Preliminary Validation Results of DigEmergo for Surge Capacity Management

Jonas Rybing
Department of Computer and Information Science (IDA), Linköping University (LiU)
jonas.rybing@liu.se

Johan Larsson
Linköping University (LiU)
johla125@student.liu.se

Carl-Oscar Jonson
Department of Clinical and Experimental Medicine, Centre for Teaching & Research in Disaster Medicine and Traumatology, Linköping University (LiU)
carl-oscar.jonson@regionostergotland.se

Erik Prytz
Department of Computer and Information Science (IDA), Linköping University (LiU)
erik.prytz@liu.se

ABSTRACT

This paper presents preliminary analysis from a validation study of a novel emergency medicine command and control training and evaluation simulator: DIGEMERGO®. The simulated emergency scenario was a surge capacity event at a generic emergency department, in which the participants took on a management role as the emergency department's coordinating head nurse. A between group validation design with medical expert and novice participants was used. Initial analysis examined three triage measures associated with surge capacity management performance: time to triage, amount of patients triaged, and triage accuracy. The results show that experts were significantly more accurate at triaging in-hospital patients, but not incoming trauma patients. No significant differences in time or number of patients triaged was found. These initial results partially indicate simulator validity, but trauma patient triage accuracy suffered from a confounding variable in the triage system used. Analysis of additional measures is undergoing to further investigate validity claims.

Keywords

Simulator validation, between group analysis, command and control, performance measures, emergency medicine, surge capacity

INTRODUCTION

Virtual simulations are commonly used techniques for training and evaluating emergency medicine management skills (Andrews, Brown, Byrnes, Chang, and Hartman, 1998). Command and control (C2) is one essential component to efficiently manage emergency medicine (see e.g. Blackwell and Kaufman, 2002; Hick, Barbera, and Kelen, 2000; Brehmer, 1992), as it involves abilities such as understanding the situation, planning actions, making medical decisions, and coordinating resources. There are many different systems that are currently used worldwide to train medical management, e.g. Emergo Train System®[1], MACSIM[2], ISEE Hospital[3], and many more. However, there is a general lack of rigorous scientific validation of such simulators, despite the

[1] http://www.emergotrain.com/
[2] http://www.macsim.se/
[3] http://iseehospital.com/

importance of validation (Bewley and O'Neil, 2013; Feinstein and Cannon, 2002). Rather, these types of systems often rely on experience-based validation, meaning basically: if the system seems to work and is perceived as useful by users – the simulation is probably valid. However, this line of reasoning is scientifically problematic. The *validity* of a simulation can be described as the extent to which the simulation accurately captures those real life phenomena that are intended to be simulated. Feinstein and Cannon phrased *validation* as: "the process of assessing that the conclusions reached from a simulation are similar to those reached in the real-world system being modeled" (2002, pp. 427). In the case of training and evaluation simulations, this would translate to assessing that the skills being trained and evaluated in simulation are the same skills that would be applied in a corresponding real life scenario.

Validating training simulators used for C2 skills is no easy task. Most such studies rely on analysis of simulation output - i.e. performance measures (see e.g. Weaver, Dy, and Rosen, 2014). To validate training simulations, longitudinal studies measuring real life performance would be preferred, as these would indicate successful transfer of above-mentioned skills. However, longitudinal studies are impractical and costly in terms of time and resources. This is especially true for validating simulators under development, or for simulations that aim to train rarely used skills such as those required to manage the medical response to mass-casualty incidents. This is problematic as timely validation of systems under development is important to ensure operation functionality, efficiency, and to guide further development (Sokolowski and Banks, 2010). A different approach to validate training simulators is to compare the simulation performance of domain experts and novices (see e.g. Mathis and Wiegmann, 2007; Pucher, Aggarwal, Srisatkunam, and Darzi, 2014; Gallagher and Satava, 2002). In such validation studies the simulator is considered to exhibit validity if experts perform better than novices in the simulated scenario, as this would indicate that the simulator has captured those phenomenon that require domain expertise to be successfully executed. Controlled experiment validation studies are comparatively easy and fast to perform and should thus be suitable for validation of simulators under development, and to the authors' knowledge these are rarely performed for management oriented training simulators.

