Online Simulation for the Operational Management of Inpatient Beds

David Oakley

Thesis submitted for the degree of Doctor of Philosophy

Lancaster University Management School

Department of Management Science

United Kingdom

2019

Thanks to Dave and Stephan for their unwavering support, enthusiasm, and most of all, patience. I couldn't have asked for better supervisors.

Contents

List of Figures

List of Tables

Abstract

In many modern hospitals, resources such as beds, theatre time, medical equipment and staff are shared between patients who require immediate care and must be dealt with as they arrive (emergency patients), and those whose care requirements are known to the hospital some time in advance (elective patients). Caring for these two types of patients poses a logistical challenge, since some portion of each resource must be set aside for *emergency* patients when planning for the number and type of *elective* patients to admit. Failing to strike this balance can result in negative outcomes, such as patient-stays on non-ideal wards, or increased waiting time for elective procedures (in the case of public health services).

The potential benefits of using discrete event simulation (DES) models in healthcare are well established, and they are often preferred to other modelling approaches because of their ability to emulate the randomness seen in real systems, at a level of detail which is necessary for models to be convincing. However, their use is often limited to strategic or tactical decision making, and few have attempted to produce models which can help hospitals with short-term (operational) decision making. This is where Online Discrete Event Simulation (ODES) can help.

An ODES (also known as symbiotic simulation) takes all the components of a DES model, and adds the ability to load the state of the real system at run-time to make predictions about how the real system might evolve in the short-term.

This thesis reports the development of a whole-hospital, proof-of-concept ODES to assess the impact of elective admissions decisions, on wards which are shared with emergency patients. The model is parameterised by analysing 18 months of patient administrative data from an Australian General Hospital.

Since ODES is a relatively new method, this research focuses on formalising the model development process, resulting in a new “black-box” validation method for handling conditionally distributed simulation outputs. Additionally, a new probabilistic routing method is developed to better represent inter-ward dependencies during peaks in bed demand. A statistical analysis of the relationship between ward transfers and ward occupancy is conducted on real hospital data to parameterise so-called “Dynamic Transition Matrices” for this purpose. Finally, the ODES is used to demonstrate how additional patient-level information (which might only become available after admission) can affect the predicted bed census. Clinicians’ discharge date estimates fit this criterion, and the case is made for more scientific use of this type of information, as part of an operational ODES model.

Chapter 1
Introduction

1.1 Overview

In many modern hospitals, resources such as beds, theatre time, medical equipment and staff are shared between patients who require immediate care and must be dealt with as they arrive (emergency patients), and those whose care requirements are partly known to the hospital some time in advance (elective patients). Caring for these two types of patients poses a logistical challenge in the sense that some portion of each resource must be set aside for *emergency* patients when planning for the number and type of *elective* patients to admit. Hospitals have guidelines for the number of emergency patients they might expect to see in each planning period, although the exact number is unknown. If too many elective patients are admitted, the hospital's ability to treat emergency patients will be reduced, potentially resulting in negative patient outcomes, such as "outliers" - a term which refers to patients whose ward might not be ideally suited to their condition. On the other hand, if too few elective patients are admitted, patients can be left on waiting lists unnecessarily in the case of public health services, or represent a loss of income in the case of private health services.

In terms of bed resources, research has been conducted to better understand the relationship between the number of occupied beds and negative patient outcomes. Bagust et al. (1999) developed a computer-based model of a hypothetical UK hospital, finding that when the number of occupied beds averages 85% of the model's maximum bed capacity, there is a noticeable increase in the risk of a "crisis day", where at least one patient requiring immediate admission cannot be accommodated. In addition, their model suggests that crisis days will occur regularly if average bed occupancy surpasses 90%.

Although the relationship between bed occupancy and negative patient outcomes is expected to vary between hospitals, NHS England's bed availability statistics (NHS England, 1988-2017) show an increase in the proportion of NHS Organisations exceeding the 90% threshold for general and acute sectors. Between the years 2000 and 2010[1], the proportion of hospitals whose average occupancy was above 90%, increased from 20% to 31%. The year ending March 20161 saw a further increase to 44%. Despite rising occupancy, governments are also asking health services to make better use of their funding, as part of austerity measures which aim to reduce budget deficits. To this end, the NHS reportedly made £2.9 billion worth of so-called "efficiency savings" (NHS Improvement, 2016a) between April 2015 and March 2016.

With increasing demand for service, and without commensurate funding increases to public funding, hospitals require (preferably low cost) methods for maintaining high patient throughput whilst minimising negative outcomes, such

---

[1] Annual publication at the organisation-level ceased after 2009, in favour of quarterly publication. Annual statistics after 2009 are derived by the author as the average of each year's four constituent quarters, for continuity.

as outliers. However, changes to the management, or possibly even the capacity of a given hospital resource require some assessment of the expected improvement. However, testing competing management options in an operational hospital may not be possible for several reasons; including potential service disruptions, and most importantly, maintaining patient safety on wards which operate with low margins for error, such as Intensive Care.

The field of Management Science is concerned with answering questions about how organisations are affected by management decisions, by using analytical methods. One method, is to develop a model of the real organisation (or system) which can be used to test the impact of competing options. Generally, models of this type are mathematical or computer-based in nature, although the chosen approach is influenced by the type of question the researcher or manager wishes to answer, and the details of the system being modelled.

Although many researchers have produced models which focus on specific hospital facilities (e.g. Accident and Emergency), or specialisms (e.g. Cardiology), few have attempted to build models which encompass broad patient types from admission to discharge. Fewer still have attempted to produce models which can help hospitals with short-term decision making. Most of the hospital modelling literature is concerned with longer-term decisions, such as determining the required capacity of each resource, or the creation of cyclic elective admissions schedules which optimise resource use *on average*. While this information is likely to be essential for the long-term “health” of a hospital system, it does little to help a hospital which is already experiencing abnormally high demand.

The aim of this research is to develop a proof-of-concept model for operational (short-term) decision-making, to help hospital staff maintain an effective balance between two broad patient types; elective (planned) and emergency (unplanned) inpatients. Decision-making in the short-term benefits from the latest information, therefore this research favours a data-driven approach which allows the model to be re-calibrated when new data becomes available. An added benefit of this approach is that the same type of model could be developed for any hospital which collects similar data about its patients' stay in hospital.

1.2 Discrete Event Simulation

Although several Operations Research/Management Science (OR/MS) techniques have been applied to hospital modelling problems, including Queueing Theory, Mathematical Programming (Optimisation/Heuristics) and System Dynamics, Discrete Event Simulation (DES) offers several advantages, making it particularly well-suited to the aims of this research.

One of the most important advantages of using DES, is that it can emulate the randomness seen in the real system (such as the arrival of emergency patients) which allows the distribution of any performance indicator to be approximated. Approximating the distribution of a given performance indicator is likely to be useful in short-term planning scenarios where the user is more interested in the probability of an event occurring, rather than estimating average performance.

DES also offers the greatest degree of flexibility in defining the model's structure and parameterisation. Specifically, DES permits the use of both empirical and closed-form functions for defining arrival and service time distributions, whereas queueing-based analytical models often require closed-form distributions, or specific classes thereof, to produce tractable results. The flexibility to use empirical distributions also aligns closely with the data-driven nature of the modelling approach, since the observed data can be used to construct the necessary empirical distributions, without the need for additional curve-fitting software.

Finally, DES can operate at a level-of-detail which allows information about individual simulation entities to be included in the model. This is important because the task of balancing emergency and elective numbers in the short-term is likely to depend on patient-level information, such as the day and ward of admission, and expected length-of-stay. Methods such as System Dynamics, which would deal with patient movement between wards as “flows” and “stocks” at an aggregate level, are therefore unsuitable.

1.3 Online Discrete Event Simulation

An Online Discrete Event Simulation (ODES) takes all the components of a DES model, and adds the ability to load the state of the real system at simulation run-time. In contrast, non-terminating DES models typically start “empty and idle” and run through a warm-up period before data collection begins. The data generated during the warm-up period is often disregarded, since it does not represent the normal operating conditions of the system being modelled, however in an ODES model, an operational state is loaded at initialisation, therefore data collection begins immediately.

However, the motivation to create an ODES model is not simply to circumvent the warm-up period. Rather, the data generated during the warm-up period becomes the modeller's primary focus, since it represents a set of possible "futures" from the initialised state. In this sense, ODES models bear some similarity to terminating simulations, which retain data generated during the warm-up period and run until a set of termination criteria (such as the completion of n jobs, or the end of a working day) are met. However, ODES is designed to model non-terminating systems, for a sequence of transient periods which are initialised with the real system state whenever a run is requested. As might be expected, increasing the run-length of an ODES reduces the dependence of the simulation outputs on the initial state, so that eventually the results might not be statistically distinguishable from that of a non-terminating simulation. Therefore, the value of an ODES model is in its ability to test how a system might evolve over relatively short periods of time, based on a set of recently observed initial conditions.

1.4 Expected Contributions

ODES has been most widely applied within the manufacturing sector, where sensors connected to machines are able to track the location of jobs and their progress. A tracking system allows an ODES model to be loaded with the state of the system being modelled at any time, however, even if it was feasible, the real-time tracking of staff and patient locations in a hospital is likely to be considered too intrusive. Nevertheless, patient location and progress information *is* collected on a "semi-real-time" basis, given that manual data entry might be required before the information becomes widely available. Therefore, an overarching contribution of this research is to demonstrate that with reasonable assumptions about the frequency and the type of data collected, an online simulation for the operational management of inpatient beds can help hospitals to make better decisions, without the need for additional data acquisition systems.

Theoretical contributions to the ODES methodology are also possible, since methods for validating models prior to their operational use have not been sufficiently covered by the ODES literature. Specifically, the simulation outputs of an ODES are known to be conditionally distributed (conditional on the model's initial state, for example) due to the model's short run-length, and therefore require a different treatment to that of a classical DES model. For this reason, this research is also concerned with investigating how an ODES can be validated, whilst respecting the conditional nature of the simulation outputs.

Finally, it is well-known that the occurrence of outlier patients suits neither the patient, nor the alternative ward on which they are placed. Although various models in the literature have reportedly incorporated outlier placements, none (to the author's knowledge) have done so by utilising a hospital's patient database to inform the way outliers occur during peak bed demand. A model where patient placement dynamically responds to the number of occupied beds is likely to provide a better representation of the impact of short-term decisions across the ward network.

Methods for validating online simulations via their conditionally distributed outputs, and methods for dynamically routing simulation entities through a network, are also likely to have applications beyond hospital modelling. The former is likely to be broadly applicable, since checking a model's performance is an important step in its development, regardless of what the model might be used for. The latter could be used to model systems where customers can select alternative servers (or leave the system altogether) when their current or preferred server becomes too busy; a behaviour referred to as "jockeying" in the queueing theory literature. While methods exist for modelling systems which allow jockeying to occur, the ability to detect behaviour such as this (from real data) means no assumptions need to be made about the thresholds at which customers will switch servers, or the complexity of the system being modelled.

1.5 Thesis Structure

The remaining chapters of this thesis are organised in the following way:

Chapter 2 provides an overview of the hospital modelling research which has been conducted within the field of Operations Research/Management Science (OR/MS), and discusses the types of control strategies hospitals have at their disposal for improving service delivery. The notion of "operational" decision-making in hospitals is further defined, and some of the medium to short-term planning models reported in the literature are examined. The chapter concludes that using an ODES model offers several advantages over other modelling methods, and the research questions for the thesis are established.

Chapter 3 describes the backdrop against which this research takes place, along with the involvement of the Australian General Hospital (AGH) whose patient database forms the basis of the modelling work. Although this database contains information generated by a single hospital and its patients, similar information is known to be routinely collected by most modern hospitals. An overview of the database extract is provided, along with an explanation of how it begins to influence the model's scope. Any assumptions made, or cleaning steps taken in the process of readying the data, are also discussed.

Chapter 4 describes the development of the ODES model in detail, from conceptual design through to testing. Of particular importance, are the steps taken to move from a classical (offline) DES model, to an ODES model, and their respective validation methods. A newly developed method for validating ODES models is described and demonstrated, which aims to account for the conditional distribution of the simulation outputs.

Development of the ODES continues in Chapter 5, with a focus on emulating the placement of outlier patients to better reflect the routing behaviour between wards during periods of high demand. To this end, the relationship between occupied bed numbers and ward placement is investigated statistically, using data supplied by the AGH. It is demonstrated that the results of such an analysis can be implemented in an ODES, and the effect of doing so is compared against a model where this information is not included.

Chapter 6 demonstrates how the ODES could be used in practice, using two case studies. In the first case study, the model is used as an early-detection system to anticipate days when an effective balance between emergency and elective workloads is *not* likely to be struck, given the elective arrivals which are scheduled over the period in question. In the second case study, the impact of obtaining additional information, outside of what is collected in the patient administration database, is investigated. This information could include Estimated Date of Discharge (EDD), which is frequently assessed by doctors and nurses, although may not be formally entered in any patient database. The value of including this type of information is assessed at varying levels of accuracy, given its potentially subjective nature.

Chapter 7 closes the thesis with a summary of its findings, a review of its research contributions, and a discussion of further work which could be carried out in this area.

Chapter 2
Literature Review

2.1 Introduction

The purpose of this chapter is to gain an understanding of the bed management literature within the field of Operations Research/Management Science, and to position this research within the field. The review starts with a discussion of the control strategies and the associated “planning levels” which can be employed by hospitals. The planning levels provide a means to classify the bed management literature and help to distinguish operational planning from other planning types. The inpatient bed management literature is then reviewed, with a focus on models with similar aims to this research, namely, bed management via admissions control. Finally, the ODES literature (not necessarily pertaining to hospitals) is reviewed, to gain an understanding of the theoretical work which has already been carried out, and some of the practical considerations which come with developing a model of this type. The chapter concludes with a summary of the review findings, and a statement of the research questions.

2.2 Control Strategies and Planning

Broadly speaking, there are two ways of managing resources for improving hospital performance. Either the *supply* of resources is controlled and matched to the stream of admitted patients, or the *demand* for service is controlled and matched to available resources. Gemmel and Van Dierdonck (1999) refer to these strategies as “Chase Strategies” (supply control) and “Level Strategies” (demand control) respectively.

In practice, the choice to exercise either of these strategies is dependent on the mix of elective and emergency patients, and the scope of the changes the hospital is willing to commit to. For example, an Emergency Department (ED) would have little use for demand control, since the arrival of emergency patients is generally unscheduled and to a large extent, uncontrollable. Therefore, if improvements are made in an ED, the relevant decision variables are likely to be related to managing the *supply* of resources; such as bed capacity, staff numbers and staff schedules.

Table 2.1, taken from Groot (1993), provides an overview of the types of decisions made within a hospital, and the typical planning horizons required to implement them. The two highest tiers are described as the “strategic” planning level, in which decisions can require years of planning to execute. This level would include the construction or expansion of facilities, or changes to the types of specialist care which are made available to patients. Next, the capacity allocation and capacity scheduling levels form the “tactical” planning level, in which short-to-medium term decisions are made. Such decisions might include the reallocation of beds between specialties, or the number of theatre hours which are available per day/week. Finally, the scheduling of elective patient arrivals sits within the “operational” planning level. Although it is possible for elective patients to remain on waiting lists for extended periods of time (in the case of public health services), the detailed scheduling of procedures takes place only once the patient nears the front of the waiting list queue. Scheduling the admission might take place weeks before the patient is due to arrive, however decisions to *postpone* elective treatments can occur with very short notice. Hans et al. (2012) note that reacting to unforeseen events, such as bed shortages, is a typical feature of planning at the operational level.

Questions	Decision Makers	Level	Horizon
What is the future direction of the hospital?	Board of Directors	Strategic Planning	2 – 5 Years
What will the hospital activities be in the coming period?	Top management	Main patient flow planning	1 – 2 Years
How are the capacities allocated to functions or departments?	Top and middle management	Capacity allocation	Months – 1 Year
How are the capacities scheduled in time?	Middle management	Capacity scheduling	Weeks - Months
Which patient is treated at what time?	Admission planner	Operational Planning	Days - Weeks

Table 2.1: Production control decisions in a hospital (Groot, 1993, p.16).

In addition to defining the terms “strategic”, “tactical” and “operational”, Table 2.1 also suggests that demand control is most relevant at the operational planning level, since the only example in the table is reserved for the operational planning tier. While *supply* strategies certainly exist within operational planning (such as the management of nurse rotas), decisions of this type are likely to be impacted by the number of patients admitted in the planning horizon. Therefore, this research focuses on demand control strategies; specifically, the admission of elective inpatients, in an operational planning context.

2.3 Bed Management Literature

Given the importance of achieving reasonable levels of efficiency in hospitals, bed management has been an active topic of research in OR/MS for a long time, resulting in numerous approaches to the problem and vast quantities of related literature. Some of the earliest literature surveys include Milsum et al. (1973) whose surveyed papers investigate the relationship between admissions scheduling policies and hospital resources; Magerlein and Martin (1978) who focus on the surgical scheduling literature, including “multiple constraint” models which account for bed numbers and nursing staff; and England and Roberts (1978) whose survey covers the use of computer simulation across various healthcare systems, including admissions control and bed management.

With the proliferation of more powerful personal computers and programming languages, simulation has become one of the most popular tools to aid healthcare decision making. So pervasive are simulation studies, that numerous literature surveys have tracked the progress of this modelling method almost exclusively, including Klein et al. (1993), June et al. (1999), Fone et al. (2003), Fletcher and Worthington (2009) and Günal and Pidd (2010). The prevalence of simulation is due (in part) to its flexibility, which facilitates the modelling of complex systems, such as hospitals. However, a significant body of work within healthcare OR/MS exists outside of simulation, highlighted by literature surveys by Rais and Viana (2011) who primarily focus on optimisation problems, and Lakshmi and Iyer (2013) who focus on applications of queueing theory.

To help narrow the body of research, the bed management literature is reviewed by first focusing on the class of methods and models in which elective hospital admissions are treated as a decision variable. Not all the literature in this class is intended to be used at the operational planning level, however, an understanding of how beds might be managed, via admissions control, remains important. The review then focuses on the bed management models which are designed to inform operational decision making, before examining the small number of reported ODES models used in a healthcare setting.

2.3.1 Bed Management via Admissions Control

While simulation has been broadly applied in healthcare modelling, few simulation models regard the elective admissions schedule as the primary decision variable. Since the admission of elective patients can be viewed as a scheduling problem, the literature in this domain tends to favour analytical methods which aim to provide optimal (or close to optimal) schedules given a set of constraints. However, simulation studies do exist in which the relationship between elective admissions and bed is investigated; sometimes by comparing competing admissions *policies*, and sometimes by drawing from waiting lists which can be user-specified or randomly generated.

Early work in this area includes Smith and Solomon Jr (1966), who developed a simulation model of a hospital treating narcotics addition. While the type of patient differs significantly from the acute patients this research is concerned with, the model's structure bears similarity to an acute hospital through the admission of non-authorized (unplanned) and authorized (planned) patients. The authors also note that the authorized patient stream is the most easily controlled, and therefore the decision variables for the model are based around their admission. Three types of admissions policies are analysed; Type A allows a fixed number of admissions per day. Type B allows the number of admissions to vary proportionally to the number of discharges. Type C allows the number of admissions to vary with the number of discharges plus or minus a constant. The aim of the work is to minimise variation in the bed census while maintaining reasonable occupancy levels, which was achieved by using a constant admissions rate (Type A). With no details of a ward network or disaggregation of length-of-stay, the success of a simple admissions policy seems reasonable, although for a more complex system, some disaggregation of the overall admissions rate (into wards or patient types) is probably required.

Other early work includes Robinson et al. (1968), who use a simulation model to test three routines for the development of an automated scheduler for elective admissions, with a particular focus on how estimating each patient's length-of-stay might improve scheduling decisions. Four scheduling methods are tested. The first method is myopic, and schedules patients while giving no consideration to expected length of stay. The second schedules a patient only if a pre-determined occupancy level is not exceeded during the patient's expected time in hospital. The third determines a conditional probability of occupying a bed. For each day the patient remains after the expected discharge date, the hospital census is updated using the probability of remaining in the hospital, given LOS so far. Finally, the fourth method allows the initial LOS estimates to be revised every three days, and admissions are scheduled based on updates to the estimated hospital census. The results are evaluated using a cost function which penalises empty beds, overflows and turn-aways. The results show that admissions schedules derived from an estimated census in which revisions to LOS estimates can be made, minimises the cost function. However, only elective patients are considered, and the bed pool is treated as homogenous, which could have a significant impact on the ability to admit on the scheduled day, when applied to multi-ward hospital.

Although the work described by Bagust et al. (1999) models only emergency admissions, they claim that it can be generalised to accommodate both emergency and elective streams. The work is interesting in that negative patient outcomes (crisis days and proportion of patients not admitted) are treated as a function of bed occupancy for a hypothetical acute English hospital. Several different experiments are run. Scenarios include changes to the rate of patient arrival, changes to the number of available beds and changes to the discharge rate among others. One of the most cited conclusions of the paper is that hospitals operating at 90% occupancy or higher will suffer regular crisis days, and that operating staffed but empty beds is a necessity for absorbing stochastic variation associated with emergency arrivals. However, no mention of how to maintain such an occupancy level is offered, and in practice this may be a challenge without a method to help balance emergency and elective workloads.

