Comparing de-congestion scenarios using a hospital event simulation model

W. Hou^a, <u>S. Qin</u>^a and C. Thompson^b

^a College of Science and Engineering, Flinders University, Australia ^b School of Medicine, University of Adelaide, Australia Email: <u>shaowen.qin@flinders.edu.au</u>

Abstract: The ever increasing demand for hospital services has resulted in more and more congestion episodes occurring in hospital emergency departments (ED). Congestions lead to serious problems such as delayed treatment, increased mortality, as well as stressed hospital staff. It is thus important to explore suitable solutions to manage congestions when they occur. This study seeks to identify effective ways to resolve the serious situation of congestion in a hospital by testing a range of de-congestion strategies using simulated scenarios. All scenarios were investigated through a sophisticated simulation model HESMAD (Hospital Event Simulation Model: Arrivals to Discharge) that captures the characteristics of patient flow based on an existing patient journey database from a large tertiary hospital in South Australia. Simulation models can demonstrate the changes and the impacts of different operational parameters within a complex hospital system before, during and after congestions. In this study, eight scenarios, suggested by senior hospital staff, were investigated using the simulation model, in addition to the base scenario where no intervention was applied. These scenarios include diverting ambulances, postponing elective patients, and discharging certain groups of patients earlier when congestions occur. Parameters such as the number of congestions over a one-year period, the duration of a congestion, and the 10 am hospital occupancy of each scenario were examined. Simulation of each scenario was replicated 20 times under the same conditions to obtain an average behaviour that would allow meaningful comparison of the results from different intervention scenarios. The results show that, for the scale of actions taken, hospital occupancy remained almost the same for all scenarios, while the duration of congestions and the frequency of their occurrence exhibited different levels of reduction. The scenario of temporarily diverting ambulances was most effective in reducing the number of congestion days (from 76.5 days in the base case to 44.2 days; a reduction of over 44%). However, other simulation scenarios, such as removing particularly long-staying patients and postponing the admission of elective patients may be preferable to diverting new emergency admissions when trying to reduce congestion duration and frequency. Although the aim was to shorten the duration of a congestion when it occurred, an important side benefit was observed for all interventions tested, that is, the number of congestion episodes in the long term (over one-year simulation period) was also reduced.

Keywords: Hospital Event Simulation Model, Congestion Days, De-congestion Scenario Analysis

1. INTRODUCTION

It is a fact that the demand for hospital services has been increasing more rapidly in this modern era. Consequently, a larger number of congestions occur and this leads to serious problems such as delayed treatment, increased patient mortality and hospital infections (e.g. Virtanen, 2008 and Richardson, 2006). Thus, exploring valuable solutions to tackle this negative issue is becoming crucial. Existing studies have discussed access block and its effects upon patients' mortality (e.g. Sprivulis, 2006 and Richardson, 2006) and inpatient length of stay (LOS) (e.g. Liew, 2003 and Richardson, 2002). Among responses to overcrowding in Australian hospitals (Richardson, 2003), attention has generally been focused on bed usage aiming to achieve efficient inpatient flow and reduce hospital congestion (e.g. Bagust, 1999 and Green, 2002). Simulation modeling, an important systems analysis tool, is particularly suitable for examining the behaviour of a complex system as a whole while demonstrating the causal relationships among the issues experienced by its parts. It is a risk free approach to determine the relative effectiveness of isolated specific interventions aiming towards system improvement. Consequently, an increasing number of simulation studies have been conducted in hospital settings, such as those to examine causes and effects of overcrowding (e.g. Derlt, 2001, Fatovich, 2003 and Trzeciak, 2003) and those to test measures to address problems of hospital Emergency Departments (Asplin, 2003, Samaha, 2003).

In this study, instead of considering a single department in a hospital, a comprehensive simulation model named HESMAD (Hospital Event Simulation Model: Arrivals to Discharge) that captures the patterns of patient flow through a large tertiary hospital in South Australia (Ben-Tovim, 2016) was used to investigate the impact of a range of de-congestion strategies. Simulated scenarios corresponding to the implementation of these proposed strategies were tested, with the aim of comparing the effectiveness and efficiency of different de-congestion strategies activated after congestions occur in the hospital. Parameters such as the number of congestions over a one-year period, the duration of a congestion, number of patients affected, and the 10 am hospital occupancy of each scenario are used as bases for such comparisons.