This paper presents the study design and preliminary results, focusing on patient prioritization (triage), from a validation study that adopted the expert – novice comparison approach to do initial validation of the DIGEMERGO® prototype system (Rybing, Nilsson, Jonson, and Bang, 2015a; Rybing, Prytz, Hornwall, Jonson, Nilsson and Bang, 2015b). DIGEMERGO is a digital simulator, based on the Emergo Train System® (ETS), for training and evaluating management aspects of emergency medicine. In this study the system was deployed as a simulator of emergency department surge capacity in which participants performed as the coordinating head nurse, tasked with managing the patient flow, including triage. The study was designed to have medical management expertise (C2 skills) as the independent variable and management performance (simulator output) as the dependent variable. The aim of the study was to do initial validation of the system and to test the applicability of the expert – novice validation approach on a medical management simulator under development. To do this, three quantitative measures (time to triage, number of patients triaged, and triage accuracy) of medical management were analyzed.

STUDY DESIGN

The Simulator

DIGEMERGO is a digitization and further development of the ETS-concept (Rybing et al., 2015b; Rybing et al. 2015a), which is an analogue (non-computer) simulator for training and evaluating medical emergency and disaster management skills. ETS is an internationally used technique and methodology to train and evaluate medical command and control, major civil incident doctrines, triage practice, patient management, medical surge capacity, and more (Nilsson, Vikström, and Rüter 2010; Nilsson, Jonson, Vikström, Bengtsson, Thorfinn,Huss, Kildal, and Sjöberg, 2013; Rybing et al., 2015a).

DIGEMERGO is built to run on one, or more depending on scenario size, Windows computers. Users interact with patients and resources (personnel, beds, rooms, etc.) figures via touch input on large multi-touch screens, see Figure 1. Users perform actions (triage, treat, and transport) on patients by first allocating the resources required via drag and drop and then selecting which actions are to be performed via a patient menu. The simulator is purposely designed to give a low-fidelity walk-up-and-use impression, similar to that of ETS (Rybing et al. 2015a). In this study, the participants worked on a 65" touchscreen (1920x1080 pixel resolution) that visualized a top view of the virtual emergency department.

Figure 1. DigEmergo® System. An expert participant is interacting with the patient menu.

Participants

The study consisted of two user groups: novices and experts. The novice group consisted of 21 participants (9 men and 12 women), and the expert group of 20 (6 men and 14 women), for a total of 41 participants.

The novice group was composed of students from different programs at a large university in southern Sweden. The age varied from 19 to 27 years (M = 23.9 years, SD = 2.2 years). None had official medical training, although two participants had worked as nurse´s assistants but not in an emergency department. Five had been trained in basic cardiopulmonary resuscitation. Self-assessed computer experience (on a scale of 1=very low to 5=very high) varied from 1 to 5 (M = 2.5, SD = 1.2) and self-assessed touchscreen experience varied from 3 to 5 (M = 4.7, SD = 0.6).

The expert group was recruited at a local hospital. All were registered nurses, physicians, or medical residents at the ED. 14 were head nurses, six of whom were also specialized emergency nurses. Three of the experts were emergency medicine residents, two were specialized prehospital nurses and one was a specialized emergency nurse. All had previous experience of working in the role as head nurse. Self-assessed computer experience varied from 1 to 5 (M = 2.0, SD = 1.3) and self-assessed touchscreen experience varied from 4 to 5 (M = 4.8, SD = 0.4).