The simulation model described by Everett (2002) considers both bed and operating theatre resources. A detailed waiting list is continually updated by draws from Poisson distributions governing each patient type and urgency, with rate parameters estimated from historical data. Alternatively, real waiting list data can be read in from a file, and used in this way, the model can reportedly be used to support real-time decision-making. However, details of how the model might be initialised to facilitate real-time decision support are not provided. The surgery hours and bed days required for each patient are drawn from a Normal distribution, and patients are selected at the beginning of each day so that the sum of their expected theatre times does not exceed the total time available for that day. An index which represents both the urgency of the procedure, and the time spent on the waiting list so far also influences patient selection. Although this model contains many of the components deemed to be important for this research, its scope is high-level; including multiple hospitals which draw from a centralised waiting list. At this level-of-detail, beds are treated as a homogenous resource, therefore dependencies between wards within a single hospital cannot be modelled.

While details of a simulation model are also provided, Harper (2002) focuses on the development of a generic *framework* for modelling hospital resources, and outlines a number of modelling considerations for OR/MS practitioners working in this domain. The cornerstone of the framework is the Classification and Regression Tree (CART) analysis to construct homogeneous patient groupings, for which modelling parameters, such as arrival rates and lengths-of-stay, can be estimated. A model which incorporates the prescribed framework is developed and used to assess a set of competing theatre scheduling policies, and their downstream effect on bed occupancy. The model is also used to estimate the mean number of occupied beds per month using a stochastic representation of hospital processes and shows that this can differ significantly from estimation methods which only make occupancy estimates based on averages. While the simulation only considers surgical beds, similar software also appears to have been used in the simulation described by Harper and Shahani (2002), in which a multi-ward hospital is modelled. However, this model appears to treat the elective admissions as a stream of exogenous demand, rather than a decision variable.

Helm et al. (2009) also report the development of a comprehensive hospital simulation framework. One of the most interesting features of the framework is the existence of a feedback loop between the state of the hospital and admissions decisions; allowing the admissions policy to dynamically respond to the state of the simulated hospital. While other models can do this to a degree, via patient deferrals when no beds are available, the feedback within the framework allows the testing of a policy in which additional patients are admitted from short-notice waiting lists during times of low bed occupancy; improving efficiency by making better use of unoccupied beds.

Günal (2008) reports the development of a whole-hospital simulation, designed at a level of genericity such that it could be parameterised and applied to most modern hospitals. The whole-hospital model (DGHPSim) contains four component models which simulate accident and emergency facilities (AE~Sim), bed management (BM~Sim), waiting lists (WL~Sim) and outpatient facilities (OP~Sim). With the ability to load a user-defined waiting list, the admission of elective patients can be treated as a decision variable. BM~Sim receives elective admissions from the waiting list component and emergency admissions from the accident and emergency component, which are used to generate output statistics which include time spent on waiting lists, elective cancellations and the number of patients which become outliers.

Since scheduling problems are typically approached by using analytical methods, it is common that the literature investigates the relationship between elective admissions scheduling and bed management in this way. While simulation offers the most flexibility in defining the model's structure, analytical/mathematical methods have the potential to generate "optimal" schedules or provide exact solutions to well-specified problems. However, defining optimality necessarily depends on the problem formulation, and settling on a definition can be particularly difficult for systems with multiple objectives, like hospitals.

The analytical model presented by Gallivan and Utley (2005) covers many of the modelling elements considered to be important for this research. Elective patient admissions are treated as a decision variable, while the emergencies arrive at random according to a Poisson distribution, in which the rate parameter can vary by day-of-the-week, if necessary. Additionally, no class of length-of-stay distribution is assumed (as is often the case with queueing models), and instead so-called "length-of-stay persistence" distributions are used, which are essentially discrete survival distributions. Mathematical expressions are derived for the mean and variance of the number of patients occupying a bed for each day of the planning cycle, depending on whether the real hospital treats emergency patients. An optimisation problem is formulated from these results, which is used to find a cyclic admissions pattern which maximises minimum reserve capacity, thus smoothing the bed census for the planning cycle. While the results could be used to improve the use of a single bed pool, it is not clear how the model could be extended to a multi-ward hospital, in which there is expected to be interactions between wards.

Work by Adan and Vissers (2002) considers three major inpatient resources (beds, operating theatres and nurses) as constraints in an integer linear programming formulation. The objective of the model is to meet target throughput whilst minimising the difference between the actual utilisation and target utilisation of these resources. Instead of classifying patients by their condition, patients are classified by the quantity and type of resources they require. The formulation is solved for an orthopaedics unit, resulting in a one-week “admissions profile”, which sets out the number and type of patients to admit on each day of the week. However, the authors concede that the main weakness of the model is that it does not make allowances for emergency patients. Adan et al. (2009) extend the model described in Adan and Vissers (2002) by treating length-of-stay as stochastic, rather than deterministic, which results in smaller deviations from the target utilisations of each resource. Again, this model does not account for the workload generated by emergency patients, although the target utilisations are reportedly calibrated such that sufficient space is allowed.

Helm and Van Oyen (2014) develop one of the most comprehensive analytical models in terms of elements considered to be important for this research, and is one of the few which captures the entire emergency/elective pathway, whilst also modelling the bed census at the ward level. The authors start by developing a census model to capture the number of patients present in each ward for emergency and elective patients separately. Capacity constraints are then “overlaid” on these models to mimic the occurrence of outliers. The probability of a given ward containing a certain number of emergency patients is derived from the emergency census model and these probabilities feed into a mixed integer program which maximises the number of elective patients which can be admitted, subject to constraints on the number of outliers permissible. An alternative formulation is also presented whose objective is to minimise the number of outliers.

The analytical methods summarised so far have treated the elective admissions schedule as a decision variable; working to develop cyclic schedules for the number and type of elective admissions per day, which “optimise” the downstream use of hospital beds in some way. However, the process of scheduling elective admissions is closely related to the task of scheduling surgeries. Working from a surgical perspective, research has been carried out which pays more attention to the details of scheduling theatre time, whilst also keeping post-operative bed management in view. For surgical scheduling, an increase in temporal detail is most commonly achieved by considering operating room (OR) blocks, which represent subdivisions of a day in the planning horizon. OR blocks can be assigned properties of the modeller’s interest, such as block duration, length-of-stay and recovery ward, and these properties are used to determine feasible assignments of OR blocks to days in the planning horizon.

Models of this type include Beliën and Demeulemeester (2007) who use discrete empirical distributions to govern the number of patients, and length-of-stay distributions per-block for each surgeon. Blocks are assigned to each day in the planning horizon under different solution methods, including mixed-integer programming and simulated annealing, with the objective of minimising the total expected bed shortages. However, the models proposed do not account for stochastic variation introduced by the inclusion of an emergency patient stream, and beds are treated as a homogeneous resource. The model formulation is extended by Beliën et al. (2008) by grouping beds into wards, and by allowing block duration to vary by surgeon.

Van Essen et al. (2012) consider a very similar problem statement to Beliën and Demeulemeester (2007), but formulate the objective function in terms of minimising the number of beds required. Additionally, “what-if?” scenarios are run, in which various constraints are relaxed to assess the impact these have on bed requirements. Again, emergency patients are not considered, although they claim that the model can include them by estimating the average number of emergency surgeries and representing these as “dummy OR blocks.” However, doing so might over-simplify the situation, given the stochastic nature of emergency arrivals.

The model reported by Chow et al. (2011) consists of two main components. The first is an optimisation component; containing a mixed integer program which finds an elective admissions schedule consisting of OR blocks which minimises total maximum bed occupancy across all wards. The second component is an uncapacitated discrete event simulation which includes recovery, intensive care, day case, short stay and normal inpatient wards. The optimisation component generates elective schedules which are run through the simulation component. The simulation component generates emergency arrivals, whose admissions pattern can be arbitrarily complex. The two patient types are combined, resulting in overall bed occupancy statistics for each modelled ward. The elective schedule can be re-optimised under different bed constraints, should the combination of the two work-streams exceed available capacity, thereby leveraging the strengths of simulation and optimisation.

The research summarised in this section addresses the relationship between elective admissions and bed requirements by employing both analytical and simulation approaches. However, the insights gained as part of testing admissions policies, or by generating optimised admission/surgical schedules, are designed to be using on an ongoing basis. Assuming the modelling assumptions are not too abstracted from the real process, the insights gained from these models have the potential to improve *average* performance with respect to a chosen metric, and therefore serve an important purpose at the tactical planning level. However, most hospitals will experience abnormally high demand at some point, if only due to normal stochastic variation in the emergency arrivals process. In these situations, operational bed management techniques come into focus.

2.3.2 Operational Bed Management

Without employing Online Discrete Event Simulation, methods have been reported in the literature which aim to help hospital planners make better operational decisions. In terms of bed management, this necessarily involves some knowledge about the currently occupied beds, and potentially information about when they might become free. Therefore, the examples of operational bed management research summarised in this section are those which consider the current bed-state as part of the decision-making process.

Early work in this area includes Connors (1970), who reports an algorithm for managing bed occupancy by controlling the elective admissions process. Components of the algorithm bear some similarity to the typical components of an optimisation problem. It starts by extrapolating the bed census for the planning period by using a status file (which stores the status of all patients currently occupying a bed) and the length of stay distributions for each patient. The algorithm then looks for a feasible admission date by searching the bed census according to a constraint hierarchy. The hierarchy includes constraints for the total number of hospital beds, beds within the applicable specialty, and rooms containing patients with ailments most closely related to those of the admittee. A set of feasible admission days are then evaluated by an objective function, weighted by deviations from the requested admission date, and deviations from the target occupancy. This use of a status file and length-of-stay distributions to estimate the short-term bed census is conceptually very similar to loading and running the initial conditions of an ODES. However, without modelling a ward network, the downstream effect of patient transfers will not be captured in the predicted census. The model also assumes that emergency patients are allocated separate beds, and resource sharing between the emergency and elective patients does not occur. While this may be true for some types of hospitals, this research is concerned with modelling wards which are shared between these two patient types.

Kolesar (1970) describes a Markovian decision model for the bed census which accounts for both emergency and elective patient streams. The state of the model over time is defined as the number of occupied beds; therefore, the model's initial state can be matched to the real bed census. The state at the next time-point is equal to the current system state, plus the number of emergency arrivals, plus the number of elective arrivals, minus discharges. The number of emergency arrivals is a random variable, and the number of elective arrivals is treated as a deterministic decision variable. A linear program is formulated around the bed census model, to admit elective patients in such a way that average occupancy is maximised, whilst constraining patient overflow. While the Markov model for the bed census could be used to generate short-term predictions, the elective schedules derived from the linear program are more suitable for tactical planning, rather than operational planning, since they are based on the model's steady-state properties. The author also assumes that the arrival and discharge processes are stationary, although empirical evidence (Audit Commission, 2003, p.17) suggests that it is common for hospitals to experience a decrease in discharge rates over the weekend.

The bed availability model described by Kusters and Groot (1996) bears some similarity with Kolesar (1970). A state equation represents the number of available beds over time, which is the summation of current occupancy levels, along with elective and emergency admissions, less the number of discharges in each discrete time-period. Emergency patients arrive according to a Poisson distribution, whose rate parameter can change for each day in the planning horizon. Elective arrivals are treated as deterministic since they have already been scheduled some time in advance. Discharges occur according to length-of-stay distributions, which are conditional on the time already spent in hospital for the patients occupying a bed at the start of the prediction window. Expressions for the mean and variance of each of these terms are derived, which allows the same to be computed for the master equation as a linear combination of random variables. Results of 3-day ahead predictions are presented in terms of mean bed occupancy, suggesting good performance, however the model seems to only consider the number of patients under each specialism, rather than their physical location, and assumes that specialism is a constant from admission to discharge.

Koestler et al. (2013) formulate a bed census forecasting model for a neonatal intensive care unit (NICU). The "throughput" structure is familiar, with the forecast in k period's time equal to the number of current patients, plus the number of arrivals, less discharges in the same period. However, the formulation of the arrivals and discharge terms are more advanced. The arrivals term is itself a forecast, described by a Poisson Autoregressive model. The discharge term is particularly interesting, and is influenced by the time already spent in the NICU, and a generalized linear model to map patient-level covariates to length-of-stay. While the overall structure is intuitively appealing, the scope of the model is limited to a single hospital unit, in which all arrivals are probably emergencies.

Littig and Isken (2007) make heavy use of generalized linear models, to build a whole-hospital census analytical forecasting model which includes both elective and emergency patients. Five logistic regression models are fitted per hospital unit, which provide the probabilities of leaving the ward within 24, 48, 72 and 96 hours, and one for the likelihood of being transferred internally or discharged. Multinomial logistic regression models are also used to govern patient location, including the first and subsequent ward stays, should the patient be transferred. A linear regression meta-model is developed for each hospital unit based on current occupancy, inflow and outflow, to aggregate the patient-level information and provide bed census estimates. While this is one of the most comprehensive bed census prediction models reported in the literature, the quantity of patient-level covariates used to fit the various regression models is vast, and only an excerpt of the covariates is presented in the paper. This suggests that it might be difficult to recalibrate the model for another hospital if the same information is not collected. Without an event-based model of the ward network, it might also be difficult to capture the dynamic interactions which occur between wards, such as the occurrence of outlier patients.

Broyles et al. (2010) develop a so-called Discrete Time Markov Chain (DTMC) model, designed with the whole-hospital in scope. While Markov Chain models are commonly analysed via their steady-state properties, the authors focus on deriving expressions for the *transient* distributions of bed occupancy. By focusing on transient distributions, the theory can be applied to operational planning scenarios, rather than their typical use at the tactical planning level. The ability to say something about the predicted *distribution* of bed occupancy is clearly more useful than a point estimate, since any summary statistic can be derived from a distribution. While the authors present the results generated by their formulation for two hospital units, they note that these are treated in isolation, and further work could be carried out so that the bed census can be predicted in a network.

The model described by Yi et al. (2010) is designed to inform operational decision making in emergency scenarios, without explicitly capturing the state of the hospital at the start of the prediction window. A generic hospital discrete event simulation is developed which includes three main parameters; the number of beds in the hospital, the number of operating rooms and an OR efficiency index. Each of these parameters is tested at three possible levels, resulting in 27 possible experiments, although 6 are excluded as unlikely combinations. Under each combination, the simulation is "shocked" with an increase in the arrival of emergency patients to simulate an earthquake disaster, and performance statistics are collected from the transient period that results from the sharp increase in arrivals. Rather than recommending the model's use in a real disaster scenario, the transient performance statistics are regressed against each of the three simulation parameters. The resulting regression equation can be used to assess the impact of an emergency, by substituting the parameters of the hospital in question, rather than waiting for simulations runs to complete. While a model such as this can generate quick evaluations for extreme circumstances, it pays a price in terms of its flexibility, since detailed bed-state and patient-level information is not considered.

The work by Helm et al. (2011) has similar aspirations to that of Yi et al. (2010) in the sense that the details of the model are not designed to be run on an operational basis. Instead, the authors formulate a Markov Decision Process model (MDP) for the bed census, and make the case for three inpatient streams; emergency, elective, and expedited. The expedited patients are essentially elective patients who can be admitted at much shorter notice than regular elective patients. The MDP is used to determine two thresholds which divide bed occupancy into three "zones". In the lowest "call-in" zone, additional elective patients should be admitted from the expedited patient waiting list. In the "steady" zone, there is no reason to deviate from the elective admissions schedule. In the "cancel" zone, the hospital should consider cancelling planned elective procedures. While a zone-based admissions/cancellation policy could be straightforward to implement, it is not clear how the thresholds disaggregate to the ward level. The hospital might also be willing to explore the effect of other types of interventions, such as admitting elective patients to non-ideal wards, rather than outright cancellation. A flexible online model would allow for almost any type of elective schedule intervention to be assessed as necessary.

Vanberkel et al. (2011) expand on the surgical block optimisation literature by developing a queueing-based analytical model which can be used in both tactical and operational planning scenarios. Since master surgical schedules are intended to be used on an ongoing basis, the model can be used for tactical planning to assess the steady-state workload for each day in the cycle. Additionally, information about the actual patients in recovery can be treated as an input (along with the proposed surgical schedule) to make short-term bed forecasts for operational planning. An important feature of the model is that it can be used to derive exact distributions of the bed census, rather than restricting the user to point estimates. The authors present examples in which the bed census distributions provide the 90th percentile of surgical bed occupancy for each day in the cycle. However, it is unclear how the surgical recovery wards interact with other wards in the hospital, since their occupancies are estimated in isolation. The model also only accounts for the bed requirements of scheduled elective patients, although the authors note that the inclusion of non-elective patients is a feasible extension.

2.3.4 Bed Management Summary

The research summarised in the preceding sections suggests that simulation-based models which can assess the impact of elective admissions on inpatient beds, are most commonly found at the tactical planning level. In these applications, simulations are typically used to assess the relative effect of admissions policies, or the effect of admissions from a waiting list. However, the simulation methodology involves a degree of trial-and-error, therefore analytical methods have also been developed to find cyclic admissions patterns which are "optimal" in some sense. While these models have the potential to improve longer-term bed management, Vanberkel et al. (2011) note that reducing the chance of problems occurring does not guarantee that they will not occur. This notion necessitates the development of operational models.

In reviewing the literature, the models classified as "operational" are those whose results depend on the hospital's current state. This allows hospital planners to take action based on the most up-to-date information available. However, the research in this area tends to focus on analytical models which necessarily sacrifice some part of the hospital-wide operational bed management problem in order to be solveable in some sense. While simulation-based models cannot generate exact solutions, as some of the reported analytical methods do (such as distributions of the bed census), fewer concessions must be made in terms of the model's scope in order to generate results.

2.4 Online Discrete Event Simulation

The use of classical (offline) discrete event simulation to model complex systems is well established, and healthcare applications can be found as far back as the late 1960s (Fetter and Thompson, 1965, Smith and Solomon Jr, 1966, Barnoon and Wolfe, 1968, Goldman et al., 1968, Robinson et al., 1968, Fetter and Thompson, 1969), near the time when simulation languages were first becoming available (Nance, 1996, p.376). The term "offline" is used in this context to describe simulations which may be parameterised with data generated by a real system, although no persistent relationship exists between the simulation and the data source. In typical cases, offline DES is used to generate estimates of the long-run behaviour of a system (under a set of user-defined scenarios), and models of this type can provide useful results without requiring a deep understanding of analytical modelling methods, or indeed the simplifying assumptions needed to make analytical models viable.

Given the flexibility of offline DES, it is perhaps natural that practitioners would adapt the methodology to be able to make predictions about the evolution of complex systems in the short-term. However, the ability to do so depends on the availability of information about the system being modelled. As with offline

DES, historic data can be used for an initial parameterisation of an *online* DES model. Although crucially, very recent or “up-to-the-minute” information is also required, if only to define the starting point for any forecasts which might be made. With the proliferation of computer systems in almost every aspect of human endeavour, this requirement is becoming increasingly easy to satisfy.

In this section, some of the theoretical contributions to the ODES methodology are discussed, along with applications of the method within the healthcare arena. It is important to note that the literature often refers to models of this type as “real-time” or “symbiotic” simulations, as well as “online”. However, the term “real-time” implies a level of temporal granularity which might not be possible to achieve with the data acquisition systems commonly found in hospitals. Similarly, “symbiotic” implies a close relationship between the real system and the parameterisation of the model, as one improves the performance of the other. This type of “auto-validation” process (discussed in the next section) is out-of-scope of this research. For these reasons, the term “online” is preferred, although it is unclear whether an appreciable distinction exists in the field presently. Therefore, research under all three naming conventions are within the scope of this review.

2.4.1 Online Simulation Methodology

One of the first, and most comprehensive online simulation frameworks was developed by Davis (1998). The author’s schematic, presented in Figure 2.1, provides an overview of the framework, and proposes the use of parallel models operating under alternative control policies, along with a single model operating under the current control policy. Under this framework, the performance of each model is analysed, and the real system adopts the policy which generates the "best" simulated results (given the real system's current state) for the next planning period. The framework also includes an "auto-validation" process which updates the simulation parameters as new information is gathered about the real system. While this framework covers most, if not all the components one might include as part of a developing an ODES, not all of them are strictly required for a model to be considered "online", and Davis notes that the technology required to implement all the components did not exist at the time of writing. In particular, the auto-validation component is out-of-scope for this research, since the focus is on developing a proof-of-concept model first, before considering the details of ongoing maintenance.