2. METHODS

HESMAD (Hospital Event Simulation Model: Arrivals to Discharge) is a computer simulation model that captures the patterns of patient flow through a large tertiary hospital in South Australia (Ben-Tovim, 2016). HESMAD adopted a modular structure (a module represents either a physical unit or a process) and captures the journey of patients through the hospital from the time they first present at the ED, or are admitted as elective patients, to the time that they are discharged. It models all patients in the ED and all specialties, surgical, general medical and overnight elective (surgery & medicine) patients. Patient "agents" are defined by a series of parameters which are assigned at their creation. Such parameters include age, arrival method and mode and their triage status. Such parameters are used for tracking milestones in the patient journey. These milestones include which sections of the hospital they enter and at what times they complete various aspects of their journey. The model allows for a realistic representation of patient flows, at a level of resolution that was deemed appropriate, by the hospital's data management experts, for balancing complexity and fidelity. The model has been validated against historical data and through consultation with health care and hospital experts. An overview of the structure of HESMAD is given in Figure 1.



Figure 1. Overview of the HESMAD structure.

In this study, simulations of different scenarios were run to demonstrate their impacts upon different operational parameters within a complex hospital system before, during and after congestions. Eight scenarios, suggested by senior hospital staff, were investigated using the simulation model, in addition to the base scenario where no intervention was introduced when congestions occur. The scenarios include diverting ambulances, postponing elective patients, and discharging a small selected group of patients earlier when congestions occur. Each scenario simulates two years of hospital operations. However, only the results of the second year were collected to minimise the impact of the 'warm-up' period. In addition, each scenario was replicated 20 times under the same conditions to obtain an average behaviour that would allow meaningful comparison of the results from different intervention scenarios.

2.1 De-congestion Scenarios

All de-congestion scenarios proposed in this study are listed as follows:

- Scenario 0: Base case/Standard Hospital Operation (no interventions)
- Scenario 1: Moving 15% of long stay patients (LOS >= 21days) to a peripheral hospital
- Scenario 2: Postponing 15% of all elective admissions
- Scenario 3: Postponing 15% of elective surgery admissions
- Scenario 4: Postponing 15% of elective surgery with an anticipated LOS > 1 day
- Scenario 5: Diverting 20% of ambulance arrivals
- Scenario 6: 24 hours/day discharge
- Scenario 7: 24 hours/day discharge of patients younger than 65 years
- Scenario 8: Removing 15% of long stay patients (LOS >= 21 days) and diverting 20% of ambulance arrivals at the same time (a combination of scenarios 1&5)

The interventions were activated after a congestion occurs. Without doubt, defining the criteria for what is considered a congestion in a hospital was essential for this study. Unfortunately, there is no widely accepted criteria we could reference and apply to the simulation to determine the onset of a congestion. We selected three indicators for this purpose. They are: occupancy level of inpatient beds, utilization of treatment space in ED and the number of patients in the queue (queue length) waiting for ED treatment. We assumed that the frequency of congestion in the base case (Scenario 0) is approximately once a week, almost all spaces (>=17 beds) in ED admission side (18 beds in total) are utilized, and inpatient bed occupancy rate >= 0.97. Using a so-called reverse engineering approach, an average queue length of 28 was determined. In other words, one can set thresholds for ED and hospital occupancy, as well as a frequency of congestions per simulation period (e.g., one to two times per week) in the base scenario, and derive an average queue length for congestions of such severity. This queue length, combined with the corresponding levels of ED and hospital occupancy, is then applied to other scenarios to identify whether a congestion of such severity has occurred or not. It is important to keep in mind that the definition of a congestion will always be subjective. For the purpose of comparing the effects of different de-congestion strategies on a relative scale, it suffices to apply a consistent definition throughout all scenarios. For example, we can set the queue length criterion to a smaller number with congestions occurring many more times than once a week during the simulation period, and still compare the relative effectiveness of the de-congestion strategies.

Scenario 1 involved moving 15% of long stay patients (LOS ≥ 21 days) to other care facilities when congestion happens. In this scenario, the transfer will start 1 hour after the onset of a congestion because it will take at least this amount of time for hospital staff to check the availability of other care facilities. Scenario 2 to 3 tested the effects of deferring the admission of elective patients when congestions occur. For the realization of scenario 2, the model postponed admission of 15% of all medical and surgical elective patients during a congestion episode. Instead of removing these patients, the model put them in a backlog of elective patients waiting for rescheduling when beds become available in the hospital. Scenario 3 is slightly different from Scenario 2 because this scenario only postponed 15% of elective surgical patients. Compared to Scenario 3, Scenario 4 postponed 15% of elective surgical patients with an anticipated LOS over 1 day owing to the fact that about 52% elective surgical patients complete their treatment in less than 1 day.

