Simulating hospital patient flow for insight and improvement

Anthony Clissold1, Jerzy Filar1, Mark Mackay2, Shaowen Qin1, Dale Ward1
1 School of Computer Science, Engineering and Mathematics
Flinders University of South Australia
2 School of Medicine
Flinders University of South Australia
anthony.clissold@flinders.edu.au, jerzy.filar@flinders.edu.au,
shaowen.qin@flinders.edu.au, mark.mackay@flinders.edu.au, dale.ward@flinders.edu.au

Abstract
It is a well-known problem that Australian public hospitals are struggling to meet the increasing inpatient care demand, which often results in patient flow congestion and poor health care outcomes. Health care management practitioners and researchers have been trying to improve overall flow of patients with various modelling and data analysis approaches. The implementation and uptake of modelling as part of normal hospital management practice has, however, not been overly successful to date for a variety of reasons. Consequently, there remains a gap between the development of models and the buy-in from health care professionals in using them for management decision-making. This paper shares our recent experience of creating a discrete event based simulation model of patient flow in Flinders Medical Centre (FMC) that accurately describes the progression of patients through the emergency department (ED) into wards and then discharge. It describes the development of various aspects of the simulation, including data preparation, arrival patterns, queuing, ward allocation, weekend discharge delay, and model validation. The model has been used to test the impact of the likely increased ED arrivals arising from proposed Government policy change. It will be used to investigate other scenarios and strategies for improving patient flow through the hospital and reducing waiting times in the emergency department. The animated visual representation of such simulation has proven to be an effective way to engage hospital staff, which should lead to better buy-in of improvement changes.

Keywords: Simulation modelling, stochastic process, patient flow, hospital improvement.

1. Introduction
1.1 The problem
The Australian population is continuing to grow along with its average life expectancy. According to Kowal, Towers and Byles (2014), the population aged 60 years or more is set to grow by more than 80% between 2013 and 2050. Over the same period, the number of people aged 80 or more years is expected to rise by 200%.

A recent study found that a half of the Australian population had at least one chronic health condition such as cardiovascular disease, arthritis, endocrine/nutritional/metabolic diseases, or psychological conditions (Harrison et al., 2013).

Kowal, Towers and Byles (2014) suggest that during this period of population ageing, and assuming that all current factors of health use and provision remain the same, an increase in use of the health system would be expected. Not surprisingly, Fitzgerald et al. (2012) has reported that ageing of the Australian population and the prevalence of chronic disease has and will continue to contribute to higher demands for emergency and inpatient services in the country's hospitals.

Currently many metropolitan teaching hospitals are often operating at 90 per cent or greater occupancy (AMA, 2013). Forster et al. (2003) has previously demonstrated that the time spent in the emergency department is influenced by the hospital’s occupancy. Most importantly the extension of time spent in the ED is greatest when hospitals operate at more than 90% of occupancy. Schilling et al. (2010) have found that patient outcome, as measured by in-hospital mortality, is associated with hospital occupancy. The AMA (2013) also reported that high levels of hospital occupancy are factors affecting ED waiting times and patient safety. Consequently there is a need for decision tools that can facilitate the improved understanding and management of hospital occupancy and patient flow.

1.2 Discrete Event Simulation of Health Services
Discrete Event Simulation (DES) is a technique that can be used to create and visualise models of complex workflow based systems. Simulations of this form are built up from a sequence of discrete events that occur at different frequencies with various constrains. Events can occur deterministically and stochastically in DES, which allows simulation of stochastic processes. Multiple entities, such as patients, can travel through the system process, and many types of statistics related to the entities in the simulation, such as waiting times, resource use, queue sizes, etc., can be collected. The collection of statistics, in turn, allows for the results of the simulation to be validated against real-world data to test how realistic the simulation is with respect to the process being modelled. Once an adequate model has been created, multiple scenarios can be tested to quantify what is their effect on
These simulations can be run quickly allowing for the study of a system’s variability and underlying pattern to take place in a virtual environment. This is obviously an advantage over experimenting with the real system where intervention is required and we can only study at most a handful of instances.

DESs are becoming more and more frequently used (Brailsford et al. 2009; Fone et al. 2003, Gunal and Pidd 2010) in health care settings. A more recent search of Google Scholar of patient flow related terms (e.g., emergency department, access block, patient flow, hospital) with modelling terms (e.g., DES), shows that academic interest, based on the publication rate, has increased dramatically since the year 2000 in general, and also in Australia (see Figure 1).