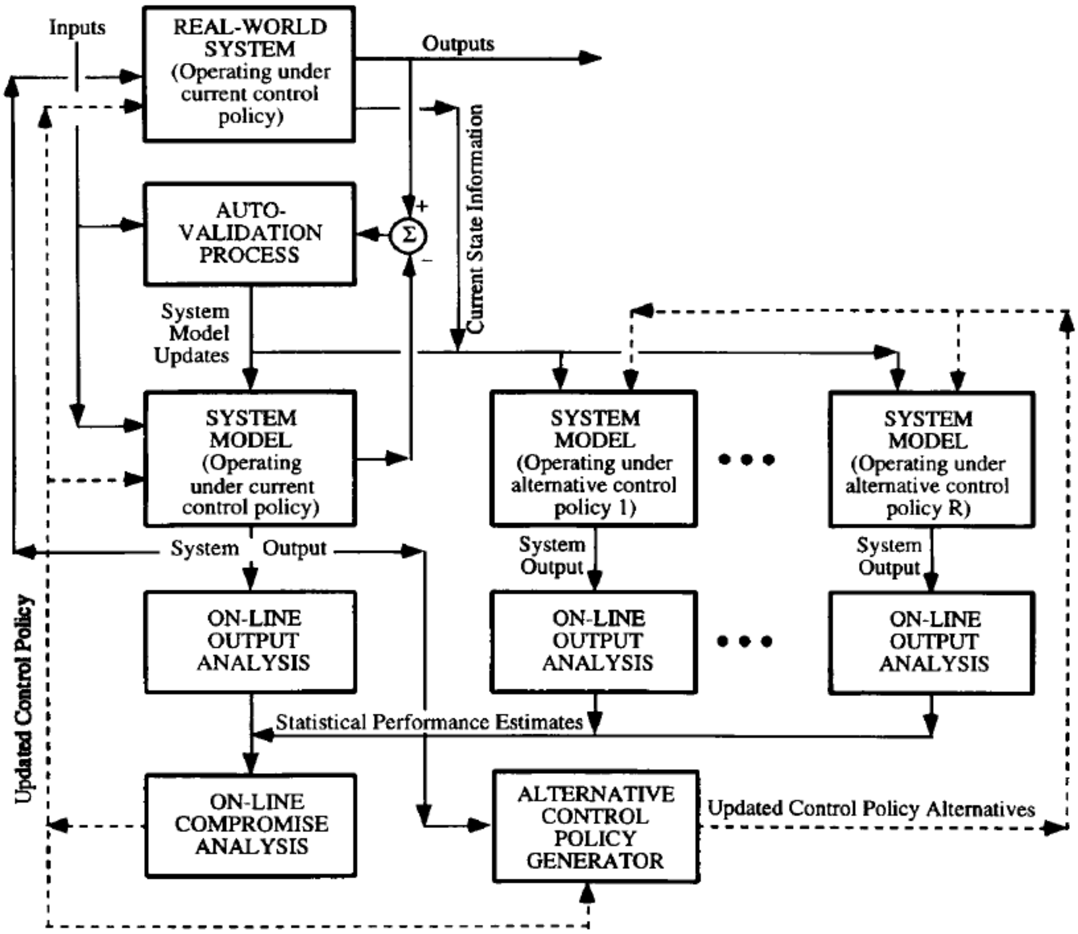


Figure 2.1: Schematic for the on-line planning/control process using real-time simulation (Davis, 1998)

Other components within the framework simply require clarification of what they represent in the hospital context. For instance, an “alternative control policy generator” is most likely to be a hospital admissions planner who is aware of resource constraints and patient scheduling considerations, rather than a computer-based component of a simulation. A similar approach is reported by Vanberkel et al. (2011, p.1857):

“Although a local search heuristic may have found an acceptable (or optimal) solution more quickly, it would have required the many surgical

department constraints to be modelled and would not have garnered the same level of staff understanding and support as this more manual process."

With manual adjustments to the elective admissions schedule being expected, the R alternative policies are more likely to be run sequentially, rather than in parallel, with the user improving the predicted performance for the planning period via iterative attempts. This approach is easily accommodated within Davis's framework, which was designed under the assumption that there may not be time for iterative adjustment before a decision must be made, hence the parallel execution of alternatives. The operational time-scales considered in this research occur over multiple days, therefore the emphasis on obtaining nearly instantaneous results is reduced.

Another important contribution of Davis (1998) is the discussion of "reactive" versus "proactive" decision-making using online models. In reactive mode, an ODES is used to develop a plan at a point in time (a so-called "decision point"), possibly in response to a critical state in the real system, which is implemented in the real system until the next decision point occurs. The alternative is a "proactive" mode, in which the plan is updated between decision points as the real system evolves. While either of these modes of operation could be applied to operational bed management in hospitals, they are dependent on the rate at which the hospital's databases can be synchronised with actual bed occupancy. For instance, if it is known that up-to-date data entry occurs only once per week, the hospital may be limited to reactive planning at weekly decision points.

Hanisch et al. (2005) further develop the theory of online simulation by considering some of the challenges associated with their initialisation. Since ODES models are initialised with a state reflecting the real system (rather than starting “empty and idle”) and analysed via their transient behaviour, the accuracy of the initial conditions has a direct effect on the results. However, in systems where the state descriptors change quickly over time, the current state becomes a moving target. The authors describe two initialisation methods. The first involves maintaining a continuously synchronised parent model, from which any number of child models can be generated and run at any time. The second is more simplistic and generates a model from a specially formatted file whenever a new simulation run is requested. Since the state of inpatient beds is thought to evolve at a slower rate than the example applications described by Hanisch et al. (traffic and pedestrian flow modelling), the initialisation method envisaged for this research bears more conceptual similarity with the second method. Additionally, hospitals may be able to choose times during the day when arrivals, discharges and transfers between wards are less likely to occur, thereby reducing the chance that the bed-state will change before the results are obtained.

Aydt et al. (2008) formalise the use of reactive and proactive modes of operation; described by Davis (1998). In reactive mode, an ODES can be triggered to assess alternative control policies only when a performance indicator of the real system exceeds a critical value. In proactive (or preventative) mode an ODES is run periodically as an early warning system, and alternative control policies are assessed if the forecasts exceed the same critical value. Although it is tempting to say that proactive mode is preferable, forecasts can be wrong, in which case the control policy could be changed unnecessarily. The authors develop a metric for comparing reactive and proactive response modes, and by considering two types of forecasting error (failure to detect a problem or detecting a non-occurring problem) expressions are derived for a probability threshold above which a proactive model would suggest taking preventative action. This represents a very useful way of calibrating ODES models, although it relies on a well-defined set of control policies which can be compared under reactive and proactive modes. This set is likely to be difficult to define in the context of elective admissions, in which admissions planners may have to consider multiple subjective criteria.

In terms of more application-focused work, one of the biggest areas of research interest for online simulation is in the manufacturing arena, including Gupta et al. (2002) (plastics processing), Potoradi et al. (2002) (semi-conductor manufacturing) and Low et al. (2007) (high-tech manufacturing and service networks). Additional areas in which the method has been applied include military operations (Hill et al., 2001), pedestrian flow (Hanisch et al., 2003) and traffic flow (Mazur et al., 2004). Despite some of the parallels that could be drawn between manufacturing and healthcare systems, applications of online simulation in the latter domain are relatively scarce, possibly due to differences in the data acquisition systems used between the domains. For instance, manufacturing plant systems can collect up-to-the-minute information about real processes by using sensors and detectors, however data collection at this level-of-detail is not commonplace (and is potentially unwelcome) in hospitals.

2.4.2 Online Simulation in Healthcare

Although relatively scarce, online simulations have been developed for use in healthcare applications; most commonly for modelling patient flows and treatment processes within Emergency Departments, where planning horizons are very short (on the scale of hours).

Hoot et al. (2008) report the development of a generic ED model which uses discrete event simulation to make forecasts of seven performance indicators; including waiting count, waiting time, occupancy level, length-of-stay, boarding count (the number of patients awaiting admission) and boarding time (the time between requesting a hospital bed and receiving it). The authors pay particular attention to validating the model, reporting the use of a "sliding window" technique which partitions their data into fitting and testing subsets which do not overlap. The window contains four weeks of ED data which parameterises the model and moves forward in time in 10-minute increments, updating the simulation parameters as it advances. The outputs generated by the simulation are the mean of 1000 replications compared against their counterparts from the testing subset via their steady-state distributions (independent of time), and Pearson's r coefficient of correlation at 2, 4, 6 and 8-hour forecasts. The correlation coefficient indicates how much of the variation in the testing data can be explained by the simulation model, and each estimate is benchmarked against the autocorrelation at the same intervals from the testing data alone. While this shows that the simulation model is likely to outperform an autoregressive forecasting model, it may be difficult to diagnose issues during development using this statistic. The authors also conduct a residual analysis to show that the forecasts are unbiased, although other properties of the distribution of the simulation outputs (such as the variance) are lost if the replications are averaged to obtain a forecast. Additionally, the ability to update the simulation parameters automatically (typically via an auto-validation component) may not be a trivial development for complex simulations. Therefore, the sliding window method may be difficult to apply in a more general context, especially during early stages of development.

Tan et al. (2013) also focus on the Emergency Department, developing a comprehensive model which aims to improve both the supply of ED resources and the management of patient demand. On the supply side, a symbiotic simulation is developed which generates demand forecasts based on the current ED state (or “snapshot”) along with historical data. The snapshot contains current queue conditions, doctors’ availabilities, patients’ statuses and arrival rates. The demand estimates from the symbiotic simulation are used to generate an optimised schedule for the supply of resources, such as doctors, over relatively short planning horizons. The author’s development of an symbiotic simulation which informs an optimisation component (and vice versa) is the only known application of this type in a healthcare setting, although the scope is solely concerned with the Emergency Department.

Marmor et al. (2009) and Espinoza et al. (2014) also develop real-time simulations of Emergency Departments, albeit with slightly different focuses. Since EDs normally have one of the highest throughput rates of any hospital department, both models disaggregate daily arrival rates into hourly rates to facilitate decision making over planning horizons of less that one day. Interestingly, both papers report challenges with initialising the models, due to incomplete data to represent the real system's state at run-time. This is probably to be expected given the frequency at which the models are intended to be run, coupled with manual data entry by ED receptionists. Both models overcome this challenge in a similar way; by running a warm-up period to populate the model first, and then by injecting the information associated with patients who *are* observable. Espinoza et al. (2014) refer to models in which the initial conditions are imputed in this way as "mixed input" simulations. Additionally, the authors investigate various levels of data completeness in order to assess the feasability of using their approach in practice. While data availability is clearly an important issue when modelling an ED on time-scales of less than one day, this research is concerned with the management of *inpatient* beds, where patients typically stay days or weeks before being discharged. On time-scales such as this, data availability issues are not expected to be encountered as regularly.

Bahrani et al. (2013) develop a real-time simulation to aid operational decision making over similar planning horizons (4-8 hours) to Hoot et al. The model focuses on a subsection of the clinical pathway for cardiac patients who arrive as emergencies and uses three metrics to assess the performance of each simulation; patient total waiting time, total hospital cost, and percentage of patients discharged. The performance metrics are computed under different scenarios which can be defined by the user, or from a pre-selected list, including the base case (running under the current configuration of the real system), additional ED staff, additional cardiac staff, additional beds or reductions in these resources. Using a similar approach to Vanberkel et al. (2011), the estimates of the three performance indicators are intended to be judged by hospital staff in the light of the criteria considered to be most important at the time, before implementation in the real system.

Finally, Mousavi et al. (2011) report the development of a system for the real-time monitoring of patient quality-of-care throughout a hospital, using a Healthcare Quality Index (HQI) of their formulation. Observed events in the real system are tracked over time, which correspond to parameters in a discrete event simulation model. The simulation is run under the latest parameter values, and the performance statistics which are generated form the basis of the aggregate HQI calculation, to provide a real-time indication of quality-of-care.

2.5 Conclusions and Research Questions

The research examined in the preceding sections, to help with tactical and operational bed management, suggests that models which encompass both elective and emergency workloads, while keeping the entire ward network within scope, are relatively few. The models which do, are largely intended to inform tactical planning decisions, rather than being influenced by the current state of the hospital. At the operational planning level, bed management tends to be addressed via analytical methods, which often consider only a subset of acute patient types or pathways, or treat bed resources as homogeneous; possibly in aid of mathematical tractability. An important advantage of using simulated-based methods for prediction, is that fewer of these simplifying assumptions are needed to generate estimates of key performance indicators, or to approximate their distributions.

One of the trade-offs of using simulation-based methods is that they rely on educated trial-and-error to produce improved control strategies, rather than solving for them directly via a mathematical formulation. With the elective schedule as a decision variable, this might seem like a problem, since there are many possible combinations of patient numbers, wards and admission days which could be tested. However, in practice, schedules for short-term bed management are likely to be constrained by quotas for different patient types, staff availability, and the surgical schedule used to book patients in advance. Therefore, it is likely that an ODES can be used in an iterative way, via manual adjustments to the admissions schedule by hospital planners who are already aware of resource constraints.

In terms of the *online* simulation literature, a whole-hospital bed management model, or indeed a methodology for developing one, is not known to exist. Therefore, the first research question for this thesis is:

> ***RQ1:*** *How can an on-line simulation, which provides estimates of bed demand, be developed for the operational management of hospital beds at the ward level?*

Since the aim of this research is to develop an ODES for *ward-level* bed management, one of the challenges is to represent the ways in which hospital wards might interact with one another. For example, if a simulation is initialised with a bed-state from the real hospital which includes a ward near to, or at its maximum capacity, this may have short term implications for other wards, through the occurrence of outlier patients. Without detailed consultation with hospital staff on the nature of this interaction, including the possible locations of diverted patients, this behaviour becomes difficult to model in a meaningful way. Additionally, the results of any consultation could reveal diversion “rules” that are subjective in nature and would therefore need to be summarised and simplified for the purposes of modelling. If this information can be extracted from data which is routinely collected, the consultation process can be circumvented, and the diversion rules can be summarised into a component of the simulation. The second research question is therefore:

> ***RQ2:*** *Can the effect of hospital busyness on patient-to-ward placement decisions be detected in patient administrative data, and can this be incorporated in a simulation model? If so, what effect does it have?*

In addition to the patient administrative data collected by most hospitals, supplementary patient-level information can also be recorded on a less formal basis, such as patient notes, or staff whiteboards, or it can evolve over time based on the recovery process of the individual patient. One such example is Estimated Date of Discharge, which may or may not be formally recorded, and is subject to change throughout a patient’s stay. Information of this type lends itself well to inclusion in ODES models which are also designed to be regularly updated. Prevailing patient-level knowledge can then be loaded as part of the initial conditions. Therefore, the final research question for this thesis is:

> ***RQ3:*** *How can additional patient information, made available at run-time, affect the estimates of bed demand from an online simulation?*

The three research questions formed in this section focus the subsequent chapters in this thesis, and are addressed in Chapters 4, 5 and 6 respectively.

In the next chapter, collaboration with an Australian General Hospital is discussed, along with some of the details of the data extract which forms the basis of the ODES development work.

Chapter 3
A Real Hospital and its Data

3.1 Introduction

With the research questions defined (Section 2.5), the next phase in developing the online simulation involved partnering with a real hospital whose staff shared our interest in developing a tool for operational bed management. This chapter provides an overview of the hospital we worked with, and the collaboration that took place to ensure the research focuses on real issues, while also using realistic data. The details of the pre-processing steps which ready the data are also discussed, along with the scope of the included patient episodes, followed by a ward-level analysis based on statistics which are typically used to parameterise simulation models. The resulting database informs the model development and validation work reported in subsequent chapters of this thesis.

3.2 Background

Because online simulations are designed to have a close and persisting relationship with the systems they model (and their end-users), it was important from the outset of this research to work in partnership with a hospital for at least three reasons:

1. To ensure our area of academic interest can be applied in practice.

2. To gain subject-matter expertise and guidance so that the model is relevant to the challenges faced by hospital staff.
3. To access any data already collected by the hospital and confirm its correct interpretation during analysis and model development.

To this end, discussions surrounding a collaborative project began with staff at an Australian General Hospital (AGH) who expressed interest in participating in the research. The AGH is a 300-bed public hospital which provides acute care facilities for over 67,000 residents and treats over 24,000 inpatients annually. Its services include Cardiology, Renal, Gastroenterology, Haematology-Oncology, Rehabilitation, General Surgery, Ear/Nose/Throat Surgery, Plastic Surgery, Orthopaedics, Radiology and Paediatrics. The facilities also include an Emergency Department and an Intensive Care Unit.

On the 1st of February 2012 the first telephone conference took place with managers and staff at the AGH and our group in Lancaster. This first conversation was not aimed at clarifying many technical details, however we gained some insight into the areas they deemed important and found that these aligned closely with our research aims. Specifically, staff expressed interest in an early warning system for detecting overcrowding on inpatient wards; for which an ODES model is well suited. Additionally, we learned from the AGH nursing managers that there is frequent overflow of outlier patients from medical to surgical wards, which poses several challenges including relocating equipment, and potentially causing staff to work outside their area of expertise. The frequent occurrence of outlier patients also coincided with an area of potential research interest, since methods for modelling these patient flows are not often reported in the bed management literature.

After these initial talks, we continued to work with the AGH for approximately one year. During this time, we were supplied with an anonymised extract of the patient administrative database. Following a preliminary analysis of the data and the development of a small-scale simulation, a progress report was sent back to the AGH containing information about historic patient activity levels and examples of how these can be modelled using simulation. One further telephone conference was conducted to discuss the contents of the report; giving staff the opportunity to provide feedback on our modelling approach and use of their data. However, despite early enthusiasm, we lost contact with the AGH after this call, due to changes to key personnel and a lack of resources on their part. Because of the lack of contact beyond the early development phase, the name of the (previously) participating hospital is suppressed throughout this thesis.

Although further input would have been welcomed, our early interactions with the AGH meant that the preliminary stages of model development were more grounded than they would have been, had we accessed alternative data sets without any consultation process. The remainder of this chapter explains the analysis and filtering applied to the raw data we obtained from the AGH, to ready it for parameterising an ODES model.

3.3 Scope and Filtering

The patient administrative (PA) data sets supplied by the AGH contain information for all patients occupying a bed between the 1st of January 2010, and the 30th of June 2012. The data is split into two parts, known as the "InpatientStay" and "Inpatient" data sets. The InpatientStay data set contains information related to each patient's location, and additional rows are created when these details change within an inpatient visit or "episode". The Inpatient data set has a different structure; containing visit-level information with one row per patient, per episode. Because of this structure, any overarching information which applies to an entire patient stay (admission to discharge) is recorded here, including demographic details such as age and sex, and episode-specific information such as admission type, date/ward of admission, date/ward of discharge and specialty at admission/discharge. Because this information is summarised for each episode, the Inpatient data set is used to determine the scope of the data which parameterises the ODES model. Figure 3.1 shows a five-row excerpt of the Inpatient data set, and some of the key fields used in the preliminary analysis of the data.

	Patient_Id	Inpatient_EpisodeId	Admission_Type_RefId	Admission_DateTime	Discharge_DateTime
1	524	690378	Urgency status not assigned	29MAR2012:07:00:00	30MAR2012:14:20:00
2	542	678433	Urgency status assigned - emergency	22FEB2012:13:38:00	06MAR2012:14:03:00
3	582	231239	Urgency status assigned - emergency	27MAY2010:17:23:00	01JUN2010:15:08:00
4	692	217336	Urgency status assigned - elective	01FEB2010:22:50:00	03MAR2010:10:05:00
5	692	599611	Urgency status assigned - emergency	04JUL2011:23:05:00	08JUL2011:10:45:00

Admission_LocationId	Discharge_LocationId	Admission_SpecialtyId	Discharge_SpecialtyId	Actual_SameDay_Flag
Ward 4B	Ward 4O	Obstetrics	Obstetrics	0
ED	Ward 6D	General Medicine	General Medicine	0
ED	Ward 6D	Cardiology	Cardiology	0
Ward 01	Ward 01	General Practice/Primary Ca	General Practice/Primary Ca	0
Ward 01	Ward 01	General Practice/Primary Ca	General Practice/Primary Ca	0

Figure 3.1: Five inpatient episodes recorded in the Inpatient dataset. Patients with more than one episode are recorded under distinct episode identifiers (Inpatient_EpisodeId) but maintain their individual patient identifier (Patient_Id).

One of the most important variables in the Inpatient data set is Admission_Type_RefId, which categorises patients into three main admission statuses: "Urgency status assigned – elective", "Urgency status assigned – emergency" and "Urgency status not assigned". The data set also includes five other admission types, however these make up less than 4% of the total number of inpatient episodes.

The majority of the "Urgency status not assigned" episodes appear to be maternity-related, with over 70% having an Admission_SpecialtyId of Obstetrics (41.8%) or Paediatric Medicine (29.8%). A potential reason for maternity-related cases having this admission type, is because they are not emergency patients, nor do they arrive from a waiting list; meaning the scope may be limited for managing this type of patient. For this reason, the not-assigned admissions are considered to be out of scope, and these records, along with all rarer admission types, are removed from the Inpatient data set to focus on only emergency and elective patients.

Similarly, patients admitted under the "Renal" specialty have also been removed from the dataset. The AGH has a separate Renal Unit, and patients rarely require treatment on other wards. Because of the simplistic care-pathway, and relatively low levels of interaction with other wards, patients under this admission specialty have also been excluded.

Finally, the "Actual_SameDay_Flag" field has been used to remove episodes which do not require an overnight stay. One of the most well-known metrics for inpatient bed usage is the "midnight bed census" and this is also the focus of the modelling work (discussed further in Chapter 4). The midnight census counts the number of patients occupying a bed at midnight, however patients who are admitted and discharged on the same day will not contribute to this metric. Therefore, episodes for which Actual_SameDay_Flag = 1 have been removed from the Inpatient data set.

The InpatientStay data set is linked to the Inpatient data set by the Patient_Id and Inpatient_EpisodeId fields. Therefore, the same set of exclusions are applied to InpatientStay, using the patient and episode identifiers which remain in the Inpatient data set. Since one of the aims of this research is to develop a model for ward-level bed management, patients' location data are central to determining the structure and parameterisation of the ODES model. To this end, a subset of the InpatientStay data set (which tracks patient locations over time) is created based on the filtered Inpatient data set. Figure 3.2 shows an example of the InpatientStay data set for the same five episodes shown in Figure 3.1.