Scenario 5 investigated the impact of diverting ambulance arrivals during a congestion episode. The model reduced ambulance arrivals by 20% for 2 hours each time congestion criteria are met (checked every hour). Scenario 6 estimated the scenario of changing hospital processes when congestion happens. In this scenario staff are able to discharge patients outside business hours when they complete their treatment. Compared on Scenario 6, Scenario 7 only applied after hours discharge to patients who are younger than 65 years. Scenario 8 tested the combination of two de-congestion scenarios 1 and 5, that is, removing 15% of long stay patients as well as reducing 20% of ambulance arrivals for two hours.

Hou et al., Comparing de-congestion scenarios using a hospital event simulation model

2.2 Congestion Measures

The effect of each scenario on de-congestion was mainly estimated through the number of congestion days and congestion frequency over a one-year period. A congestion day is defined as a day when one or more congestions had happened during that day. The fewer the congestion days, the more preferable the scenario might be. Therefore, the number of congestion days is one of the more important measures to gain insight into the potential impacts of each de-congestion strategy. Furthermore, congestion time represents the duration of congestion. Reducing congestion time allows the hospital to operate at its designed capacity and maintain the quality of its service. Hence it is desirable to decrease the duration of each congestion. To provide a more objective comparison of the scenarios, the efficiency of each de-congestion strategy, defined as the relative reduction of congestion day or congestion time per patient affected (expressed as a percentage), was also calculated and shown in Figure 2 and 3. Apparently, the most effective de-congestion strategy would be the one that affects the fewest patients but achieves best results in terms of reducing congestion frequency or duration.

Other measures reported were 10 am hospital occupancy, patients discharged early, patients discharged outside of business hours, postponed elective patients and diverted ambulance arrivals. Occupancy was defined as the number of beds occupied in the hospital (330 base beds and 8 flex beds in the modeled hospital for the year). The model calculated 10 am hospital occupancy because the period between 10 am and 12 noon is the time when overcrowding in the hospital tends to occur. The metric "patients discharged early" shows the average number of patients discharged in Scenarios 1 and 8. "Patients discharged outside of business hours" is an indicator to count how many patients were discharged after normal operating hours in Scenarios 6 and 7. The model also recorded the metric "postponed elective patients" to record the total number of elective patients postponed in response to a congestion episode in Scenarios 2-4. Diverted ambulance arrivals represents the number of ambulance diversions in Scenarios 5 and 8. Not all of these diverted patients would have been admitted to hospital. All of these parameters were utilised to measure the number of patients affected by each scenario and hence allow calculation of the efficiency of the different de-congestion strategies.

	Scenario 0	Scenario 1	Scenario 2	Scenario 3	Scenario 4	Scenario 5	Scenario 6	Scenario 7	Scenario 8
Mean Occupancy(10AM) (Number of beds)	327.96	326.27	327.06	327.54	327.89	327.08	326.68	326.80	325.67
Mean Std Dev (Number of beds)	9.83	9.82	10.13	9.37	9.75	9.90	10.14	9.97	10.31
Number of Congestion Day (Days)	76.50	51.65	54.75	63.65	64.15	44.2	57.55	64.52	38.37
Maximum Congestion Time (Hours)	43.60	30.39	30.20	33.05	40.90	38.55	45.80	51.60	27.50
Minimum Congestion Time (Hours)	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00
Mean Congestion Time (Hours)	9.20	7.96	7.74	8.02	8.72	8.03	8.12	8.74	6.77
Median Congestion Time (Hours)	7.08	5.67	5.85	6.18	6.55	5.25	5.43	6.93	4.65
Patients Discharged early (Persons)	-	52.90	-	-	-	-	-	-	53.8
Patients Discharged Outside of Business Hours (Persons)	-	-	-	-	-	-	76.35	47	-
Postponed Elective patients (Persons)	-	-	45.70	25.95	17.69	-	-	-	-
Diverted ambulance arrivals (Number of Ambulances)	-	-	-	-	-	233.38	-	-	166.50

Hou et al., Comparing de-congestion scenarios using a hospital event simulation model

3. **RESULTS**

The simulation results of different de-congestion scenarios are presented in Table 1. The mean hospital occupancy at 10am is shown in row 1, with its standard deviation shown in row 2. Row 3 presents the average number of congestion days. The next four rows are measures of congestion duration in hours.