Scenario and Procedure

The validation scenario was a mass casualty incident scenario in which a bus accident had occurred, an event designed to overwhelm the hospital's resources and stretch its surge capacity. Participants in both groups took the role as head nurse and were asked to handle the initial management of the patient flow in the hospital's emergency department (ED) to the best of their ability. The ED consisted of three trauma rooms, 11 examination rooms, one waiting room, one personnel room, and one ambulance hall. The head nurse were to perform in-hospital triage on patients (using the RETTS system, formerly METTS; Widgren and Jourak, 2011), allocate patients to rooms, allocate appropriate staff to the patients (medical doctors, nurses, etc.) to start treatments, and to dispose patients from the ED by either sending them home or to another ward. Patient triage is a medical decision process in which a set of patient parameters is examined to assign the patient a priority, commonly: red (most urgent), yellow, and green (non-urgent). The simulator was set up to automatically select appropriate treatments for each patient given the injuries of the patient. Each treatment required specific, allocated resources, e.g. a nurse to take samples, a medical doctor to do basic sutures, or a specialized medical doctor to treat specific injuries. Besides the trauma patients, arriving from the bus accident, the ED was populated by in-hospital patients that occupied the ED's resources from scenario start. Further in-hospital patients were waiting for medical assessment in the ED's waiting room.

Each session started by having the participant read and sign a written informed consent form. The participants then completed a demographics questionnaire. After the questionnaire, the participants read a short instruction about the upcoming scenario and how to operate the simulator. The participants performed a 5-minute introduction tutorial in DIGEMERGO, after which the 40-minute validation scenario was started. As part of the unfolding bus accident scenario, an experiment leader introduced a series of interventions during the scenario; e.g. new information from the scene of the accident, declaration of the mass casualty incident, and incoming trauma patients in need of treatment. These interventions were to challenge the participants' management skills and were the same for all participants, following a predetermined schedule.

Validation Measures

There is a multitude of parameters associated with surge capacity management performance that can be extracted and analyzed using the DIGEMERGO system. *Length of stay* and *patient volume* parameters along with *triage accuracy* are some of the most important and often used categories (Franc et al., 2014). Length of stay is divided into time from patient arrival to (1) triage; (2) room assignment; (3) assignment to medical staff; and (4) disposition. Patient volume included the total number of patients during the simulation that were; (1) triaged; (2) assigned to a room; (3) assigned to a medical doctor; and (4) disposed. These parameters are milestones in patient flow and important for both simulation and real-life environments (ibid.).

In the current, preliminary analysis time to triage, number of triaged patients, and triage accuracy data has been extracted and analyzed as measures of management performance. The simulation is assumed to exhibit validity with regards to *triage skills* if experts exhibit shorter *time to triage* for trauma and in-hospital patients, have a higher *number of patients triaged*, and have higher *triage accuracy* as compared to the novices. All data was logged by the DIGEMERGO system and analyzed using SPSS (v. 23.0.0.0). Two-tailed independent samples t-tests were used to test for differences between the groups with $p < 0.05$ as threshold.

RESULTS

There were no significant differences for *time to triage* of trauma patients, $t(30) = 1.216$, $p = 0.23$ (95% CI, -344.42, 87.39), between the novices (M = 225.12 seconds, SD = 244.66) and the experts (M = 353.63 s, SD = 338.55). Similarly, there were no significant differences in time to triage the in-hospital patients, $t(35)= .14$, $p = 0.89$ (95% CI, -277.79, 319.93), between the novices (M = 1239.42 s, SD = 463.10) and the experts (M = 1218.35 s, SD = 431.58).

There were no significant differences for *number of triaged* trauma patients, $t(39) = 1.130$, $p = 0.27$ (95% CI, -1.55, 0.44), between the experts (M = 2.65, SD = 1.42) and the novices (M = 2.10, SD = 1.70). Nor was there any significant difference in number of triaged in-hospital patients, $t(39) = .039$, $p = 0.483$ (95% CI, -3.50, 3.37), between the experts (M = 8.40, SD = 5.71) and the novices (M = 8.33, SD = 5.12).