	Patient_Id	Inpatient_EpisodeId	Stay_Segment_Start_DateTime	Stay_Segment_End_DateTime	Bed	LocationId
1	524	690378	29MAR2012:07:00:00	29MAR2012:14:00:00	Room 5	Ward 4B
2	524	690378	29MAR2012:14:00:00	30MAR2012:14:20:00	Bed 33	Ward 4O
3	542	678433	22FEB2012:13:38:00	22FEB2012:21:56:00	DEM Bed	ED
4	542	678433	22FEB2012:21:56:00	23FEB2012:22:15:00	Bed 06 Isolation Roo	Intensive Care Unit
5	542	678433	23FEB2012:22:15:00	01MAR2012:13:10:00	Bed 05	Intensive Care Unit
6	542	678433	01MAR2012:13:10:00	06MAR2012:14:03:00	Bed 25	Ward 6D
7	582	231239	27MAY2010:17:23:00	27MAY2010:19:04:00	DEM Bed	ED
8	582	231239	27MAY2010:19:04:00	28MAY2010:02:31:00	Bed 13 - NCU	Ward 6D
9	582	231239	28MAY2010:02:31:00	01JUN2010:15:08:00	Bed 12 - NCU	Ward 6D
10	692	217336	01FEB2010:22:50:00	02MAR2010:12:26:00	Bed 03	Ward 01
11	692	217336	02MAR2010:12:26:00	03MAR2010:10:05:00	Bed 06	Ward 01
12	692	599611	04JUL2011:23:05:00	08JUL2011:10:45:00	Bed 02	Ward 01

Figure 3.2: Five inpatient episodes recorded in the InpatientStay dataset. This data set focuses on patients' location details, and a new row is created whenever these details change within an inpatient episode.

Figure 3.2 shows how each inpatient episode can be disaggregated into "stay segments" when patients' bed or wards details change during their stay. These changes are represented by a set of records for each episode which are contiguous over time; meaning the end time of the previous segment corresponds to the start time of the next segment. This structure allows each patient's location details to be completely accounted for between admission and discharge.

Although the InpatientStay dataset contains information relating to individual bed-stays, this level of detail is finer than is needed for a ward-level bed management model. For example, Figure 3.2 shows that during episode 217336, patient 692 switches from Bed 03 to Bed 06, while remaining on Ward 01. For ward-level modelling, it is assumed that beds within the same ward form a homogeneous group. Therefore, adjacent stay segments on the same ward (LocationId) are collapsed into one row; populated with the start time of the first stay segment, and the end time of the last stay segment. Finally, the admission type field (Admission_Type_RefId) is combined with the stay segment data, thereby classifying each set of segments within an episode to either an emergency or elective admission.

3.4 Preliminary Analysis

As mentioned previously, the midnight bed census is one of the most important inpatient metrics, therefore it makes sense to derive its time series from the PA data to check for any long-term trends or seasonality. The midnight census is derived by counting the number of patients whose admission and discharge dates span midnight for each day in the observation period. The resulting series also serves as a benchmark for model validation in the chapters which follow.

Figure 3.3 charts the midnight bed census (or midnight occupancy) over time for the emergency and elective admission types, for all hospital wards, after filtering the data. The AGH clearly has a greater number of emergency patients resident at midnight throughout the observation period, and on average, the ratio of emergencies to electives is approximately 5:1. There also appears to be some non-stationarity in both the emergency and elective series. For the elective patients, a slight downward trend occurs for most of the observation period. In the emergency series, the non-stationarity is more noticeable; with an upward trend in the first four months, and a decline during the last nine months. However, the pattern of decline between October 2011 and February 2012 does not appear to occur in the previous year, suggesting reasons other than seasonality.

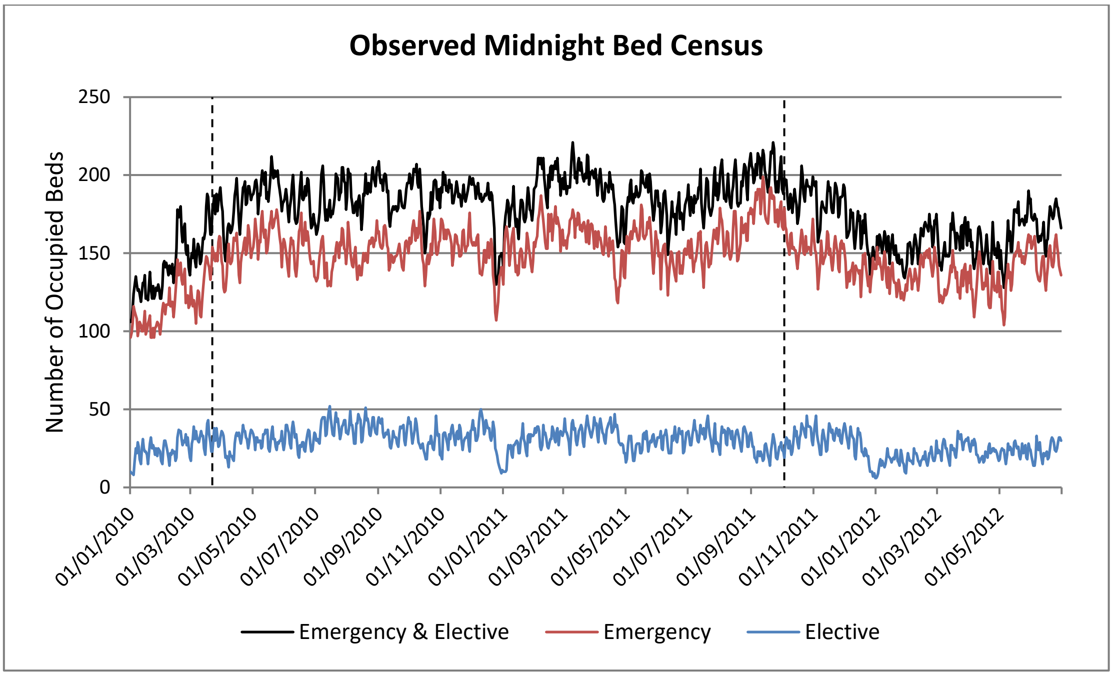


Figure 3.3: The emergency and elective midnight bed census during the observation period in the PA data.

The existence of longer-term trends in key simulation parameters (such as patient arrival rates) are important considerations when developing an ODES. However, this research is firstly concerned with developing a proof-of-concept model. Additional complexity, such as trends caused by seasonal effects (or other reasons) can be added at a later stage, after demonstrating the model's capabilities over the planning horizons it is designed for. For this reason, a set of time-based exclusions are also applied to the InpatientStay data set to remove some of the trending behaviour seen in Figure 3.3. The result is a subset containing the inpatient episodes occurring between the 22nd of March 2010 and the 3rd of October 2011 (indicated by the period between the vertical lines in Figure 3.3) which is less likely to be affected by external or systemic factors which change over time.

The final dataset contains the stay segments which occur during each inpatient episode, with each segment coded to one of 20 wards, and classified by admission type (emergency/elective). Table 3.1 summarises the data by ward, with statistics derived from the midnight census, along with arrival rate and length-of-stay information which is typically used in DES modelling. The summary statistics are in descending order of the average midnight census, with the top five wards contributing to over 70% of the hospital-wide census for emergency and elective patients. In terms of the proportion of occupied beds, these five wards experience occupancy levels of approximately 80% or higher on average, which could indicate a greater likelihood of encountering capacity-related issues. Further down the table, some of the wards exhibit very low occupancy levels, with the bottom eight rows displaying midnight census levels which average less than one patient. This is caused by the exclusions applied earlier (in the case of Ward 4O, Ward 4N and the Renal Units), or by the rarity of overnight stays (in the case of Theatres and the Day Procedure Unit), rather than indicating frequently empty wards. It should be noted that the proportion of occupied beds at midnight (third column) is calculated from the highest midnight bed census observed in the filtered PA data, rather than the total number of physical beds on each ward. Using the total number of beds in the denominator would overestimate the capacity available to the within-scope patients.

Wards (PA Data)	Avg. Midnight Census	Avg. Midnight Occupancy (% of Observed Maximum)	Avg. Emer/Elec Split at Midnight (%)	Avg. Arrival Rate (per Day)	Average LOS	Number of Stay Segments
Ward 5D	28.90 ± 0.16	88%	83/17	4.13 ± 0.19	7.08 ± 0.33	2344
Ward 5B	28.52 ± 0.27	86%	65/35	7.56 ± 0.25	3.79 ± 0.15	4256
Ward 4D	26.00 ± 0.22	84%	95/5	4.98 ± 0.24	5.40 ± 0.30	2813
Ward 5A	25.79 ± 0.19	81%	66/34	5.93 ± 0.20	4.43 ± 0.26	3345
Ward 6D	25.10 ± 0.22	78%	96/4	2.76 ± 0.16	9.34 ± 0.75	1572
Ward 4K	11.93 ± 0.32	50%	81/19	4.45 ± 0.19	2.70 ± 0.21	2504
Ward Northside	11.79 ± 0.25	59%	99/1	1.10 ± 0.09	11.04 ± 1.06	626
Intensive Care Unit	7.84 ± 0.14	65%	81/19	1.96 ± 0.12	4.04 ± 0.40	1097
ED	7.31 ± 0.34	33%	99/1	18.69 ± 0.35	0.39 ± 0.01	10475
Ward 3R	5.29 ± 0.21	41%	92/8	0.82 ± 0.09	6.47 ± 1.00	464
Hospital in the Home	3.31 ± 0.17	33%	79/21	0.26 ± 0.04	13.37 ± 2.31	144
Ward 4O	1.13 ± 0.09	19%	65/35	0.42 ± 0.06	2.69 ± 0.44	239
Ward 4N	0.94 ± 0.14	10%	92/8	0.13 ± 0.04	7.91 ± 2.40	82
The Manor Transitional Unit	0.33 ± 0.04	16%	50/50	0.01 ± 0.01	22.88 ± 24.56	8
Theatre Main	0.14 ± 0.03	7%	96/4	0.58 ± 0.07	0.23 ± 0.05	326
Day Procedure Unit	0.09 ± 0.04	1%	48/52	4.86 ± 0.34	0.02 ± 0.01	2723
Ward 4B	0.06 ± 0.02	3%	94/6	0.11 ± 0.03	0.54 ± 0.16	61
Renal Unit	0.02 ± 0.01	2%	92/8	0.92 ± 0.09	0.02 ± 0.02	513
Theatre Angio	0.02 ± 0.01	2%	90/10	0.34 ± 0.06	0.05 ± 0.03	193
Renal Unit - North West	0.01 ± 0.01	1%	0/100	0.01 ± 0.01	1.33 ± 3.79	3
Total	**184.51 ± 1.14**	**85%**	**83/17**	**28.75 ± 0.68**	**6.53 ± 0.16**	**16276**

Table 3.1: Summary statistics for each ward with stay segments in the filtered PA data. The ranges associated with the midnight census, arrival rates and lengths-of-stay are 95% confidence intervals for the means.

The split between emergency and elective patients (fourth column of Table 3.1) shows that for most wards, the emergency patients outnumber the elective patients in terms of average midnight occupancy, although the proportion fluctuates by ward. In terms of absolute occupancy, Ward 5B has the highest number of elective patients, averaging approximately 10 occupied beds at midnight, closely followed by Ward 5A which averages approximately 9 occupied beds. The only ward where elective bed occupancy could outnumber that of the emergency patients is the Day Procedure Unit. However, overnight stays at this location are rare, making it less important from the perspective of inpatient bed management. The other locations with higher elective proportions only have a handful of stay segments in the filtered PA data, therefore no meaningful conclusion can be drawn about the emergency/elective split.

As might be expected from the high proportion of beds occupied by emergency inpatients hospital-wide, the arrival rate at the Emergency Department (ED) is easily the highest among the wards in the PA data (fifth column of Table 3.1), and more than doubles the next highest arrival rate (Ward 5B). It should be noted that the arrival rates to each ward includes internal transfers from other wards, as well as new admissions to the hospital. However, the total row includes only new admissions, since internal transfers would not be classed as arrivals at the whole-hospital level. Therefore, the sum of the ward-level arrival rates is necessarily greater than the whole-hospital arrival in the Total row.

Of the wards exhibiting higher levels of average occupancy, Ward Northside has the highest average length-of-stay (sixth column of Table 3.1) of approximately 11 days. All patients admitted to the Northside ward do so under the Psychiatry specialty, in which patients often require longer hospital stays that those presenting with physical disorders (Mechanic et al., 1998). Two locations exist which have average lengths-of-stay higher than Northside ("Hospital in the Home" and "The Manor Transitional Unit"), however their contribution to overall midnight occupancy is negligible. The Manor Transitional Unit has the highest average length-of-stay of all wards in the filtered PA data. However, the very low number of observed stay segments (along with the influence of potential outliers in the sample) causes the confidence interval to be wide relative to its mean, therefore the estimate may not be reliable.

Although 20 wards are referenced in the data, some of them make relatively small contributions to the hospital's midnight census (on average) based on the patient episodes which are within scope. For this reason, modelling every ward as an individual location in the simulation may not be practical, especially if the number of observed stay segments is small. Since these types of considerations also inform the model's structure, further discussions about selecting wards to be individually modelled take place in the next chapter (Section 4.4), where other structural elements of the ODES are characterised.

Chapter 4
Model Development 1: From Offline to Online

4.1 Introduction

The purpose of this chapter is to describe the development process which results in an online simulation for operational bed management. The process starts with a discussion of the requirements for an online simulation, and how these apply in the context of hospital planning (Section 4.2). These requirements help to inform the subsequent development phases, beginning with a conceptual model which outlines the structure, components and level of detail needed for an operational planning simulation (Section 4.3). In Section 4.4, the parameterisation of the simulation takes place by analysing the arrival patterns, length-of-stay distributions and ward transitions in the PA data. These findings, along with the conceptual model, are combined in the Micro Saint Sharp simulation package, resulting in an “offline” model of the AGH. This model undergoes the first of two validation steps to check for problems which could be detected over longer simulation runs (Section 4.5). Section 4.6 discusses the necessary additions to move from a stochastic *offline* simulation to a stochastic *online* simulation. These include defining the system state and expressions for sampling from conditional length-of-stay distributions. The second validation step takes place in Section 4.7, using novel methods which account for the conditional distribution of the performance measures. Methods for both discrete and continuous performance measures are developed, and the discrete case is demonstrated for midnight bed occupancy. The chapter ends with a discussion of how Research Question 1 has been answered, and the conclusions which can be drawn from the model development process (Section 4.8).

The focus of this chapter is the development of the core of the simulation model, while Chapter 5 and Chapter 6 address the incorporation of two additional modelling components which arise as part of answering Research Questions 2 and 3. These additional components could be seen as part of the model development process (and therefore contribute to answering Research Question 1), however they also stand alone as research contributions to the online simulation methodology and to patient flow modelling in their own right. These additional components are therefore discussed in separate chapters.

4.2 Online Simulation Requirements

In Chapter 2, an outline of the framework developed by Davis (1998) was presented, which describes the generic components, and component dependencies of an online simulation. This framework (Figure 2.1), describes what could be considered an ideal online simulation, however, not all the components are strictly required for a model to be considered "online". Indeed, Davis acknowledges that the technology required to implement all of the components in the framework did not exist at the time of writing.

Conversely, there are some components which are not discretionary if a discrete event simulation is to be considered "online". Hanisch et al. (2005) list three requirements for online simulation control systems and these requirements form the basis of the model development process described in this chapter, however the requirements are generic, and some discussion is necessary in order to clarify how they might be applied in a hospital context.

Requirement 1: *A validated simulation model of the real system in which the level of detail of the simulation model must be equivalent to structures in the real system.*

This requirement covers the conceptual modelling and parameterisation stages of development, using the same methodology that would be applied in the development of a non-terminating or steady-state simulation. The resulting model is referred to in this thesis as an "offline" simulation, which is initialised with an “empty and idle” state. Programmatic adjustments can then be made to the offline model to accept initial conditions *other* than empty and idle.

One of the expectations of Requirement 1 is that a *validated* model of the real system can be obtained, however it is not immediately clear how a model should be validated for online use. Therefore, one of the research contributions of this chapter is the development of a two-stage validation process which is influenced by online modelling concepts. In the first stage, validity of the offline model is considered for the purpose of ruling out any unexpected behaviour in the long run, such as systematic bias, which may be difficult to detect in the comparatively short runs of an online simulation. In the second stage, the simulated and observed data are both treated as draws from conditional distributions which depend on the hospital state and the elapsed time from initialisation. These dependencies inform the discussion, development and application of a new method for validating online models (Section 4.7).

Requirement 2: *An online connection of the simulation with the real system.*

Requirement 2 ensures that the simulation can be initialised to the state of the real system being modelled, and in doing so, be brought online in order to investigate how the real system might evolve given its current state. In general, having an online connection with the real system means that the real system state can be queried by the simulation model at any time, however in a hospital setting this may not be possible since data entry into the patient administration system may not occur automatically. Unlike manufacturing systems where sensors can be used to track jobs and automatically update centralised data stores, hospitals often rely on nurses, clinicians and admin staff to enter this data manually. Unless this is carried out diligently, the PA data may not be synchronised with the real state of the hospital, but in order for an online simulation to be of use in the operational planning process, it must be assumed that synchronisation occurs at some regular points in time. At these times, it is appropriate to initialise the online simulation, but unlike other online simulations, it is not assumed that initialisation is always possible. The steps and considerations necessary to carry out the periodic initialisation of the offline model, thus bringing it online, are discussed in Section 4.6.

Requirement 3: *The simulation engine has to be fast enough to deliver results in a period of time that allows using the results in the subsequent decision process.*

For applications of online simulation where the planning horizon may be very short, Requirement 3 is important to consider. However, in the context of managing inpatient beds, ensuring that the simulation can complete before the next decision point is reached is not so important since operational planning in this setting occurs with daily/weekly frequency, and this is more than enough time to complete a batch of simulation runs.

The process of meeting Requirements 1 and 2 form the basis of the model development described in this chapter. To meet the first requirement, an adequate conceptual model is developed and then implemented in a simulation software package with the modelled sub-processes (such as arrivals and service) suitably parameterised. The resulting offline model is validated using standard techniques. The validated offline model is then brought online by assuming a frequency at which up-to-date data is available and initialising the model at these times. This is a slight modification of the second requirement in the sense that the true system state may not be able to be queried at any time; however the underlying rationale for using online simulation remains the same; that events in the near future are dependent on the (knowable) system state in the present.

4.3 Conceptual Model

Before the development of a simulation can begin using simulation software, some consideration must be given to the aspects of the real system which are going to be included, and how they should be modelled. This phase of development is known as “conceptual modelling”. A conceptual model forms the theoretical basis of the programmed simulation and is largely independent of software choices. In this section, a conceptual model is formulated based on a discussion of the level of detail required to meet the aims of the research, along with the time-scales associated with ward-level inpatient bed management.

4.3.1 Level of Detail – Temporal

Inpatient beds are generally distinguished from other bed types, such as day beds, by the amount of recovery time required by the patient occupying it. Day beds, as the name suggests, are mainly used for the treatment of day cases, where the patient being treated is expected to leave on the same day that they arrive. Inpatient beds on the other hand, will generally be occupied for one or more nights. Therefore, a key performance measure associated with inpatient beds is the number of occupied beds (and its complement; the number of available beds) at midnight each day, also known as the “midnight census”. This information is currently collected and published by the NHS (for example) on a quarterly basis. The midnight census also features as a performance indicator in a number of bed management models in the literature, such as Kolesar (1970), Esogbue and Singh (1976), McClean and Millard (1995), El-Darzi et al. (1998) and Helm and Van Oyen (2014).

Given the prevalence of the midnight census as a metric for inpatient bed occupancy in both the literature and official statistics, it has also been chosen as the key measure of ward-level bed occupancy for this model. Therefore, the greatest frequency of data collection from the simulation will be daily collection at midnight, and higher degrees of temporal level-of-detail (such as hourly performance indicators) will not be modelled. Since the shortest time interval of interest is one day, the simulation runs in discrete time, with each time unit representing one day of hospital operations.

4.3.2 Level of Detail – Structural

The structural level of detail of a simulation determines which of the physical processes in the real system are to be included as modelled processes in the simulation. Günal (2008) discusses structural level of detail in terms of the level of aggregation at which physical processes of an A&E department are included in the simulation. At high levels of detail, many of the processes of an A&E can be modelled, however, this requires equivalently detailed data collection to adequately parameterise them. Günal notes that detailed data requirements can be circumvented by modelling at lower levels of detail, although this level must not be so low that the modelling objectives can no longer be met. Therefore, a balance must be struck when considering the level of detail of the structural elements of a simulation.

Since this project is concerned with estimates of inpatient bed occupancy at the ward level, the minimum level of structural detail includes a network of wards. Patient stays in the PA data can be disaggregated into ward stay segments to parameterise each ward in the simulated network. However, modelling every ward which appears in the PA data is not considered sensible. For example, there is little point in modelling wards which rarely allow overnight stays since this research is concerned with estimates relating to the midnight census. On the other hand, omitting these wards would break the links in the ward network. A pragmatic approach is to aggregate information relating to these wards into one pseudo-ward in the simulation (referred to as ward *Other*), meaning the population of interest is captured entirely, while modelling effort is reserved for wards which are individually significant. The details of this process are described in Section 4.4.1.

It should be noted that it is possible for a ward to have high patient throughput and not be included as an individual ward in the simulation (aggregated with other wards instead). This would occur if many patients were treated on the ward, but most were transferred to another ward when an overnight stay is required. Since the midnight census has been chosen as the performance measure, it makes sense that it forms the basis of the inclusion criteria for the modelled wards. Although this means that wards with high midnight occupancy are treated as being more important from a modelling perspective, than say, those with high throughput or high patient turnover.

4.3.3 Uncapacitated Wards

Often when service networks are modelled (using computational or analytical methods) each node in the service network has a fixed capacity for the number of customers which can be served concurrently. If the number of customers requiring service exceeds this number, customers can either queue for service or they can be diverted to another service node.

For patients within the scope of this research (that is, designated emergency/elective inpatients spending one or more nights in hospital), it is unlikely that formal queueing behaviour would occur in practice. Instead, it is more likely that if the first-choice bed is not available, patients either remain in their current bed, or are diverted to a potentially less suitable bed on an alternative ward.