It can be seen that applying different de-congestion strategies has little effect on 10 am hospital occupancy. The best scenario (Scenario 8) only caused about 0.7% decrease in occupancy (from 328 inpatients in the base case to 326). By contrast, these scenarios exerted noticeably different effects on congestion duration and frequency, many of which would carry clinical significance. The combination of removing 15% of long stay patients and reducing ambulance arrival rate by 20% (Scenario 8) resulted in the greatest reduction (~50%) in the number of congestion days (from 76.5 in base case to 38) and in the mean congestion time (~25% from 9.2 hours in base case to 6.8 hours). Implementing Scenario 5 also seems to be a profitable way to decrease congestion frequency - the number of congestion days can be brought down from 76.5 to 44 by diverting 20% of Ambulance arrivals while a congestion episode lasts (Scenario 5). However, this intervention has less effect on mean congestion time (about 13 % reduction). Separately removing 15% of long stay patients and postponing 15% all elective patients (Scenario 1 and 2) each reduces the number of congestion days (by 32% and 28% respectively). These two scenarios are also able to lower the mean congestion duration (Row 6) by about 13% and 16% respectively. 24 hours/day discharge for all patients or just those younger than 65 years (Scenario 6-7) can shorten congestion frequency (around 25% and 16% respectively).

The efficiency of each de-congestion scenario, expressed as percentage improvement per patient affected, is shown in Figures 2 and 3. Figure 2 shows that postponing 15% of elective surgical patients with a LOS over 1 day (Scenario 4) is the most effective way to reduce congestion frequency while affecting the least number of patients. Figure 3 illustrates that postponing 15% of all elective surgical patients irrespective of their anticipated LOS (Scenario 3) produces the greatest improvement in shortening congestion duration. Discharging patients younger than 65 years around the clock (Scenario 7) produced the least impressive improvements in shortening congestion duration.



Figure 2. Improvement/person in number of Congestion Days

Figure 3. Improvement/person in Congestion duration

4. **DISCUSSION**

This study aimed to investigate potential solutions to reduce the number and duration of congestions in today's hospitals. A selection of scenarios corresponding to the implementation of certain de-congestion strategies when congestions occur were investigated using HESMAD, a hospital simulation model. The findings indicated that all de-congestion interventions are able to improve the annual congestion frequency and duration to a variety of extents although the efficiency of each of these scenarios is distinct especially if expressed as a function of the patients affected by each scenario. Unfortunately, there is ongoing congestion in most hospitals despite the implementation of different scenarios such as those presented in this paper (Richardson, 2003). These scenarios consume different degrees of resources and affect differing numbers of current or future patients. Hence, not only paying more attention to decongestion scenarios but also investigating their efficiency further may optimize management of hospitals and lead to a higher quality of hospital service. In addition, modelling may allow us to determine whether certain de-congestion scenarios work better than others when applied in parallel.

This study estimated different methods to tackle hospital overcrowding whether it was by removing long stay patients, postponing elective patients, diverting ambulances or implementing a 24 hours/day discharge policy. Arguably, the best de-congestion scenario affects the fewest patients but achieves best results in terms of reducing congestion frequency or duration. This is because the more patients are affected, the more patient complaints may occur and the more delays to treatment such as elective procedures.

4.1 Moving long-staying patients

Moving a certain percentage of long stay patients to a peripheral hospital is one possibility to reduce hospital-overcrowding situations. Based on our results, removing as few as 53 long-staying patients over a year (Scenario 1) reduces congestion frequency by about one third. Moreover, the efficiency of this scenario (Figure 2 and 3) to reduce congestion frequency and duration is also high meaning very few patients are affected when considering the magnitude of the effect on congestion. Although the ratio of long stay patients (LOS exceeds 21 days) was only about 2% of all inpatients, this small number of patients significantly impacts congestion circumstances. Many long-staying patients in hospital remain there for administrative purposes rather than clinical; they lack a sufficiently-supportive low acuity discharge destination and do not suffer from acute illness.

4.2 Managing elective patients

The hospital patient journey data shows that the majority of elective arrivals occur at or close to 9AM and 3PM. Accordingly, the postponement of elective patients was also effected between these times in the model. Therefore, only a limited number of elective patients were postponed in this study. Our results indicated that postponing about 45 elective patients over a year could result in clinically important improvements in congestion frequency (28%) and duration (16%). Moreover, the efficiency of postponing admission of elective patients, particularly elective surgery patients, is even higher than moving long stay patients. These postponed patients may be disgruntled, eventually need admission and the timing of that subsequent admission is not straight-forward. Consequently, postponing admissions of elective patients may bring outstanding benefits in respect of congestion reduction for the hospital but comes at a poorly-definable cost.