1.3 Translation of Research into Practice

Lewis (2011) has commented upon the use and impact of health services research. He noted that while the Dartmouth studies on healthcare use and cost had previously identified significant variations across the U.S., the same issues continue to exist 30 years later. The identification of problems in the health services and the development of better tools to understand these problems have not yet translated into meaningful changes.

Concerns regarding the translation of modelling efforts into practice have also been previously raised. For example, Fone et al. (2003) have also noted that while there had been a growth in health care modelling, relatively little information was available on the usefulness of the models in practice.

1.4 The DES Patient Flow Model

Our work attempts to describe not only the creation of a DES patient flow model, but also how it can be used for practical purposes. Using the simulation, it is be possible to quantify the effects of policy changes on the underlying system, while still controlling for the other variables present in the model. These model/results can be given to staff in the hospital to better inform them on the impact of a policy change or to inform policy debate in other forums.

2. Model Description

2.1 The Data

The data used in this project were drawn from a hospital patient database. This database captures limited data about the journey of patients through the hospital system of FMC, from when they first arrive to their discharge. For example, it includes date and time stamps of arrival, movement between wards and discharge. It also contains limited clinical and demographic information. This allows for detailed analysis and modelling of these journeys.

2.2 The Model

A DES model was created to explore the dynamics of patient flow through FMC. The model captures the journey of patients through the hospital from the time they first present at the ED, or are admitted as elective surgery patients, to the time that they are discharged. It models all patients in the ED and all specialties, surgical, general medical and elective surgery patients. Despite the complexity of the patient journey, the DES has a modular structure that can be easily represented as a flow diagram. An overview of the layout of the DES is given in Figure 2.

The vast majority of patients admitted to FMC are those that enter via the ED. These patients are then treated in the ED and the decision is made to either admit them to the hospital as inpatients or to discharge them directly. Elective surgery patients are admitted immediately as inpatients, as their length of stay is known in advance of their admission to hospital. All inpatients are transferred from the ED towards within the hospital, which may or may not be their home ward - the ward where specialist care for their particular condition is available. Patients do not need to be in their home ward to receive treatment, but they have better access to necessary services if they do. Once treatment is...
completed, a patient is discharged from the system and removed from the simulation.

The DES model parameters was calibrated with both pre-existing patient flow data as well as input from senior hospital staff. These include arrival times, triage types, admission rates, treatment times and resource capacities, and are captured explicitly with distributions such as length of stay (LOS). It is important to note that some aspects, most notably waiting time before each patient movement, are emergent outcomes that are not explicitly defined within the model. In other words, the model can act as a thinking tool and bring issues to light that may not be directly obvious until the modelling is undertaken. The simulation is run to imitate the system, where average waiting times, occupancy rates and many other statistics are monitored and collected. Measurements should be calculated from multiple runs of the simulation to take into account the stochastic behaviour of the model. An average can be taken on these measurements to form a baseline measurement.

The development of the model was carried out using the Anylogic Software Suite. A desirable feature of Anylogic is its multi-method modelling platform and its ability to create and view models at multiple levels of abstraction (Anylogic, 2014). Other key features that made it appealing include its customisability through use of additional java scripting, ease for designing animations and optimisation and experimentation designs.

To facilitate understanding of the scope of the DES model, descriptions of its main components are provided as follows.

### 2.2.1 Processes

There are many processes which form a part of the DES model, the most significant of which is the process which governs the arrival of patients to the Hospital ED. Patients arrives at the ED at different rates from hour to hour and from day to day. The arrival process in the simulation itself is split into two sub-processes, arrival by

![Figure 3: Poisson distributions of arrival rate for Mondays 9am-10am, grouped by presentation type](image)
ambulance or by another means. The arrival rate is modelled by a Poisson distribution; with an exponentially distributed inter-arrival rate for each day, hour and mode of arrival combination. The parameters of these distributions were saved into a file which is loaded into the DES model on simulation start-up. As the arrival rate distribution is updated every hour during the simulation, there are 168 distributions for each mode of arrivals. Each of these distributions corresponds to one of the day-hour and arrival mode combinations in the calendar week. As previously mentioned, the distributions have similar characteristics, but differing parameters. An example of a pair of these distributions, for both arrival modes during the same hour, is given in Figure 3.