While a model which diverts patients away from full wards has the potential to capture ward dependencies known to exist in most general hospitals, no consensus has been reached in the literature for modelling this behaviour. The lack of agreement on how to divert patients in a model can be attributed to the many factors which influence bed placement decisions, such as the sex of the patient, their condition, nursing constraints and estimated length of stay, therefore simplifications are often made in order make modelling patient diversions feasible.

Examples of models which account for patient diversions include Günal (2008) in which patients are routed to alternative wards by setting the probability of arriving at a ward when it is full to zero, then rescaling the remaining ward probabilities to sum to unity. Harper and Shahani (2002) asked bed managers to provide information relating to patient priorities, and these come into effect when attempts are made to place a simulated patient on a full ward.

An alternative to diverting patients is to use so-called “uncapacitated” wards. As the name suggests, each modelled ward has theoretically infinite capacity, therefore every simulated patient stays on their most appropriate ward. Examples of this approach include de Bruin et al. (2010), Chow et al. (2011), Helm and Van Oyen (2014) and Monks et al. (2016). Monks et al. and de Bruin et al. use the infinite capacity assumption to help determine what appropriate facility sizes should be, while Chow et al. and Helm and Van Oyen use so-called “soft” capacity bounds on infinite capacity wards to determine the number of excess patients that would occur. Exceeding one of these bounds flags an instance where demand for a ward might need to be reduced and also allows for simple estimation of the probability of exceeding ward capacity i.e. the number of simulation runs in which the capacity at midnight is exceeded, divided by the total number of simulation runs.

Since the probability of exceeding capacity on a given ward (at midnight) is likely to be a key metric in evaluating the quality of any scenario the user might wish to simulate, a method which affords simple calculation is preferable. For this reason, an uncapacitated approach has been chosen to model the wards in the AGH.

4.3.4 Decision Variables (Elective Schedule)

As argued in Chapter 2, one of the most important operational decisions a general hospital must make is the number of elective admissions to schedule. Admitting too many elective patients limits the capacity to treat emergency patients, potentially causing blockages at the ED. On the other hand, scheduling too few elective admissions can leave patients on waiting lists unnecessarily (in the case of public health services) and is an inefficient use of expensive resources such as nurses and theatre time.

Given the impact that the elective admissions schedule has on the operational management of inpatient beds, and the potential complexities involved with balancing both elective and emergency bed demand, it stands to reason that a model capable of quantifying the impact of competing schedules is likely to be of value. For this reason, elective admissions occur deterministically by fixing the day and ward of arrival for each elective patient in the scheduled planning horizon. This information can either be read from the historic schedule occurring in the PA data or from a modified schedule supplied by the user. User-define schedules allow the relative effect on the midnight census to be investigated for a set of plausible alternatives.

4.3.5 Uncontrolled Variables

While hospitals can exercise a degree of control over the elective admissions schedule, there are other factors influencing bed demand which cannot be fully controlled, such as the arrival of emergency patients or the variability associated with patient recovery times. Because of these uncontrollable factors, hospital planners endeavour to schedule elective admissions in such a way that allowances are made for unforeseen events, to mitigate the under or over-utilisation of beds. To adequately represent the uncertainty associated with the management of inpatient beds, three factors impacting ward-level bed occupancy are treated as random variables in the simulation:

1. Number of emergency admissions per day
2. Number of midnights spent on each ward by each patient (ward length of stay)

3. Location of the next ward stay (or discharge) once the current ward stay is complete.

In contrast to the elective admissions which occur deterministically, the number of emergency admissions per day is treated as a random variable, in keeping with the random nature of emergency arrivals in a real hospital. Since the simulation runs in discrete time, it is not necessary to model the inter-arrival times of the emergency patients. Instead a random number of daily arrivals can be sampled from the history of daily arrival numbers, and this determines the size of the “batch” of emergencies occurring on any given day.

After being admitted to the hospital, patients occupy a bed for some period before being discharged or transferred to a bed on another ward. This period is known as the patient’s *ward* length-of-stay. In the case of elective admissions, it may be well estimated by clinicians and planners responsible for scheduling procedures (investigated further in Chapter 6). However, it is not unreasonable to expect variation in length-of-stay from patient to patient. For emergency patients arriving at the hospital, length-of-stay may be even less predictable due to the unscheduled nature of their admission. For this reason, ward lengths-of-stay are modelled as random variables whose discrete distributions (since the simulation runs in discrete time) are derived from the PA data.

Once a patient’s ward stay is over, they may be discharged from hospital or they can be transferred to another ward. If a transfer to another ward is necessary, the choice of ward is not only dependent on the patient’s clinical requirements, but also the availability of the resources needed to treat the patient, such as beds, nurses and monitoring equipment. Ideally, a patient’s path through the network of hospital wards would only depend on their clinical requirements; however, this is not always the case. The stochastic nature of demand for some resources (caused at least in part by the unplanned nature of emergency arrivals) means that unforeseeable circumstances can arise in which a patient's ideal ward is at full capacity, and a transfer is not possible. Patients' clinical requirements can also change unexpectedly; due to the worsening of existing conditions, or even because of their proximity to other ill patients. The potential for uncertainty in the sequence of visited wards, along with the unknown types of emergency arrivals occurring in each planning horizon, justify the use of stochastic transfers between wards in the simulation. Although random, the probabilities which govern these transfers can be estimated using PA data, thereby maintaining the average patient flows seen in the real hospital. Stochastic methods for modelling ward transfers are discussed further in Section 4.4.4.

4.3.6 Conceptual Model Diagram

The conceptual model described in the preceding sections is summarised in Figure 4.1, which represents the first step in meeting the first requirement of an online simulation described in Hanisch et al. (2005). The level of detail of the conceptual model and the way in which components are to be implemented are focused by the structures and processes of interest in the real system. For operational bed management purposes, the structures and processes of interest are those related to the ward-level midnight census. With a conceptual model developed, implementation and parameterisation of this model in the Micro Saint Sharp software package can begin.

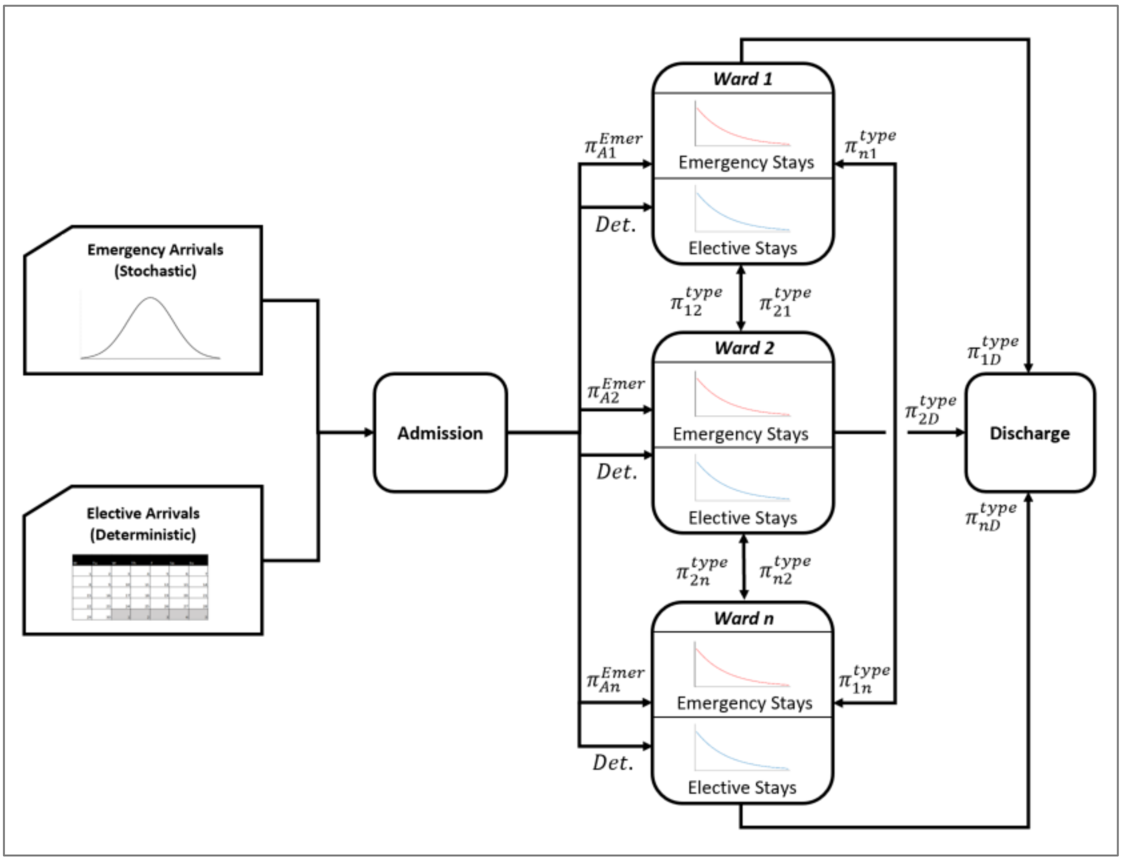


Figure 4.1: Schematic of the conceptual model to be implemented in Micro Saint Sharp simulation software. The $\pi_{i,j}^{type}$ represent the transition probabilities from ward i to j for each admission type (emergency/elective). Probabilities will not be designated for routing elective patients from the admission node, since the first ward-stay is assigned deterministically by the elective admissions schedule. No simulated time is spent at the admission and discharge nodes.

4.4 Offline Model Development

In this section, the conceptual modelling decisions from Section 4.3 are implemented in Micro Saint Sharp, and the details of parameterising the model with the PA data supplied by the AGH are described. The result is considered to be an “offline” model of the inpatient ward network, in the sense that it is always initialised with an empty and idle state, meaning that at this stage of development, it is not suitable for use in an operational decision-making context. This model does however, form the basis of the eventual “online” model, and can be used to ensure that no unexpected long-run behaviours are exhibited in the simulation outputs. For example, unexpected behaviour such as a trend in the midnight occupancy time-series (when average arrival rates are not known to be increasing or decreasing) may be difficult to detect in the comparatively short runs of the online model but would be readily seen in the time-series of midnight occupancies generated by the offline model.

4.4.1 Modelled Wards

As has already been mentioned in Section 4.3.2, modelling every ward which appears in the PA data is not sensible since not all wards are geared towards accommodating patients for overnight stays. For this reason, a pragmatic approach towards determining the structure of the ward network was chosen, where wards whose midnight census was consistently low were aggregated together to form an “Other” ward. By doing this, modelling effort is focused on the wards which have the greatest impact on the hospital-wide midnight census, while also maintaining the ability to make inferences about the number of beds occupied by *all* emergency and elective patients.

An entry criterion is used to determine which wards are modelled based on the average midnight occupancy for each ward over the period of available PA data. The average occupancies are ranked high to low, and the wards which comprise 90% of total midnight occupancy for emergency/elective patients are modelled with their own node in the simulation network. The wards which make up the remaining proportion of average midnight occupancy are aggregated together to form the “Other” ward. Table 4.1 shows the 20 wards in the PA data where overnight stays occur, along with their percent average contribution to hospital-wide emergency/elective midnight occupancy. The set of possible wards is reduced to 10 modelled wards (including “Other”) once the entry criterion is applied. Figure 4.2 shows the resulting ward network which is implemented in Micro Saint Sharp.

It is important to note that since the Other ward represents several low occupancy wards (for emergency/elective patients), it can be visited reflexively. Therefore, patients leaving Other can begin a stay on Other immediately afterwards, unlike the individually modelled wards. These reflexive transfers represent patient transitions within the lower occupancy wards. The length of stay distributions for Other are derived by pooling all LOS data for the shaded wards in Table 4.1.

Wards (PA Data)	Average Midnight Census	% of Total Hospital Average	Cumulative %
Ward 5D	28.90	16%	15.7%
Ward 5B	28.52	15%	31.1%
Ward 4D	26.00	14%	45.2%
Ward 5A	25.79	14%	59.2%
Ward 6D	25.10	14%	72.8%
Ward 4K	11.93	6%	79.3%
Ward Northside	11.79	6%	85.6%
Intensive Care Unit	7.84	4%	89.9%
ED	7.31	4%	93.9%
Ward 3R	5.29	3%	96.7%
Hospital in the Home	3.31	2%	98.5%
Ward 4O	1.13	1%	99.1%
Ward 4N	0.94	1%	99.6%
The Manor Transitional Unit	0.33	0%	99.8%
Theatre Main	0.14	0%	99.9%
Day Procedure Unit	0.09	0%	99.9%
Ward 4B	0.06	0%	100.0%
Renal Unit	0.02	0%	100.0%
Theatre Angio	0.02	0%	100.0%
Renal Unit - North West	0.01	0%	100.0%
Total	**184.51**		

Table 4.1: Average midnight census for emergency and elective patients combined. Shaded wards are represented by the "Other" ward in the simulation.

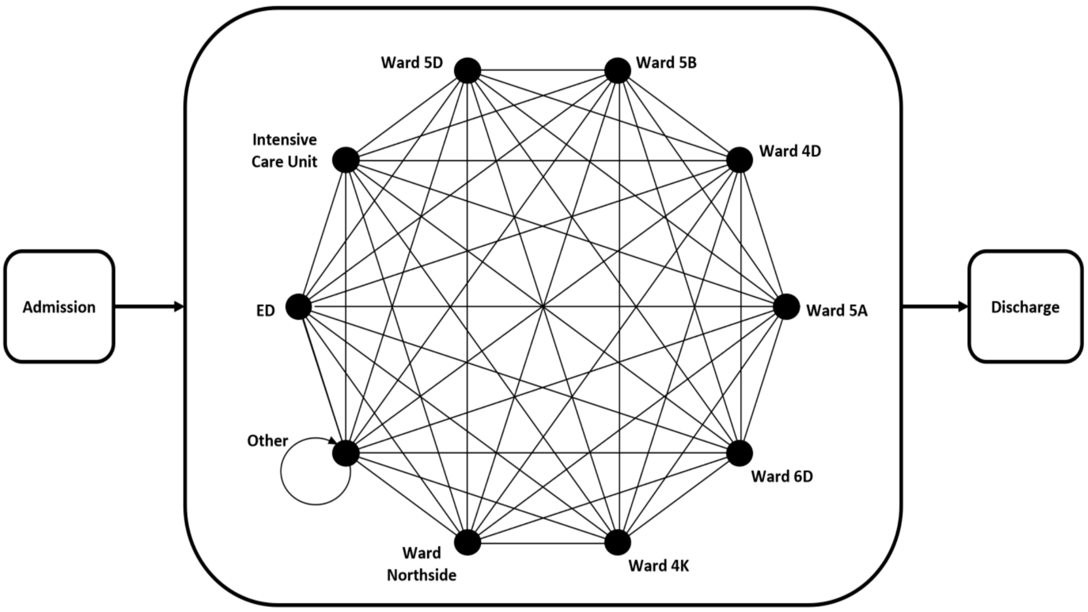


Figure 4.2: The network of ten individually modelled wards implemented in Micro Saint Sharp simulation software forms a complete graph.

4.4.2 Modelling Arrivals

As discussed in Section 4.3.5, the number of emergency patients arriving each day is modelled as a random variable, and because the simulation runs in discrete time, the time between successive emergency arrivals is not important. Therefore, a discrete empirical distribution derived from the PA data can be used to generate the number of emergency arrivals on a daily basis.

As might be expected, not all days of the week are likely to have the same average number of emergency arrivals, and if a "day-of-the-week" effect exists in the pattern of emergency arrivals, it is likely to have some impact on the day-to-day management of inpatient beds. In Figure 4.3 the PA data has been used to investigate the likelihood of a day-of-the-week (DOW) dependent pattern of emergency arrivals.

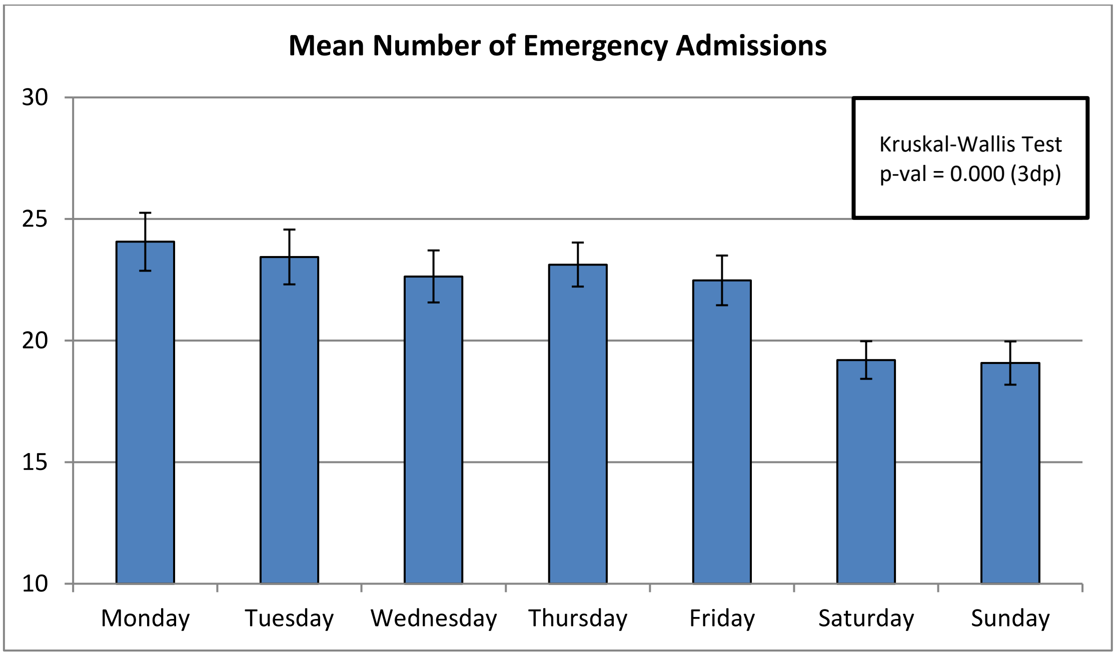


Figure 4.3: Mean number of emergency admissions to the AGH for the period in which PA data is available. Error bars are 95% confidence intervals for the means. K-W has been used since homogeneity of variance across weekday groups is not assumed.

The mean number of emergency admissions appears to decrease during the weekend, supporting the notion that a DOW pattern exists for the emergency arrivals. The Kruskal-Wallis Test has been used to test the null hypothesis that the distribution of the number of daily arrivals is not dependent on DOW; a hypothesis which is rejected at a significance level of 0.001.

Since the relationship between the distribution of daily emergency arrivals and day of arrival can be shown to be statistically significant using the PA data (and is likely to have an impact on operational bed management), this pattern has been included in the simulation. The number of daily emergency arrivals is drawn from one of seven discrete empirical distributions derived from arrivals recorded in the PA data on each day of the week.

In contrast to the randomly generated emergency arrivals, the elective admissions occur deterministically, which can be used to test the effect of competing elective schedules on bed demand for the planning horizon in question. For validating the offline model, the simulation uses 560 days of actual elective admissions. To run this observed schedule through the simulation, the number of elective patients arriving on each ward, during each day of observation is taken from the PA data.

4.4.3 Modelling Length of Stay

Once a patient has been assigned to a ward bed (by either admission or transfer from another ward), the bed is occupied for some period of time before the patient is either discharged or transferred to another ward. The amount of time a patient occupies a bed is known as the patient's length-of-stay (LOS) and this can potentially be influenced by a number of factors. The most obvious of these is the type and severity of a patient's condition, but others might include bed blocking, (the patient is required to wait until space on a more suitable ward becomes available) or the hospital's ability to discharge the patient (in cases where the patient requires help leaving the hospital).

Naturally, not all factors which influence a patient's LOS are recorded in the PA data, and even if very detailed information was available, it would not rule out uncertainty in LOS altogether. For this reason, LOS for both the emergency and elective patients are treated as discrete random variables which represent the number of midnights a patient will stay on a particular ward.

While it may be unreasonable to expect that detailed LOS information (such as delays caused by bed blocking) is recorded in the PA data, other standard information is available. As has already been mentioned, the PA data contains patient stay records with specialty, ward and admission type information along with the times at which the status of any of these changes. This means the pool of patient level LOS records can be disaggregated to form samples of similar patients from which LOS distributions can be derived.

Since the purpose of the model is to provide estimates of ward-level bed demand, the pool of all LOS observations is first disaggregated by ward. This also goes some way towards grouping patients of similar specialty, although it is not uncommon for wards to provide beds for multiple specialties. The next disaggregation occurs at the admission type level, meaning LOS for elective and emergency patients on the same ward will be drawn from distinct distributions. Further disaggregation by specialty is theoretically possible by using the specialty categorisation in the PA data; however, each level of disaggregation has the effect of diminishing the sample size of each patient group, making any statistical inference (such as analysing the relationship between midnight occupancy and transition probability) less meaningful. Therefore, discrete empirical LOS distributions for the number of midnights spent on each ward are derived at the ward and admission type (emergency/elective) level.

As has already been shown in Section 4.4.2, it is likely that some day-of-the-week dependent effect exists for the distribution of emergency arrivals. In a similar way, it might be reasonable to expect that a patient's LOS is also affected by the day of the week on which he or she is admitted. Such an analysis has already been carried out by the UK Audit Commission (Audit Commission, 2003) in which LOS records generated by NHS Trusts across England and Wales were grouped by weekday of admission. The Audit Commission found that patients admitted on a Thursday stayed in hospital for a significantly longer period than patients admitted on any other day of the week, citing the reduced availability of support and diagnostic departments, along with reduced numbers of senior staff capable of making discharge decisions over the weekend as likely causes.