Further, there were no significant differences in *triage accuracy* for trauma patients, $t(30)= 0.480$, $p = 0.64$ (95% CI, -1.786, 0.111), between the experts (M = 7.84%, SD = 25.08) and the novices (M = 4.44%, SD = 11.73). However, there was a significant difference for triage accuracy of in-hospital patients, $t(36)= 2.208$, $p = .03$ (95% CI, -0.364, -0.015), between the experts (M = 62.91%, SD = 25.68) compared with the novices (M = 43.94%, SD = 27.28).

DISCUSSION AND CONCLUSION

The main conclusion of these initial results is that there is mixed evidence of simulator validity in terms of triage. The higher in-hospital patient triage accuracy for the experts indicates simulator validity as it exhibits application of medical knowledge. However, trauma patient triage accuracy and number of patients triaged did not reveal any differences, and the time to triage of trauma patients were in favor of the novices although not significantly so. These latter results require further attention.

First, both groups scored overall very low on triage accuracy for trauma patients, and especially experts scored far below what would be expected. There are two probable explanations for this. The first is that the number of trauma patients in need of first triage were too few to generate statistical power, as most arrived to the ED already triaged. The only trauma patients without triage were four patients who walked to the ED from the bus crash. As not all participants triaged all, or even any, of these patients the available data is limited and the statistical power low. Second, a confounding variable is present in the RETTS triage system. The RETTS triage system was applied in this study as it is the most common ED Triage system in Sweden and the system the experts should be familiar with. However, according to RETTS, all trauma patients from a high-energy traffic accident should be assigned an orange or red code (highest two priorities) regardless of medical status, which the experts apparently did not adhere to. Overall the experts typically assigned a green triage to the group of trauma patients that arrived to the ED by walking. Excluding this traffic accident rule and only looking at the medical status of the patients (or using another triage system) expert triage accuracy would dramatically increase, and possibly reveal group differences, as these trauma patients only had minor injuries. Differences could probably also be found in how experts and novices allocated resources in relation to triage of trauma patients, which has yet to be analyzed.

There was no significant difference in time to triage, but if anything the novices tended to be faster than experts in the current sample, despite similar self-reporter computer and touch-screen experience. This gives rise to an interesting notion: that another aspect present in the novice group improves their performance - for instance video game skills. The DIGEMERGO system share features reminiscent of video games and serious games. Commercial video games have shown widespread application for C2 training and have been correlated with enhanced related skills, such as visuo-spatial attention (Nählinder, 2007). If this is the case, the assumption that experts should perform faster than (these) novices does not hold and thus time to triage and number of patients triaged would not be an appropriate management validation measures. Unfortunately, no data regarding the participants' familiarity with video games or gaming in general was gathered. A factor of the simulator that might have affected the results in favor of novices is the resolution (amount of details present regarding heath status) of simulated patients. Currently, this resolution may be too low which makes the triage procedure less demanding. Patient injuries and medical status were represented by static textual descriptions, which enabled a 'sounds bad - is bad' triage rule requiring less medical expertise. As the patient information were static the patients' health state could not change, which eliminated the need of expertise skill to perceive that a patients state could deteriorate and the need for re-triage. It would be interesting future work to study how the groups would perform given more complex and dynamic patients where re-triage would be required.

Based on the above considerations the immediate future work will be to: 1) analyze additional performance measures (e.g. patient disposal) and look at trends over time between groups that could further determine simulator validity, 2) supplement future studies with video game familiarity demographic data of the participants. In conclusion, the experts exhibited better performance in terms of triage accuracy. This indicates that the DigEmergo simulator exhibits validity for tasks that require a high degree of medical decision-making. However, for performance measures that does not rely on medical knowledge (e.g., speed of actions) there is no difference between novices and experts. This can be explained by shortcomings in the study design, the triage system used, and confounding variables in the participants. Further, time and quantity might not be the best indicators for performance differences of medical C2 skills; this however remains to be determined by future research. Finally, this study demonstrated that the between-group validation method is able to provide useful indicators of validity for management oriented training simulators.