2.2.2 Decisions

The DES classifies patients into groups based on severity of condition, type of condition and when the patient was diagnosed. Decisions are the simplest way to replicate the different streams and types of patients commonly found in the hospital, providing the option to use probability-based or condition-based statements to determine a patient’s journey. As shown in Figure 2, the first such decision in a patient’s journey is whether they need to be admitted from the ED to the hospital proper. At FMC, approximately 30% of all ED presentations are admitted to the hospital as inpatients. Paediatric patients are excluded from those admitted, as they have a separate treatment path which is not currently modelled in the DES. Next, specialty and surgical patients are admitted directly to their respective wards, where they can receive the best possible care as soon as possible. In the context of the simulation specialty refers to a clear diagnosis of a patient’s condition within ED. All remaining admitted patients, i.e., those who are not specialty or surgical, are transferred from the ED into the Acute Medical Unit (AMU). The function of the AMU will be described in 2.2.4. From AMU, patients are then streamed to their respective wards if possible. The general medical stream is split into two further sub-streams: long stay and short stay. The long stay and short stay streams accommodate most of the patients that are admitted to AMU, as they are the destination for most general medical patients and are where most beds are allocated. The specialty and surgical streams are for those patients that were not judged to be clear-cut cases while in the ED, but were found to fit those categories after further testing in the AMU. The aged care, transfer and discharge to home streams all involve patients leaving the modelled system directly from the AMU, and thus are not modelled any further.

2.2.3 Services

The hospital model is made up of many services such as treatment times, recovery times or bed cleaning. These
services have an associated time delay, which slows the movement of patients or releasing of resources through the hospital. These delays are modelled by a combination of real data and estimated properties when recorded hospital data was not available. The long stay and short stay treatment times have been modelled using length of stay data profiles for each of these patient types. First, the probability density functions for each distribution are created from the data and secondly, variables are created to represent these distributions in the simulation. The distributions are defined discretely as shown in Figure 4. It can be seen that a large proportion of patients in both the short stay and long stay streams stay in the hospital less than a week. Despite this, each distribution has a long tail, indicating that some patients can have quite long stays in the hospital. The longest stay in the data sets that were modelled was 33 days for a short stay patient, and 236 days for a long stay patient. Due to lack of necessary data, some of the other delay distributions were estimated based on advice and discussion with senior FMC staff members.

2.2.4 Acute Medical Unit (AMU)

Another unique component of the FMC hospital system, and the modelled DES, is the Acute Medical Unit. The AMU is practically an extension of the ED where patients whose diagnoses are unclear can undertake further assessment, while being treated, before being moved to a more appropriate ward or being discharged. It was implemented as a mechanism to reduce inpatients’ ED waiting time as well as reduce crowding within the ED. All patients who stay in the AMU are classed as inpatients. Figure 5 gives an overview of how the AMU operates as a part of the hospital system.

Each of the patient streams involves patients leaving the AMU for another ward, except for short stay patients. Their stays are generally quite short - three quarters of short stays are two days or less - so they complete their treatment within the AMU itself. The average length of stay for patients in the AMU is approximately 21 hours, inclusive of short stay patients.

2.2.5 Elective Surgery

Another service modelled in the DES is the elective surgery stream of patients. These patients are different to the other streams as they do not enter the hospital through the ED and AMU, but directly into a surgical bed. Elective surgeries performed are contingent on the availability of beds for the patient. When an elective surgery procedure is cancelled, a backlog is created that must be cleared on subsequent days when beds are available. The beds for surgical inpatients are shared with elective surgery patients.

Elective surgery patients enter the hospital on a weekday morning only, as elective surgeries are not performed on weekends. An elective surgery has a randomly distributed duration between an estimated minimum and maximum, followed by a similarly distributed recovery period. The estimates are arbitrarily chosen, as there was no data available to us at this stage. If a patient has recovered and is ready to leave the hospital before wards are closed, then they are discharged from the hospital on the day of their admission. Some patients may undergo a more complex surgical procedure or take longer to recover from their surgery, hence unfit to be discharged by the closure of wards on the same day. These patients are kept in the hospital overnight for further recovery and observations, reflecting the actual operation of the hospital’s elective surgery services.

2.2.6 Queues

In a hospital system with limited resources, delays can be experienced while waiting for these resources to become available. This necessitates the use of queues in various parts of the system simulation. For example, if a surgical bed is unavailable for a surgical patient who is currently in the ED, then that patient will remain in an ED bed until a surgical bed becomes available.

Queues facilitate this by queueing up patients in the order that they reach the bottleneck in the system, and use a first-in first-out approach when allocating these patients to newly available beds.

