facilitator chose the debriefing points based on the participant’s performance and may have included a discussion of technical skills, such as ACLS protocols, or non-technical skills such as communication and task delegation. The GRS which is a one-item, five-point response scale may not be sensitive enough to detect changes in these limited aspects of care, whereas the detail in the checklist scores may have this discriminatory capability. Alternatively, these aspects of performance may not be affected by our intervention.
There are numerous studies in the literature that demonstrate an improvement in performance in participants who have received simulator-based procedural training.25–29 There is also recent evidence that simulation-based instruction improves performance in actual clinical care settings.30 In these studies, however, the issue of whether debriefing specifically affects post-training outcomes is not addressed. Although there is speculation that simulation debriefing may improve the ‘non-technical’ aspects of human performance, there is limited scientific evidence in the literature to support this idea. However, debriefing has been shown to effectively improve performance of residents as evaluated by a valid, reliable behavioural marking system, the Anesthetists’ Non-Technical Skills (ANTS).31–32 Unfortunately, when this study commenced, the ANTS scale had not yet been developed, and therefore, it is unclear whether those subjects in the debrief group would have improved their non-technical skills when compared with the control group.

Owing to the fact that there were a small, select group of facilitators who were involved in the study, there was a concern that they might realize that they were witnessing a first or repeat session and therefore be biased in their assessment. For this reason, we trained two non-expert raters in the use of the assessment tools. After training and demonstration of a significant correlation of their scores with scores of two expert raters, not previously involved in the study, the non-expert raters viewed the videotapes of all 58 participants. The demonstration of acceptable to high inter-rater reliability between the expert and novice raters supports the accuracy of the performance ratings.

This study does not address the predictive validity of these findings to performance in clinical practice, but it would be impossible to conduct a randomized, controlled trial of this nature in an operating theatre setting. Nonetheless, the outcome of this study does demonstrate an improvement in performance in all subjects who received simulation training and a trend towards a further improvement in those who had the added debriefing component. Although it is unlikely that a single simulation training intervention will decrease errors in practice, further research may elucidate whether repetitive simulation-based training will improve anaesthetists’ performance and ultimately, patient safety.

Supplementary material
Supplementary material is available at British Journal of Anaesthesia online.

Acknowledgements
The authors would like to acknowledge the contributions of the following individuals: J. H. Devitt MD, MSc, FRCPQ E II, Dalhousie University, S. Bertram MD, FRCPQ, Sunnybrook Health Sciences Centre, University of Toronto, R. Ahranjani MD, R. Cheung, University of Toronto, and B. Morningstar MD, FRCPQ, Sunnybrook Health Sciences Centre, University of Toronto.

Funding
This work was supported by a research grant from the Canadian Institutes of Health Research (Grant #62688).

References
5 Cooper JB, Newbower RS, Kitz RJ. An analysis of major errors and equipment failures in anesthesia management: considerations for prevention and detection. Anesthesiology 1984; 60: 34–42
10 Salas E, Fowlkes JE, Stout RJ, Milanovic D. Does CRM training improve teamwork skills in the cockpit? Two evaluation studies. Hum Factors 1999; 41: 326–43
12 Issenberg SB, McGaghie WC, Petrusa ER, Lee Gordon D, Scalese RJ. Features and uses of high-fidelity medical simulations that lead to effective learning: a BEME systematic review. Med Teach 2005; 27: 10–28
Simulation education improves performance