REFERENCES

1. Andrews, D. H., Brown, J., Byrnes, J., Chang, J., & Hartman, R. (1998) Enabling Technology: Analysis of Categories with Potential to Support the Use of Modeling and Simulation in the United States Air Force, Air Force Research Lab Mesa Az Human Effectiveness Directorate.
2. Bewley, W. L. and O'Neil, H. F. (2013) Evaluation of medical simulations, Military Medicine, 178, 10, 64–75.
3. Blackwell, T. H. and Kaufman, J. S. (2002) Response time effectiveness: Comparison of response time and survival in an urban emergency medical services system, Academic Emergency Medicine, 9, 4, 288-295.
4. Brehmer, B. (1992) Dynamic decision making: human control of complex systems, Acta Psychologica, 81, 3, 211-241.
5. Feinstein, A.H. and H.M. Cannon. (2002) Constructs of Simulation Evaluation, Simulation & Gaming, 33, 4,

425-440.

6. Franc, J. M., Ingrassia, P. L., Verde, M., Colombo, D. and Della Corte, F. (2014) A Simple Graphical Method for Quantification of Disaster Management Surge Capacity Using Computer Simulation and Process-control Tools, Prehospital and Disaster Medicine, February, 1-7.
7. Gallagher, A. G. and Satava, R. M. (2002) Virtual reality as a metric for the assessment of laparoscopic psychomotor skills. Learning curves and reliability measures, Surgical Endoscopy, 16, 12, 1746-1752.
8. Nählinder, S. and Oskarsson, P.-A. (2007). Learning from Computer Games. FOI-R--2396--SE., FOI, Swedish Defence Research Agency, Linköping, Sweden.
9. Hick, J. L., Barbera, J. A., & Kelen, G. D. (2009) Refining surge capacity: conventional, contingency, and crisis capacity, Disaster Medicine and Public Health Preparedness, 3, 2, 59–67.
10. Mathis, K. L., and Wiegmann, D. A. (2007) Construct validation of a laparoscopic surgical simulator, Simulation in Healthcare, 2, 3, 178–82.
11. Nilsson, H., Vikström, T. and Rüter, A. (2010) Quality control in disaster medicine training--initial regional medical command and control as an example, Am J Disaster Med, 5, 1, 35-40.
12. Nilsson, H., Jonson, C.-O., Vikström, T., Bengtsson, E., Thorfinn, J., Huss, F., Kildal, M. and Sjöberg, F. (2013) Simulation-assisted burn disaster planning. Burns: Journal of the International Society for Burn Injuries, 39, 6, 1122-1130.
13. Peleg, K., & Kellermann, A. L. (2009) Enhancing hospital surge capacity for mass casualty events. JAMA : The Journal of the American Medical Association, 302, 5, 565–567.
14. Pucher, P. H., Aggarwal, R., Srisatkunam, T. and Darzi, A. (2014) Validation of the simulated ward environment for assessment of ward-based surgical care, Annals of Surgery, 259, 2, 215-221.
15. Rybing, J., Nilsson, H., Jonson, C.-O. and Bang, M. (2015a) Studying distributed cognition of simulation-based team training with DiCoT, Ergonomics, September 2015, 1-12.
16. Rybing J., Prytz E., Hornwall J., Jonson C-O., Nilsson H. and Bang M. (2015b) Preliminary evaluation results of DigEmergo - a digital simulator prototype for disaster and emergency management training, Proceedings of the 19th World Congress on Disaster and Emergency Medicine (WCDEM 2015), Capetown, South Africa.
17. Sokolowski, J. A., & Banks, C. M. (2010) Modeling and Simulation Fundementals: Theretical Underpinnings and Practical Domains, John Wiley & Sons, Inc., Hoboken, New Jersey.
18. Weaver, S. J., Dy, S. M. and Rosen, M. (2014) Team-training in healthcare: a narrative synthesis of the literature, BMJ Quality & Safety, 23, 5, 359-372.
19. Widgren, B. R. and Jourak, M. (2011) Medical Emergency Triage and Treatment System (METTS): A New Protocol in Primary Triage and Secondary Priority Decision in Emergency Medicine, The Journal of Emergency Medicine, 40, 6, 623-628.