If evidence suggests that a relationship exists between LOS and weekday of admission in the AGH being modelled, then capturing this relationship has the potential to improve bed demand estimates. In Figure 4.4, the Total LOS observations (that is, the amount of time spent as an inpatient from admission to discharge, regardless of ward placement) are grouped by weekday of admission to investigate the likelihood of such a relationship.

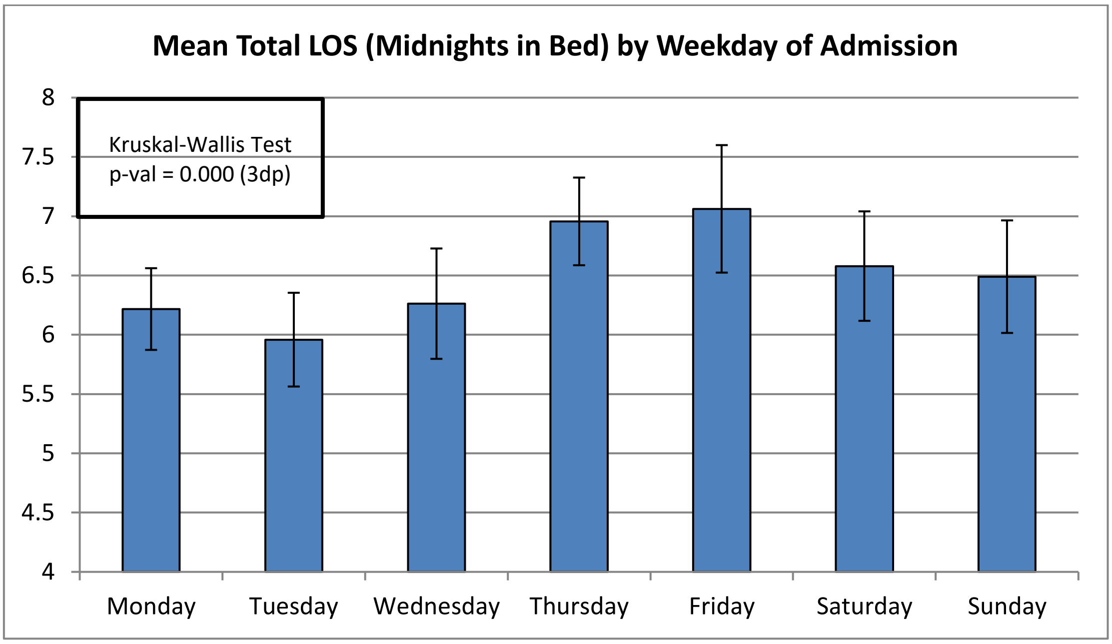


Figure 4.4: Mean Total LOS for emergency and elective patients in the observation period. Error bars are 95% confidence intervals for the means. K-W testing has been used since homogeneity of variance across weekday groups is not assumed.

On average, Total LOS appears to be greatest for patients admitted on a Friday, closely followed by those admitted on a Thursday, and these patients stay in hospital for approximately one day more than those arriving on a Monday or a Tuesday. The weekly LOS pattern is similar to the pattern reported by the Audit Commission (2003), indicating that the AGH may also suffer from a lack of resources and staff over the weekend, resulting in a decreased rate of discharge. The Kruskal-Wallis test has been used to test the null hypothesis that the distribution of LOS is not dependent on weekday of admission, and this hypothesis is rejected with a significance level of 0.0005.

Because the relationship between weekday of admission and the distribution of total LOS is statistically significant and likely to have an impact on midnight bed occupancy numbers, it makes sense to model this relationship in the simulation. However, it is worth noting that although the analysis of this relationship was based on Total LOS (all wards) and both admission types, LOS in the simulation is modelled at a lower level of detail. Rather than carrying out the analysis for each of the ward and admission type combinations, the findings of this pooled analysis are treated as being valid for all levels of detail, since disaggregation would result in smaller sample sizes and potentially insufficient statistical power on which to base a conclusion.

With 10 modelled wards, 2 admission types and 7 possible admission days, each ward LOS can be drawn from one of 140 possible LOS distributions. The LOS generation process has been simplified by assuming that each ward LOS draw is independent of all other simulated patients and any time spent on other wards. It should also be noted that although elective (planned) arrivals are treated as deterministic, their LOS is treated as a random variable at this stage of model development.

4.4.4 Modelling Ward Transitions

Given the potential complexity and uncertainty associated with assigning patients to wards (Section 4.3.5), an algorithm which accounts for *all* the factors considered by hospital staff cannot be obtained based on the information commonly contained within PA databases. However, a number of methods exist in the literature for approximating patient routing behaviour stochastically, based on data obtained from the hospital being modelled. For example, Chow et al. (2011) employ a so-called "trace-driven" approach, in which entire patient pathways (including lengths-of-stay) are sampled for each simulated patient from a database of observed hospital stays. Gallivan and Utley (2005) and Helm and Van Oyen (2014) derive "persistence matrices" from available data, which return the probability of being on a particular ward given the amount of time the patient has already spent in hospital. Günal (2008) computes "transition matrices" which contain the estimated probability of transitioning between any two wards, based patient transitions observed in the hospital data.

While persistence matrices treat patient routing as a function of time, and transition matrices treat patient routing as a function of location, it is not difficult to think of other factors which might influence routing decisions in a real hospital. For example, it is expected that transferring a patient to a ward with no available beds should be less likely than transferring the patient to a ward on which beds are available. For this reason, a preferred ward routing policy is one which can be generalised to respond to other factors which might influence patient transitions.

In this chapter, fixed or "static" transition matrices (STMs) will be used in the simulation model under consideration. This method captures not only the potential uncertainty associated with a given patient's path through the hospital network, but also the average rate of transition between any two wards. The term "static" is used here to indicate that the probability of transitioning between any two wards is estimated independently of time or any other variable which might influence the likelihood of transitioning between wards. However, transition matrices *are* expected to be able to be formulated as functions of other factors which influence patient routing, if necessary. This is because a snapshot of the hospital can be taken from the PA data at the time of any recorded transition. Persistence matrices on the other hand, only incorporate ward location information at pre-determined times after admission, meaning any information relating to what might have caused a particular transition is lost. Patient routes generated by a trace-driven approach, clearly cannot respond to the state of the modelled system, because the entire ward-stay trajectory is sampled when the simulation entity is created.

As shown in Figure 4.2, the modelled wards form a complete graph, meaning each ward is connected to every other ward once implemented in Micro Saint Sharp. The STMs govern the likelihood of a patient transitioning to other wards once the LOS on their current ward has ended. For newly arriving emergency patients, a set of entry transition probabilities govern which ward a patient is admitted to on arrival. For the elective patients, the ward of admission and weekday of arrival are considered to be part of the elective schedule; therefore, it is not necessary to estimate the probability of arriving on a particular ward for this admission type.

While all the modelled wards are connected to one another in Micro Saint Sharp, the probability is allowed to be zero if there is no evidence of it occurring in the PA data; effectively disconnecting the two wards. The transition probabilities are estimated from the PA data in the following way:

$$\hat{\pi}_{i,j} = \frac{n_{i,j}}{\sum_{k=1}^{w+1} n_{i,k}} \text{ for } i,j \in \{1, \dots, w\} \qquad \text{(Eq. 4.1)}$$

$\hat{\pi}_{i,j}$ is the estimated probability of transitioning from ward i to j, while w is the number of modelled wards. The $(w+1)$th ward corresponds to being discharged from the hospital, rather than being a modelled ward in its own right. The $n_{i,j}$ represent the number of observed transitions from i to j in the PA data. Bed transfers within the same ward are modelled as a single LOS period, therefore $\hat{\pi}_{i,i} = 0$ for all wards except Other, where reflexive transfers represent transfers between the smaller wards of which Other is composed. The $\hat{\pi}_{i,j}$ are calculated for each of the two admission types (emergency and elective), and these form two separate transition matrices to account for the likely differences in pathways through the hospital for the two patient groups.

The STMs shown in Tables 4.2 and 4.3 are used in the offline model, and the outcome of patient transfers from "row" wards to "column" wards are drawn from these distributions. Since reflexive transfers are not considered to be ward transfers, entries where the row and column wards match are zeroed. Entry and Exit are dummy wards which cannot be revisited, therefore they only occur as row and column wards respectively. It is also worth noting that the entry row of the transition matrix does not apply to elective patients, since their first ward is decided by the elective admissions schedule and is therefore deterministic.

While STMs of the type shown in Tables 4.2 and 4.3 are fixed with respect to other factors which might influence patient placement decisions, it has already been mentioned that matrices of constant probabilities such as these might be generalisable to matrices of functions of the state of the hospital. This is made possible because draws from transition matrices (static or otherwise) occur at the same time as the simulated transition occurs, meaning it is possible for the state of the hospital to influence the probability of transition at that time. This concept is explored further in Chapter 5, in which relationships between ward level occupancy and transition probability are sought from the PA data to emulate the effect of outliers in the simulation. The outputs generated by this simulation are compared against the outputs generated by the simulation using STMs in this chapter to assess the relative effect of modelling transition probability as a function of occupancy.

	ED	ICU	Ward 4D	Ward 4K	Ward 5A	Ward 5B	Ward 5D	Ward 6D	Northside	Other	Exit
Entry	86.4%	0.4%	0.9%	1.4%	0.7%	0.8%	2.3%	0.5%	4.8%	2.0%	0.0%
ED	0.0%	2.9%	16.6%	15.6%	14.1%	15.6%	10.7%	8.4%	0.1%	11.7%	4.2%
ICU	0.8%	0.0%	11.6%	3.5%	21.1%	8.9%	7.3%	9.8%	0.0%	5.5%	31.5%
Ward 4D	1.2%	1.7%	0.0%	0.0%	2.3%	1.8%	4.9%	5.7%	0.0%	18.1%	64.4%
Ward 4K	0.2%	0.9%	0.1%	0.0%	0.3%	0.6%	0.1%	0.0%	0.0%	0.9%	97.0%
Ward 5A	0.7%	4.7%	3.0%	0.0%	0.0%	12.4%	2.9%	2.0%	0.0%	4.8%	69.5%
Ward 5B	0.8%	2.7%	1.5%	0.4%	1.7%	0.0%	1.4%	1.6%	0.0%	3.7%	86.2%
Ward 5D	0.6%	2.0%	2.4%	0.0%	1.1%	0.7%	0.0%	2.2%	0.1%	4.2%	86.9%
Ward 6D	0.9%	2.3%	1.9%	0.0%	0.8%	0.3%	2.4%	0.0%	0.0%	4.5%	86.8%
Northside	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	1.0%	99.0%
Other	1.6%	6.6%	17.2%	1.8%	7.3%	6.2%	4.4%	5.5%	0.2%	11.0%	38.1%

Table 4.2: The Static Transition Matrix estimated for the emergency patients.

	ED	ICU	Ward 4D	Ward 4K	Ward 5A	Ward 5B	Ward 5D	Ward 6D	Northside	Other	Exit
Entry	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
ED	0.0%	2.8%	17.6%	10.2%	15.7%	11.1%	12.0%	7.4%	0.0%	14.8%	8.3%
ICU	0.0%	0.0%	2.4%	0.7%	54.9%	21.4%	0.7%	0.3%	0.0%	2.7%	16.9%
Ward 4D	0.0%	1.6%	0.0%	0.0%	0.5%	0.5%	3.2%	1.6%	0.0%	26.6%	66.0%
Ward 4K	0.0%	0.5%	0.0%	0.0%	0.0%	0.4%	0.2%	0.0%	0.0%	2.2%	96.8%
Ward 5A	0.0%	6.1%	0.6%	0.1%	0.0%	3.4%	0.2%	0.1%	0.0%	3.8%	85.8%
Ward 5B	0.0%	2.0%	0.1%	0.1%	1.3%	0.0%	0.2%	0.1%	0.0%	2.1%	94.2%
Ward 5D	0.0%	0.4%	0.2%	0.0%	0.0%	0.2%	0.0%	0.4%	0.0%	4.3%	94.4%
Ward 6D	0.0%	0.0%	0.0%	0.0%	4.5%	0.9%	7.3%	0.0%	0.0%	6.4%	80.9%
Northside	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	100.0%
Other	0.3%	6.2%	3.9%	1.4%	22.4%	49.5%	2.7%	2.2%	0.0%	4.0%	7.5%

Table 4.3: The Static Transition Matrix estimated for the elective patients.

4.5 Offline Model Validation

The term “validation” used in this section refers to the “black-box” type of analysis defined in Pidd (2009) in which the model outputs are compared to some historic data generated by the real system in order to confirm that the model displays similar performance characteristics when run under similar operating conditions. This type of validation is distinct from the “open-box” validation conducted in the conceptual modelling phase, where subject matter experts are consulted (where possible) to ensure that the structures and processes in the simulation constitute a suitable representation of the real system.

The model components already described in Sections 4.4.1 to 4.4.4 form the offline basis of the eventual online mode, once combined. While the creation of an *offline* model can be seen as a single step in the *online* model development process, it is important to note that the offline model should also be a good representation of the inpatient ward network in the hospital being modelled, albeit over a longer period of time. Significant differences between statistics which summarise longer periods could help to diagnose problems with the modelling assumptions or component parameterisations which may be difficult to detect in the comparatively short runs of an online simulation. On the other hand, if no significant differences are found, this indicates that the offline model is performing as expected, on average, thereby contributing towards meeting the first online simulation requirement described by Hanisch et al. (2005).

4.5.1 Run Configuration

While the details of each of the offline modelling elements (such as arrival pattern, ward lengths of stay and transitions) have been discussed earlier in Section 4.4 (Offline Model Development), Table 4.4 provides a summary of the implementation of these elements for the model being checked for validity. Their implementation, together with the chosen values of the decision variables form the configuration of the model at run-time.

	Treatment	
Offline Modelling Element	*Emergency*	*Elective*
Arrivals/Admissions	Empirical Distributions (Stochastic)	Observed Schedule (Deterministic)
Ward Length of Stay	Empirical Distributions (Stochastic)	Empirical Distributions (Stochastic)
Ward Transitions	Static Transitions (Stochastic)	Static Transitions (Stochastic)

Table 4.4: Treatment of each of the major modelling elements in the offline simulation, grouped by admission type.

As with any model validation process, the values chosen for the decisions variables should be the same as those which generated the observed data in order to carry out a fair comparison. The decision variable for this model is the elective schedule, therefore the same schedule which contributed towards the observed values of the midnight occupancy censes should also be run in the offline model. Since the lowest temporal resolution of the simulation is “daily”, the elective schedule which is executed consists of the observed number of elective patients to be admitted to each ward, each day.

Because the PA database contains 560 days of useable data, a 560-day elective admissions schedule can be continuously simulated. If the number of daily elective admissions had been drawn from empirical distributions in the same way that the emergency admissions are generated, there would be no limit on the number of days which could be simulated; however, this would also mean that the elective admissions schedule could no longer be treated as a decision variable once the offline simulation is brought online.

4.5.2 The Warm-Up Period

The notion of a warm-up period has little meaning in the context of online simulation since online models are designed to be initialised to an operational state by querying the state of the real system. However, the offline model discussed in this section contains no such "state-matching" component; each simulation run is initialised with an empty-and-idle state. This being the case, the output data generated by the offline model will contain warm-up periods at the start of each run, and this data will bias estimates of the midnight census if it is included in the analysis.

With 100 simulation runs, each midnight has 100 simulated occupancy observations with which the distribution of midnight occupancy can be approximated. The median, along with 5^{th} and 95^{th} percentiles generated for each midnight have been plotted in Figure 4.5 to provide an indication of how the distribution of ward occupancy for all wards and both admission types evolves from an empty and idle state. Visual inspection of the time series suggests that the model has a warm-up period of approximately five simulated weeks, therefore in all further analyses of the offline model the first 5 weeks of data is discarded to keep the warm-up period from biasing the results.

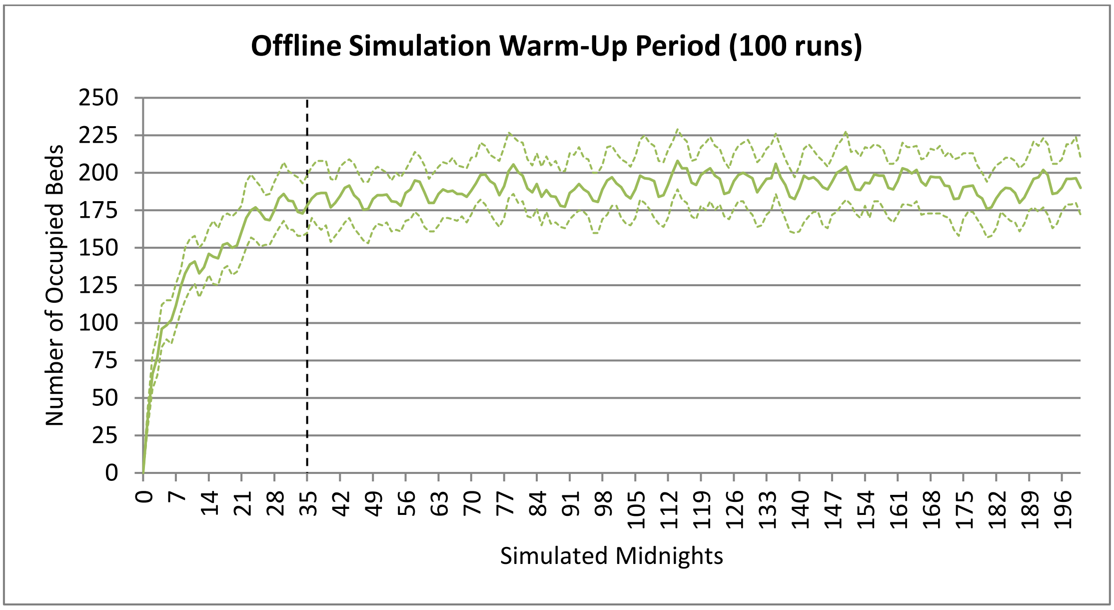


Figure 4.5: The median midnight census for 100 runs of the first 200 simulated days, along with 5th and 95th percentiles. Observations to the left of the dashed line will be treated as occurring within the warm-up period.

Naturally, the time series of actual midnight occupancy does not exhibit a warm-up period, since it is not generated by a simulation model. This means that once the warm-up period is removed from simulation data, a 35-day discrepancy exists between the simulated elective pattern and the observed elective pattern. In order compare sets of midnight occupancy observations generated (in part) from the same elective admissions pattern, the first 35 days have been excluded from the time series of actual midnight occupancy in all further checks of offline model validity.

4.5.3 Mean Midnight Occupancy

To identify bias or any other problems within the offline simulation, mean midnight occupancy over the observation period can be estimated from the PA data, and subsequently compared with mean midnight occupancy generated by the offline simulation over the same number of days. Since it is expected that weekday-dependent behaviour is associated with both the emergency arrival pattern and the length-of-stay of all patients, it makes sense to carry out such a comparison on a day-of-the-week basis. The bar charts in Figures 4.6 to 4.8 compare the observed mean midnight occupancy (by day of the week) with realisations of mean midnight occupancy derived from the simulation outputs for emergency and elective patients, both separately and combined. The error bars within the bar charts are two-tailed 90% *prediction intervals* for the mean midnight occupancy derived from the simulation. Since the observed mean midnight occupancies are fixed for this extract of the PA database, it makes more sense to consider variation coming from the simulation. The use of prediction intervals as opposed to *confidence intervals* is deliberate, since the bounds are derived empirically, through repeated simulation, rather than parametrically. The same can be said of the 5th and 95th percentiles of simulated midnight occupancy added to the observed midnight occupancy time series. The median (50th percentile) has also been plotted for an indication of central tendency.

Figure 4.6a compares mean midnight occupancy observed in the PA data with the prediction intervals constructed from the simulation outputs for the emergency patients only. The hypothesis that the mean emergency occupancies observed in the PA data for each day of the week come from the empirical distributions of mean occupancy from the simulation would not be rejected for any day of the week, at the 10% significance level.