The operation of queues can be manually overridden by other events in the simulation. This allows external entities, such as the bed manager, the ability to intervene the system in times of congestion by manually removing patients from these queues or discharging patients to free up the resources they hold.

2.2.7 Weekend Delay

Another queue-type feature incorporated in the simulation is the so-called weekend delay. Some services within the hospital are unavailable on weekends, prolonging patient stays and preventing them from being discharged. The purpose of this feature is to replicates policies within the hospital that lead to lower discharge rates on weekends. Lower discharge rates were set on weekends and the exit of a proportion of patients who have completed their treatment on weekends (Saturdays and Sunday) are delayed. The weekend delay is a small sub-process within the simulation itself. The outline of the decision making process is shown in Figure 6.

3. Model Validation

During the creation of the DES, the model must be continuously validated to ensure that it represents the FMC patient flow. For example, ED arrivals data was used to validate the arrival process, inpatient data was used to validate the admissions numbers and so on. This validation indicated that the model is quite close to the hospital data in many major metrics, including patient throughputs. However, due to unavailability of data, some parts of the model cannot be tested.

Midnight occupancy is one of the most important measures of capacity for the FMC. The DES can easily be configured to collect midnight census data on each run of the simulation. The collection of midnight census data is controlled by a discrete event that is triggered at midnight of every day during the simulation run. Each time the event is triggered the number of general medical and surgical patients currently admitted are calculated. This number, and the coded day of the week, are saved into an excel spreadsheet which can be accessed after the simulation run has been completed. To account for the
inherent stochasticity in the model, the average occupancy over ten runs was taken as the baseline measurement. Each simulation had been run for three and a half years, the same time period covered by the original data set. Though there are varying oscillations in each of the ten runs, the majority of simulated occupancy values are bounded between two constant upper and lower

Figure 7: Time series plot of the midnight occupancy of ten runs of the DES, with the average of the ten runs overlaid

Figure 8: Time series plot of the DES ten-run average midnight occupancy and the original midnight occupancy data set
bounds through the time series. This effect can be easily seen in Figure 7 on the following page.

A comparison of one of these ten simulation runs to the original data set is shown in Figure 8. It can be seen that the observed data and DES output are well matched for the first part of the time series. There is a portion in the middle where there is a noticeable difference between the two series. The series return to similar levels at approximately the beginning of 2012, after which they remain in close proximity until the end of the time series. Most of the data that controls the operation of the DES is based on the 2012 calendar year. This justifies the differences between the model and observed data during some of 2010 and 2011, as there may have been different throughputs and length of stay distributions for those time periods, possibly due to short term management decisions on adjustment of operations. Despite the fact the simulation sits a little above the measured population, the model is still a quite accurate representation of the hospital, especially for the purpose of showing trends and testing ideas and concepts.

4. Application to Real World Problems

4.1 Analysis of Change Impact

The model has been used to explore the likely impact of additional patient arrivals at the ED. This could be a consequence of reduced use of general practitioner service provision due to the introduction of a co-payment policy. In order to construct this analysis it was necessary to make a number of assumptions about how patients may respond to the change in policy. Thus, it was assumed that the patients affected by the change in policy would substitute general practitioner services with ED services, and that this would occur during the times that general practitioner clinics are generally open (i.e., Monday to Friday between 8.00am and 6.00pm). The number of additional patients expected to arrive at the ED was between 1 and 4 per hour and that this additional demand would be sustained. This seemed a reasonable assumption, as the hospital’s catchment is large and there are, on average, 20,000 regular and long GP consultations every day in South Australia. One additional patient per hour represents just 0.036% of the total GP consultations, was aimed at reducing the cost of health care, it was assumed that the current level of resources in the ED would be maintained and not increased. The results are shown in Table 1.

As would be expected, the change in demand resulted in a non-linear change in queue length and the time spent in the ED as shown in Table 2.