4.3 Ambulance Diversion

Diverting ambulance arrivals will decrease congestion frequency but not greatly affect hospital congestion duration despite this scenario being popular to reduce congestion acutely in the ED. Considering that many of these diverted patients may not be admitted in the first place, this scenario has the potential to inconvenience many with little chance of improving hospital congestion duration. However, fewer ambulance arrivals over a year will have the cumulative effect of a reduced number of congestion episodes. The combination of ambulance diversion and removing long stay patients (Scenario 8) improved congestion metrics more than either intervention on its own though the effects were neither summative nor synergistic. A 24 hours discharge scenario also has limited effects on changing the overcrowding situation in the hospital yet has the potential to displace frail elderly patients into a less-supported environment at a potentially "unhealthy" time of the day.

5. CONCLUSION

In conclusion, controlling the admission of elective patients especially elective surgery patients and transferring out long stay patients are effective ways to favourably change congestion circumstances in the hospital. It is highly recommended that clinical administrators pay more attention to the disposition of elective and long stay patients.

Although the aim was to shorten the duration of a congestion when it occurred, an important side benefit was observed for all interventions tested, that is, the number of congestion episodes in the long term (over a one-year simulation period) was also reduced. This is where simulation shows its advantage. It can reveal both short and long-term impacts of any intervention, and exhibit counter-intuitive emergent system behaviours.

ACKNOWLEDGEMENT

This work was supported by the ARC linkage grant LP130100323, jointly awarded to Flinders University, the Southern Adelaide Health Service (Flinders Medical Centre) and the Central Adelaide Local Health Network (Royal Adelaide Hospital).

Hou et al., Comparing de-congestion scenarios using a hospital event simulation model

REFERENCES

- Asplin, B.R., D.J. Magid, K.V. Rhodes, et al (2003). A conceptual model of emergency department crowding. Annals of Emergency Medicine, 42(2): 173–180.
- Bagust, A., M. Place, and J.W. Posnett (1999). Dynamics of bed use accommodating emergency admissions: stochastic simulation model. *BMJ*, 319, 155-158.
 Ben-Tovim, D., J. Filar, P. Hakendorf, S.W. Qin, C. Thompson and D. Ward (2016). Hospital Event
- Ben-Tovim, D., J. Filar, P. Hakendorf, S.W. Qin, C. Thompson and D. Ward (2016). Hospital Event Simulation Model: Arrivals to Discharge- Design Development and Application. *Journal of Simulation Modeling Practice and Theory*, 68(11), 80-94.
- Derlet, R.W., J. R. Richards and R.L. Kravitz (2001). Frequent overcrowding in U.S. emergency departments. *Academic Emergency Medicine*, 8(2): 151–155.
- Fatovich, D.M. and R.L. Hirsch (2003). Entry overload, emergency department overcrowding, and ambulance bypass. *Emergency Medicine Journal*, 20: 406–409.
- Green, L.V. (2002-2003). How many hospital beds? Inquiry, 39(4), 400-412.
- Liew, D., D. Liew and M.P Kennedy (2003). Emergency department length of stay independently predicts excess inpatient length of stay. *Medical Journal of Australia*, 179(10), 524-526.
- Richardson, D.B (2002). The access-block effect: Relationship between delay to reaching an inpatient bed and inpatient length of stay. *Medical Journal of Australia*, 177(9), 492-495.
- Richardson, D.B (2003). Responses to access block in Australia: Australian Capital Territory. *Medical Journal of Australia*, 178(3), 103-104.
- Richardson, D.B (2006). Increase in patient mortality at 10 days associated with emergency department overcrowding. *Medical Journal of Australia*, 184(5), 213-216.
- Trzeciak, S. and E.P. Rivers (2003). Emergency department overcrowding in the United States: an emerging threat to patient safety and public health. *Emergency Medicine Journal*, 20: 402–405.
- Virtanen, M., J. Pentti, J. Vahtera, J.E. Ferrie, S. A. Stansfeld, H. Helenius, M. Elovainio, T. Honkonen, K. Terho, T. Okasanen and M. Kivimaki (2008). Overcrowding in hospital ward as a predictor of antidepressant treatment among hospital staff. *The American Journal of Psychiatry*, 165, 1482-1486.
- Samaha, S., W.S. Armel and D.W. Starks. 2003. Emergency departments I: the use of simulation to reduce the length of stay in an emergency department. Proceedings of the 35th Conference on Winter Simulation, pp. 1907–1911.
- Sprivulis, P. C, J.D. Silva, L.G. Jacobs, A.R. Frazer and G.A Jelinek (2006). The association between hospital overcrowding and mortality among patients admitted via Western Australian emergency departments. *Medical Journal of Australia*, 184(5), 208-212.