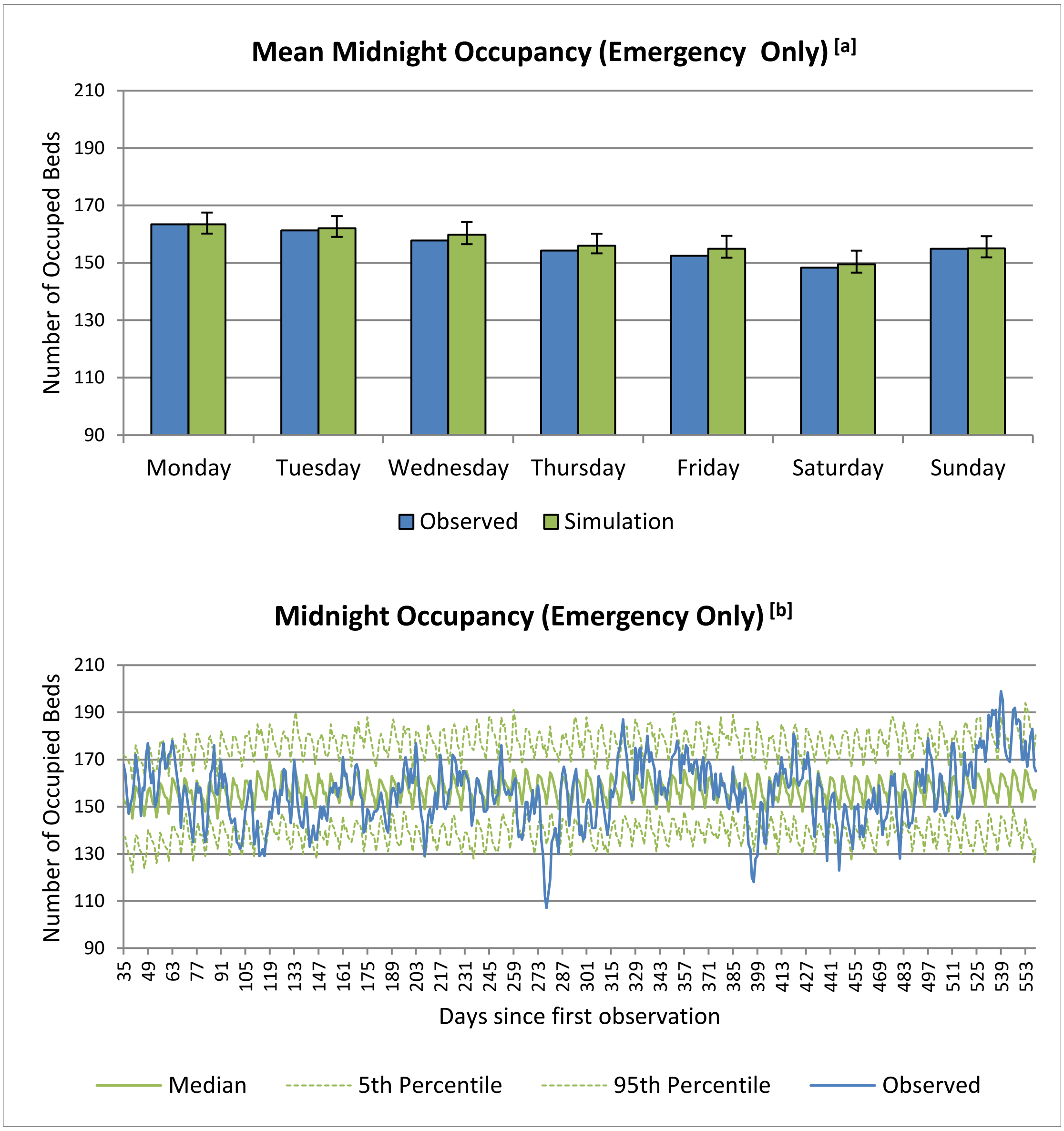


Figure 4.6: [a] Mean midnight bed occupancy by weekday. [b] Time series comparison of observed and simulated midnight occupancy for the emergency patients.

However, observed mean occupancy for the emergency patients appears to tend toward the lower end of the prediction intervals for the days later in the week. This is caused in part by the decreased number of emergency admissions around Christmas 2010, shown by the trough in Figure 4.6b around day 280. A sudden change in the arrival pattern such as this would not be mimicked by the emergency arrival pattern in the offline model, thereby contributing to the discrepancy.

Despite the drop in emergency admissions around Christmas, the midnight census time series for the emergency patients appears to sit within the 90% prediction intervals generated by the offline most of the time. In fact, for this set of 100 simulation runs, 88.3% of observed midnight occupancies fall within the corresponding prediction interval generated by the offline model, indicating that the prediction intervals are performing as expected.

Figure 4.7 compares the simulated and observed midnight census for the *elective* patients in the same way. Figure 4.7a shows that the difference between simulated and observed mean midnight occupancy for each day of the week is negligible for the elective patients, and the hypothesis that the observed mean occupancies come from the empirical distributions derived from the simulation is not rejected for any day of the week. The reason for this close agreement becomes clear when looking at Figure 4.7b, which shows that the time series of 90% prediction intervals for the elective midnight census is much narrower than the time series of the same prediction intervals generated for the emergency series in Figure 4.6b. This close agreement can be attributed to the fact that the same pattern of elective admissions is present in both the simulation and the PA data, and the stochastic elements of elective stays in the simulation come only from ward length of stay and ward transitions. Despite narrowing, they continue to perform as expected, with 90.9% of elective census observations falling within their corresponding 90% prediction interval.

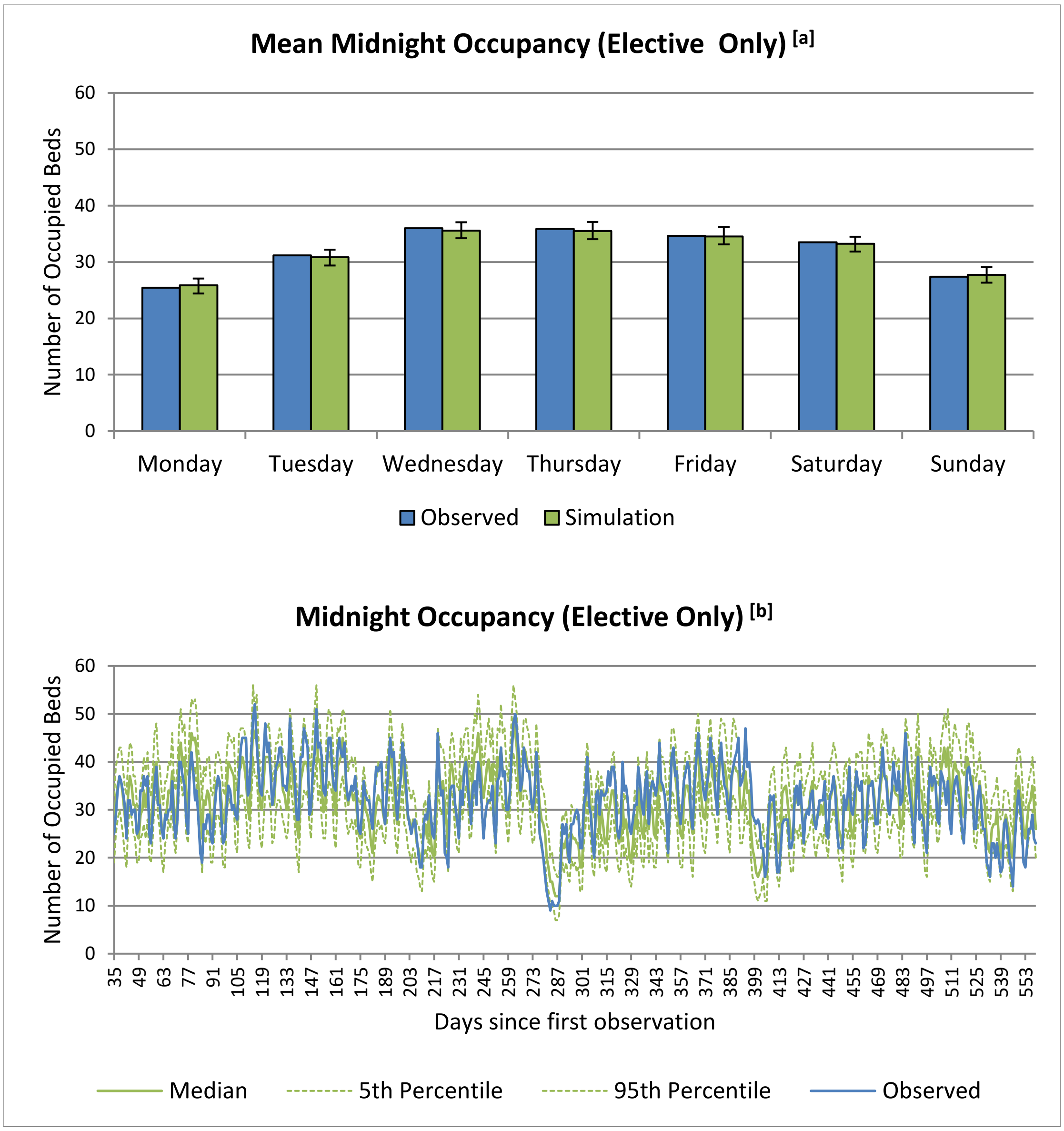


Figure 4.7: [a] Mean midnight bed occupancy by weekday. [b] Time series comparison of observed and simulated midnight occupancy for the elective patients.

Finally, Figure 4.8 compares the midnight occupancy data for the emergency and elective patients in combination. These occupancy observations are the sums of their emergency and elective parts, and as such, are not expected to show any misspecification of the model given that their summands do not.

Looking at the time series in Figure 4.8b, the median of the midnight occupancy realisations generated on each day appears to follow some of the features of the observed series (such as the trough caused by a lowered admissions rate over the Christmas period), although not as closely as the simulated time series for elective occupancy follows its observed series. This is to be expected since the elective arrivals are generated deterministically as opposed to the emergency arrivals which occur randomly, and this effectively adds noise to the elective admissions pattern. For this reason, the simulated median occupancy time series for the combination of admission types shares a greater estimated correlation coefficient with its observed time series ($\hat{\rho}_{total}$=0.395) than the emergency time series does ($\hat{\rho}_{emer}$=0.331), although it is not as highly correlated as the median occupancy time series for elective patients is with its corresponding observed series ($\hat{\rho}_{elec}$=0.832).

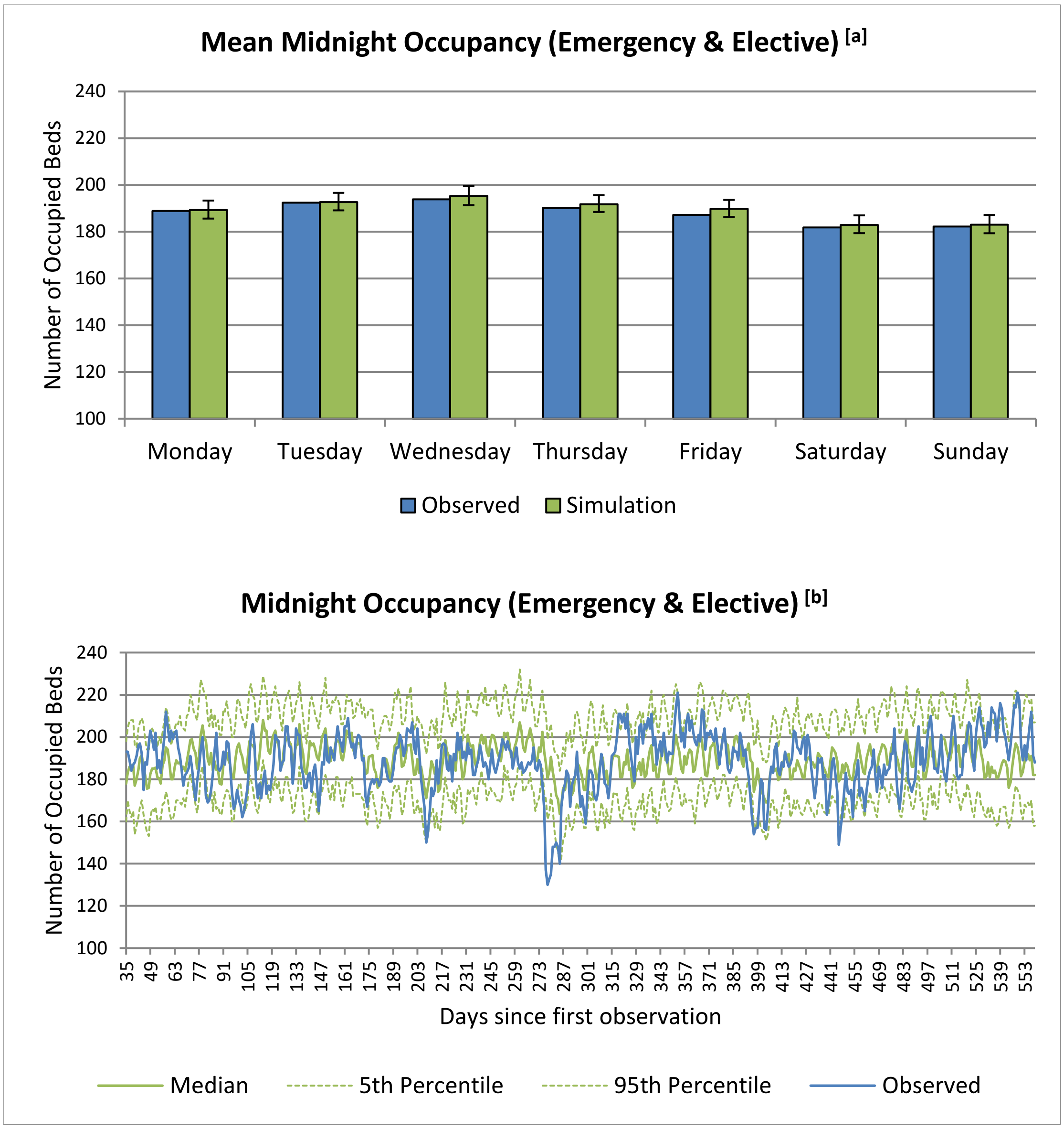


Figure 4.8: [a] Mean midnight bed occupancy by weekday. [b] Time series comparison of observed and simulated midnight occupancy for both admission types.

4.5.4 Offline Validation Summary

The analyses carried out in this section show that at the 10% significance level, there doesn’t appear to be any misspecification of the offline model in terms of mean occupancy on each day of the week at the admission type level (emergency/elective). However, it is worth noting that Research Question 1 was posed in terms of the development of a model for the operational management of hospital beds at the *ward* level. Ward-level analyses have been conducted with emergency and elective patients pooled together, and these yield similar results to those already presented at the admission type level, that is; there is very little statistical evidence to suggest that action should be taken regarding the parameterisation of the offline model. Therefore, the bar charts containing the results of the ward-level analysis have been omitted in this section but are included in Appendix A, in Figures A.1.1 to A.1.10.

In summary, the offline model described in this section appears to behave as expected with respect to the tests of validity which have been conducted. While it is possible to consider other statistical tests which might be performed, tests of mean midnight occupancy for the whole observation period, are sufficient at this stage of model development. Further checks of model validity will also be carried out in Section 4.7, after the model is brought online.

4.6 Bringing the Model Online

With a validated offline model in hand, the process of augmenting the model to meet the second online simulation requirement can begin. Strictly speaking, a connection with a real, operational database is needed to meet this requirement, however setting up such a connection is not a trivial task, and may require the creation of customised software or queries in order to facilitate communication between the simulation software and the database. Instead, while the model is still in its testing and validation phase, it makes more sense to set up a connection with the readily available historic data, to generate estimates of midnight occupancy on each ward from historic system states.

In this section, the information and steps required to initialise the offline model with an operational state (at a commensurate level of detail) are described, thus bringing it online. A method for validating the online model is then proposed based on assumptions about the length of a useful planning horizon and the fact that the true state of the hospital can be queried retrospectively from the PA database extract at any time during the observation period.

4.6.1 System State Data

While the notion of a "system state" has been alluded to in this chapter, the information that such a state might contain has not yet been described. In general terms, when an online simulation is initialised, the state data which is loaded should completely describe the simulation at a point in time (including the state of each simulated server/node and each simulated entity) while also being a snapshot of the real system. Therefore, it is necessary to determine the information which completely determines the model's state and extract this information from the PA database. For the offline model described in Section 4.4, the state at any time during a simulation run can be described completely by obtaining the following information:

1. The number of emergency patients resident on each ward.
2. The number of elective patients resident on each ward.
3. The day of the week on which each patient was admitted to the ward they occupy.
4. The amount of time already spent on the current ward for each patient at the time the state data is collected.

The first and second pieces of information are likely to be the most obvious requirements when attempting to describe the state of the model. The third piece of information relates to the way in which ward length-of-stay is modelled. More specifically, a statistically significant relationship was found between the day of the week on which a patient was admitted, and the length of time they subsequently spent in hospital (see Section 4.4.3). Thus, day of admission information is required for each patient's ward length of stay to be drawn from the appropriate empirical distribution.

The fourth piece of information ensures that the patients who are resident on a ward when the state data is captured are loaded as simulation entities who have spent the same amount of time on the ward. It is worth noting that the state data only captures the time spent on the current ward for each resident patient, rather than the total time spent in the hospital (possibly) during previous ward stays. Capturing the total time spent in the hospital as part of the state data is not necessary, since each ward length of stay is sampled independently of all previous ward stays – a simplifying assumption made during the development of the offline model (see Section 4.4.3).

4.6.2 Conditional Length of Stay

In systems where each "job" has a known service time, any job currently in service when the simulation is initialised should be loaded into the model with its remaining time. However, ward length-of-stay in this model is treated as a random variable, therefore *remaining* length-of-stay is stochastic, and cannot be known at run-time. However, remaining LOS is likely to be dependent on the time already spent on the ward. The dependence between “time already spent” and “time remaining” necessitate the use of *conditional* LOS distributions, rather than sampling all patients as new admissions.

The conditional LOS distribution for each resident patient is straightforward to derive, given their time already spent on the ward, and the marginal distribution (which accounts for admission type, weekday of admission and hospital ward) applicable to the patient if they were a new arrival. Suppose the random variable T represents the total number of nights a given patient will spend on the ward, and that when the simulation is initialised, the patient has already been on the ward for s midnights. The random variable $R = T - s$ therefore represents the number of midnights the patient remains on the ward after the simulation is initialised. The CDF $F_T(t)$ is the empirical distribution from which LOS would be drawn if the patient had just arrived on the ward (1 of the possible 140 empirical distributions mentioned in Section 4.4.3). From this, the conditional CDF $F_T(t,s) = \mathbb{P}\{T \leq t | T \geq s\}$ can be obtained using the formula:

$$F_T(t,s) = \mathbb{P}\{T \leq t | T \geq s\} = \frac{F_T(t) - F_T(s-1)}{1 - F_T(s-1)} \qquad \text{(Eq. 4.2)}$$

Since R is the difference between T and s, the sampling distribution for R is readily given by:

$$F_R(r,s) = \frac{F_T(s+r) - F_T(s-1)}{1 - F_T(s-1)} \qquad \text{(Eq. 4.3)}$$

For a given s, realisations of R can then be drawn from $F_R(r,s)$ using the inverse sampling method, and these realisations represent remaining length of stay on the ward, given length of stay already spent on the ward, at the time the simulation is initialised.

It is worth noting that conditional length of stay does not need to be considered for models whose service times are best described using exponential distributions (when service times are continuous) or geometric distributions (when service times are discrete), due to the memoryless property. However, memorylessness does not extend to the empirical distributions from which LOS is drawn in this model.

4.7 Online Model Validation

With the requirements for the system state data specified, and a method in hand for generating ward-level length of stay realisations from conditional length of stay distributions, the offline model can be initialised with system states extracted from the historic PA database to investigate properties of the online model. This process goes some way towards meeting the second online simulation requirement described in Section 4.2, which calls for an online connection to the real system to be made available, although stops short of rolling out the system to an *operational* database, in favour of connecting to a *historic* database to conduct checks of online model validity.

The online validation methods discussed in this section are considered to be distinct from the so-called “auto-validation” modules within the online modelling framework proposed in Davis (1998). While both are concerned with the validity of online models, auto-validation is intended for use in an operational environment where problems with the model must be diagnosed and fixed very quickly; preferably before the next decision point. To do this in a timely way, both the checks of validity and scope of model adjustments are pre-defined within an auto-validation module. If the model is found to fail one or more of these checks, the pre-defined adjustments are carried out automatically. For example, Hill et al. (2001) re-fit the statistical distributions which govern the processes within their simulation based on newly available data from the real system. The re-fitting can occur automatically, or the model can prompt the user to decide whether re-fitting is necessary. The possibility of re-fitting the distributions occurs when the model's predictions begin to deviate significantly from observations of the real system.

While auto-validation is likely to play an important part in the ongoing maintenance of an online simulation, it is not the focus of this research. The validation in this section is instead focused on developing methods for comparing the model to the historic data in an "online way", prior to any connection being made with the real system. This is motivated by the change that occurs in the dependence structure of the simulation outputs when an offline model is brought online. Where statistical techniques employed for offline validation might treat realisations from a single run as being independent of initial conditions, simulation time, and possibly even each other for mathematical convenience, this independence should not be assumed for realisations from an online model which uses this information to inform its predictions. For these reasons, an online validation method should account for these dependencies where possible; treating each realisation of the performance indicators as a conditionally distributed random variable. While the validation methods of Hoot et al. (2008) go some way towards this goal, information about the full distribution of the simulation outputs is lost if only the mean of each set of replications is retained.

If no significant issues arise from the online validation process, the model can then be connected to the real system in such a way that information deemed relevant by the modeller can be transferred to the simulation. This could include system state data, or historic data to be used by an auto-validation module if required. As mentioned in Section 4.2, an online connection with the real system generally means that this information can be queried by the simulation model at any time, however in a hospital setting, it may not always be the case that a patient information database reflects the true state of the hospital. In reality, updates to patients’ electronic records are likely to be carried out by hospital staff when it is next convenient to do so rather than when changes occur, creating points in time when a hospital database might not be able to accurately relay state information. Nevertheless, the output statistics generated by an online simulation are necessarily dependent on accurate state information being input at run-time, therefore it is assumed that this is achievable at a point in time prior to the execution of a given elective admissions schedule for the model to be used in practice.

While the rate at which the operational PA database is synchronised with the true state of the hospital may pose challenges in a practical setting, it does not pose a challenge for an online validation process in which model is connected to the historical PA data instead of an operational database. It is assumed that the PA data is a correct account of the evolution of the state of the hospital during the observation period, therefore the state of the hospital *can* be queried at any time, retrospectively. For this reason, restrictions about the rate of synchronicity of the operational database will be imposed where it is necessary to illustrate properties of the model in a practical setting (or an approximation to it), although these restrictions can be dropped (since system state information is always available in retrospect) for the purpose of black-box validation.