Table 2: Change in performance measures arising from additional patient arrivals during Monday to Friday between the hours of 8:00 am and 6:00 pm

<table>
<thead>
<tr>
<th>Performance Measures</th>
<th>Number of Patients Per Hour</th>
<th>Additional Patients Per Hour</th>
</tr>
</thead>
<tbody>
<tr>
<td>average number of patients in the queue (hours)</td>
<td>1.2</td>
<td>8.0</td>
</tr>
<tr>
<td>increase in the average time in ED (hours) overall</td>
<td>0.07</td>
<td>0.79</td>
</tr>
<tr>
<td>if discharged from the ED</td>
<td>0.10</td>
<td>0.85</td>
</tr>
<tr>
<td>if admitted to a hospital bed</td>
<td>0.11</td>
<td>0.89</td>
</tr>
</tbody>
</table>

Table 1: ED performance measures and the effect of increasing patient arrival during Monday to Friday between the hours of 8:00 am and 6:00 pm

while four additional patients per hour represent 0.143% of consultations. Furthermore, as the government policy

4.2 Future Model Use

It is currently planned that the model will be used to investigate, among other things:

- Waiting time and Bed occupancy implications arising from changed bed use;
- Waiting time and Bed occupancy implications arising from altering policies and work practices; and
- Implication of changes to the age mix of patients.

5. Challenges of Modelling Patient Flow

Many challenges exist in using modelling to study health system, as discussed previously by others (e.g., Fone et al., 2003). However, due to the scope of this paper, we focus our discussion to the following subset of challenge identified across the course of our work. These include:

- Data structure – the data used for modelling was not designed with the development of a DES model in mind
- This has made it difficult to accurately extract and capture all the desirable aspects of the hospitals behaviour, particularly in relation to when decision were made and subsequently acted upon, and when resources were assigned to the patient.
- Data cleanliness – there had also existed issues with using the hospital’s historic data. Numerous inconsistencies in the recording practice for the data and erroneous entries had made accurate extraction difficult. It is difficult to pinpoint a reason for these inconsistencies, as they could be symptoms of natural human error, poor uptake of recording policy in the hospital, or even a side effect of the system being under stress. This issue of data-cleanliness is further compounded by changes in the quality and type of data being recorded, as priorities within the hospital often change.
• The double edge nature of multi-disciplinary teams whose members do not share the same professional language (technical jargon) and culture but possess complementary knowledge and skills.
• Wariness of some groups as to where and how a simulation model may be used, hence hesitation in up taking modelling practice.

The last two challenges represent an interesting but important issue in regards to managing collaboration and communication during and after the modelling process. A more in-depth discussion on this issue was presented in a separate paper (Mackay et al. 2013). However it is worthwhile noting the value of an animated graphical illustration of the process modelled in addressing this issue. In particular the authors have experienced larger level of buy-in both with collaborators, peers and other non-affiliated health professionals when a visualisation has been presented.

Further to the above four points, while the DES patient flow model may be useful in describing the current system or testing proposed changes to the system, such as wait times and occupancy levels, it must be recognised that we are modelling only a small aspect of the patient journey, both within the hospital and beyond the hospital. As hospitals are merely a cog in the health care system, albeit a significant one in terms of resource cost, key changes that may affect patient arrivals cannot be modelled easily. For example, it is understood that primary health care service provision, and access to nursing homes and home care services for patients being discharged are important factors affecting hospital demand and occupancy levels (Patrick 2011). The fact that different jurisdictions are responsible for the collection of data across different services (e.g., general practice data) or that there are many providers (aged care) that patients may use means that data is not necessarily available for the inclusion in the model or is more difficult to access. Consequently, while the model can be used to aid some decision-making, it is evident that it is not suited to facilitating all decisions that might affect patient arrivals, wait times or occupancy levels, especially when health care management issues beyond the hospital are involved. Agent-Based multi-method modelling has been advocated as an emerging method with great potential for hospital simulation modelling (Gunal 2013). Fortunately AnyLogic facilitates multi-method modelling (Borschchev and Filippov 2004), and the DES approach could be integrated with system dynamics and agent based approaches to develop more accurate and useful models.

6. Conclusion

Modelling hospital patient flow process is highly challenging due to the inherent complexity and uncertainty in each step of the process. It has been shown that discrete event based simulation models of patient flow can help to identify and demonstrate problems in hospital operations and subsequently to provide “what-if” analysis and identify solutions for improvement. Although the model is developed for FMC, the methodology outlined here is generic enough to be adapted for other hospitals. Advanced software based simulation modelling has the advantage of shorted cycle time for model development, as well as appealing animation of the process dynamics for communication of model results. Close collaboration with hospital staff is vital to the correctness and usefulness of a model, which ultimately will make the difference of whether it can be put to good use in hospital operation improvement and patient flow optimization.

6.1 Acknowledgement

This research was financially supported by Flinders University and Flinders Medical Centre under the Flinders Collaborative Research Grant Scheme and ARC linkage project LP130100323 Congestion Recovery and optimisation of patient flows.

7 References