4.7.1 Run Configuration

The configuration of the model used in this section is no different to that of the offline model used in Section 4.5.1, aside from the addition of loadable system states at run-time, and the ability to generate realisations from conditional LOS distributions. This configuration is summarised in Table 4.5.

	Treatment	
Online Modelling Element	*Emergency*	*Elective*
Arrivals/Admissions	Empirical Distributions (Stochastic)	Observed Schedule (Deterministic)
Ward Length of Stay	Cond. Emp. Distributions (Stochastic)	Cond. Emp. Distributions (Stochastic)
Ward Transitions	Static Transitions (Stochastic)	Static Transitions (Stochastic)

Table 4.5: Treatment of each of the major modelling elements in the online simulation, grouped by admission type.

4.7.2 Variation in Midnight Occupancy as a Function of Run Length

As has already been mentioned, the primary feature which distinguishes the online model from the offline model is the ability to match system states at particular points in time. As might be expected, this has an impact on the variance of the realisations of the midnight census since each replication returns the same midnight census value when the matching occurs, thereby setting the variance at that point in time to zero. Figure 4.9 illustrates this effect by comparing the prediction intervals generated by running the offline model for 100 replications (with the same run configuration) with the online model, also with 100 replications.

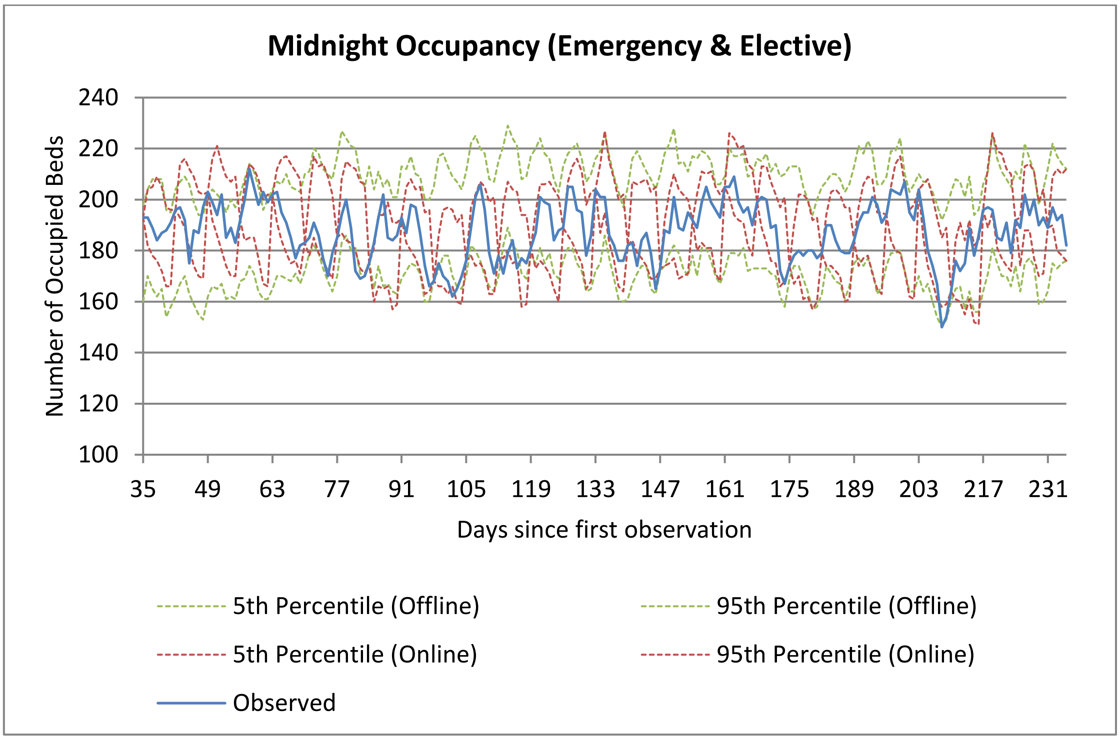


Figure 4.9: Comparison of the 90% prediction intervals generated from the offline and online models for all wards and both admission types.

In this simulation experiment, it is assumed that the PA database accurately reflects the true state of the hospital every Monday, and since the simulation runs in discrete time, the state matching must occur at midnight in conjunction with the midnight occupancy observations, meaning the online simulation is re-initialised every Monday at 00:00am in continuous time. This process is clearly seen in Figure 4.9 where the 5th and 95th percentiles (and all other possible percentiles) of midnight occupancy generated by the online model collapse to the observed midnight occupancy every seven days, since this is the only value generated by each replication of the simulation at these times.

As might be expected, the width of the prediction intervals generated by the online simulation appear to be narrower than those from the offline model, since the offline model is subject to random variation for t days, for any given time t after initialisation, while the online simulation is subject to random variation for a maximum of 6 days at a time, before it is re-initialised on the 7th day. This is more clearly illustrated in Figure 4.10, which shows the standard deviation of the midnight occupancy realisations over time for both the offline and online models. The standard deviation of midnight occupancy is lower for the sample of realisations generated by the online model, compared to the sample of realisations generated by the offline model for most simulated days. This is to be expected, since the online model is subjected to the sources of random variation in the system for fewer days at a time, however it is interesting to note that the levels of variation seen in the offline and online models are similar toward the end of the 7-day planning horizon. This suggests that predictions made by the online model beyond one week are unlikely to be influenced by the system state which is loaded at initialisation. Therefore, efforts to validate the model in an online way by considering only the first six midnights from initialisation, are reasonable.

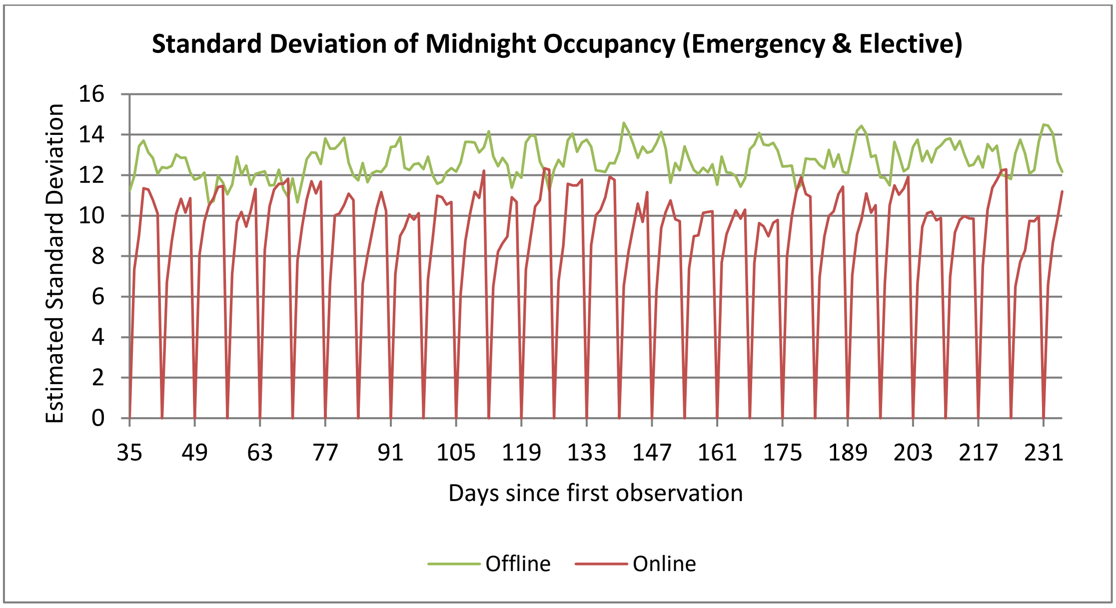


Figure 4.10: Estimated standard deviation of midnight census realisations over time for both the online and offline models.

Figure 4.11 further illustrates the degree to which variation in the midnight census across all modelled wards increases, on average, over the course of a one-week planning horizon when the online model is initialised every Monday at midnight. Since the distribution of midnight occupancy is clearly changing as a function of time, an online validation process should endeavour to check that this evolution is in some way consistent with the system it is intended to model.

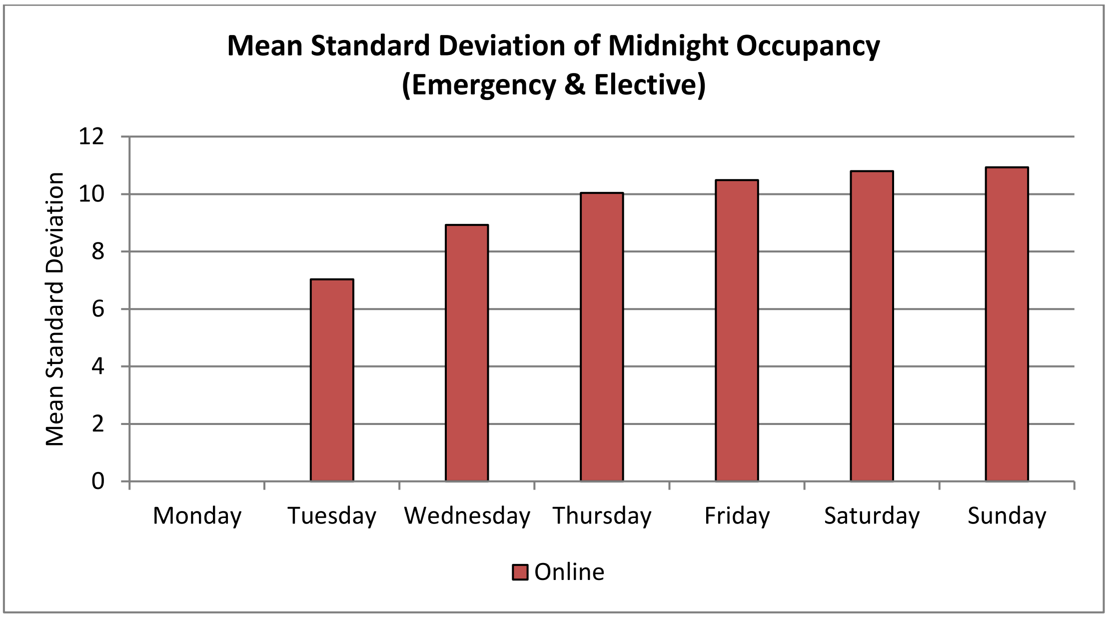


Figure 4.11: The estimated standard deviation of simulated midnight occupancy for each day in the observation period is averaged for the day of the week on which it occurs. Standard deviation on Monday is necessarily zero since each of the 100 simulation replications returns the same value on the day of initialisation.

4.7.3 The Conditional Distribution of Midnight Occupancy

In Section 4.5.3, checks of model validity were conducted by comparing mean midnight occupancy coming from the offline model with equivalent statistics from the PA data, and in this analysis, weekday and admission type were likely to be important factors influencing the distribution of midnight occupancy. In an online simulation context, the set of factors influencing the distribution of the key performance indicator is augmented to include the state of the system when it is most recently initialised, and the elapsed simulation time since this occurred, therefore the validation of an online model necessarily involves the validation of a set of *conditional* distributions of the performance indicators.

In an ideal situation, a sample of observations generated under the same set of conditions could be used to validate the conditional distributions generated by the model. However, this is likely to pose a statistical challenge when, for example, the state space of possible initial conditions is large. When this is the case, it is unlikely that any observed system state will be revisited during the observation period, resulting in a single recorded trajectory from each of these states. An observation at each time point in the trajectory is clearly insufficient for drawing conclusions about the goodness-of-fit of the conditional distributions from the online model corresponding to each of these time points. Therefore, some pooling of the observations is necessary to achieve a suitable sample size.

However, some care must be taken when pooling observations generated under different initial conditions or at different elapsed times, since these observations are not strictly realisations from the same conditional distribution. The observations should therefore be normalised in some way to account for these differences. If observations are pooled with no normalisation to account for the difference in conditional distribution, then the validation is effectively no different to an offline validation analysis, in which some of the factors influencing the distribution of the observations (such as previous system states) are aggregated.

One way of conducting a normalised comparison of the observed and simulated output data, which accounts for differences in initial conditions, is by computing the proportion of observations which are less than a chosen percentile from the simulated (hypothesised) distribution. For example, regardless of initial conditions or time from initialisation, if an online simulation is performing as expected, approximately 20% of the observations from the data should fall below their simulated 20th percentile, and this idea can be extended to a range of percentiles to assess the validity of the online model.

By working with cumulative probability as opposed to raw values, the simulated and observed data points can be normalised in such a way that a comparison between them can be conducted for the entire observation period, while accounting for the factors influencing their distribution (initial conditions and the elapsed time from initialisation). However, doing this comparison involves the inverse empirical CDF for each simulated day to count how many real midnight observations are below a chosen percentile. If the simulated distributions are discrete, as is the case for midnight occupancy, inverting these distributions can result in the same percentile being returned for a range of cumulative probabilities; resulting in what appears to be overestimation of the proportion of observations less than the chosen percentile. This is particularly evident when the simulated distributions are supported by a small range of values. For example, the 10th percentile might equal the 20th percentile for many of the simulated days. If the simulation models the real system well, the proportion of observed midnight occupancies which are less than their corresponding 10th percentile will also be closer to 20%, since these percentiles are equal. Therefore, in this example, arguing that 10% of the real observations should be less or equal to their simulated 10th percentile leads to the conclusion that the model is not performing well, when it is not the simulation at fault. This situation is also possible when working with continuous performance indicators, although it is easily mitigated by increasing the number of replications, which increases the resolution of the support of the simulated CDFs.

Because of the challenges associated with normalising discrete quantities (the number of occupied beds at midnight) in a way which accounts for the effect of the initial conditions *and* the elapsed time from initialisation, checks of online model validity in the next section will focus on comparing the simulated and observed distributions as functions of elapsed time only. The investigation of normalisation methods which allow the pooling of *discrete* observations generated by different initial conditions is left as further work.

4.7.4 The Δ-Occupancy Method for Validating Time-Dependent Distributions

Figures 4.10 and 4.11 in Section 4.7.2 illustrate how the standard deviation of the distribution of midnight occupancy coming from the online model changes as a function of the elapsed time from initialisation, which is to be expected, given that estimates further in the future are subject to random variation for a greater period of time than those near the time of initialisation. This is also true of the *observed* midnight census series relative to some previous system state, although this effect cannot be directly observed since the PA data contains only one observation of midnight occupancy each day. It can however, be observed indirectly, by considering the distribution of the difference between observed midnight occupancies h days apart. This random variable shall be referred to henceforth as the Δ_h*-occupancy* on ward w, with each realisation being defined as follows:

$$\Delta_{t,h}^{w} = M_{t}^{w} - M_{t+h}^{w} \qquad \text{(Eq. 4.4)}$$

where M_t^w represents midnight occupancy t days from the start of the observation period on ward w.

With a 560-day observation period, there are $560 - h$ realisations of Δ_h-occupancy on each ward $\left(\Delta_{1,h}^{w}, \Delta_{2,h}^{w}, \dots, \Delta_{560-h,h}^{w}\right)$ which can be derived from the PA data. The same number of Δ_h-occupancy realisations can be generated by each replication of the online model by using the elective schedule which was observed over the same period.

If the online model is initialised at each time t, then M_t^w will take the same value in both the simulation and the PA data. A comparison of the distributions of Δ_h-occupancy coming from the simulation and the PA data is therefore an assessment of how similar the distribution of ward occupancy is as a function of elapsed time from initialisation, thereby providing an indication of online model validity.

If the length of the planning horizon is assumed to be one week and the online model is initialised weekly, then there are 6 empirical distributions of Δ_h-occupancy (one for each day of the planning horizon $(\Delta_1, \dots, \Delta_6)$ from both the PA data and the simulation outputs which it makes sense to compare. If $F_{sim}(\delta_h^w)$ and $F_{data}(\delta_h^w)$ denote these empirical cumulative distribution functions over the support of Δ_h^w, denoted by δ_h^w, then the coordinates $\left(F_{sim}(\delta_h^w), F_{data}(\delta_h^w)\right)$ form a so-called *probability-probability plot* or *P-P plot*. If the distributions are similar, the coordinates will lie close to the identity line $(y = x)$, providing a visual indication of the similarity of the distributions of Δ_h-

occupancy at each possible elapsed time h from initialisation, or equivalently, on each of the h days in the planning horizon. By using P-P plots, a comparison of the simulation outputs and the data for each h can be presented in one graph, rather than analysing six (in this case) pairs of histograms (simulation vs. observed data). Having good agreement in distribution (as opposed to only having agreement in central tendency, for example) allows the user to estimate the probability of exceeding capacity thresholds during each day of the planning horizon; a metric which could be used to assess the quality of a given elective schedule.

4.7.5 Online Validation Using Δ-Occupancy

Rather than presenting the P-P plots of Δ_h-occupancy for all ten of the modelled wards, for brevity, the results from two wards (which are broadly representative of the others) are included in this section, while results for the remaining wards are included in Appendix A. Accompanying these P-P plots are histograms of raw midnight occupancy in which the time-dependent nature of the distribution of the simulation outputs is ignored (by pooling the midnight occupancy realisations over h) to illustrate the agreement between the simulated and observed data in a more familiar format.

Figure 4.12a compares the cumulative distributions of Δ_h-occupancy observed in the historic data, with equivalent distributions generated by the output of 100 replications of the online model, for Ward 5D. This ward has the highest average midnight census over the observation period (29 occupied beds), split between emergency (83%) and elective (17%) patient types.

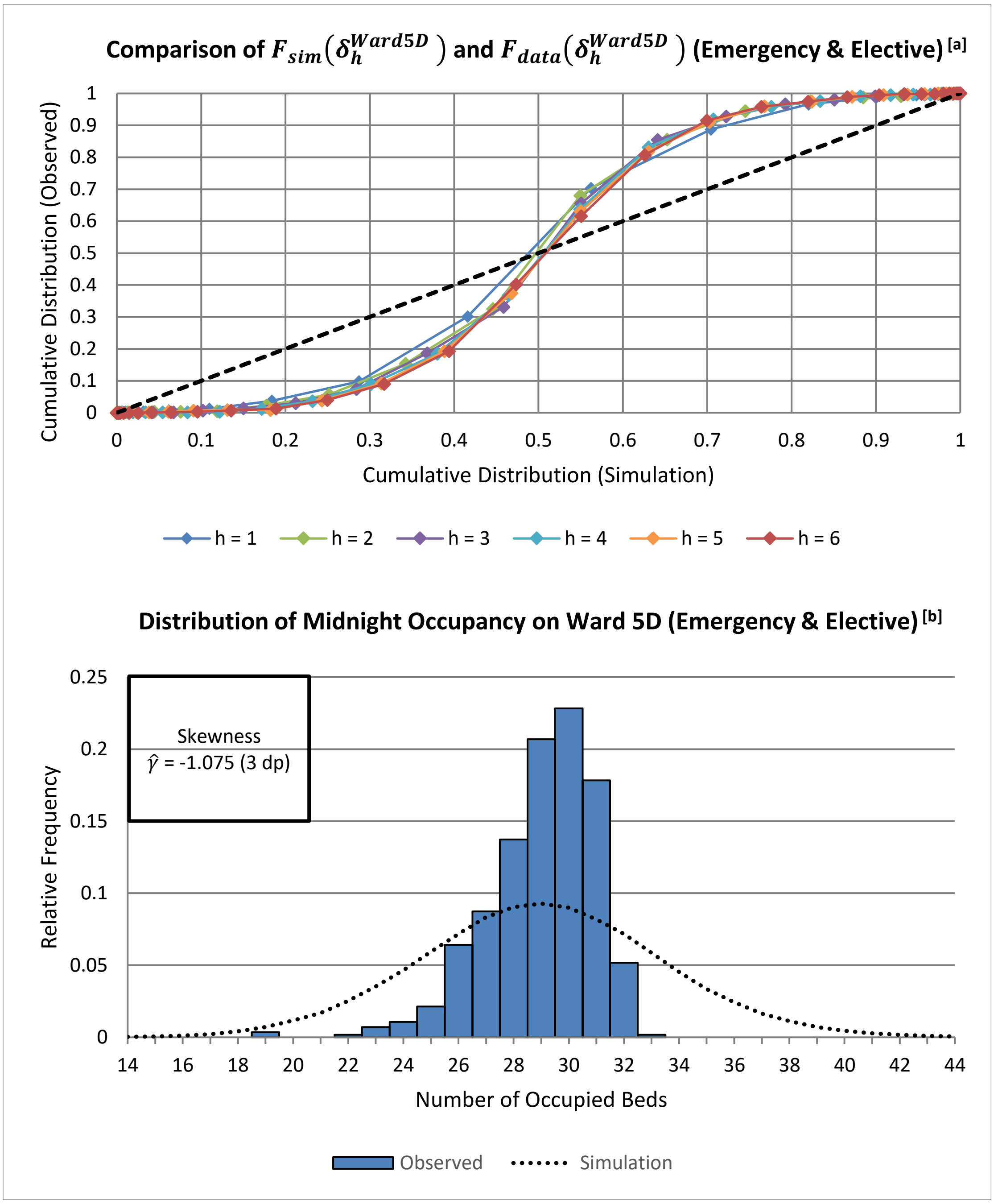


Figure 4.12: [a] The cumulative distributions of Δ_h-occupancy observed in the historic data, plotted against the cumulative distributions generated by simulation outputs for Ward 5D at each time from initialisation (h). [b] Histogram of midnight occupancies recorded on Ward 5D during the 560-day observation period, overlaid with the estimated p.m.f generated by the simulation (ignoring time-dependence).

The P-P plots show that the distributions from the observed data have less cumulative probability than the simulated distributions below their respective medians (for the same values of the support), however this difference reduces towards the point (0.5,0.5), and changes to a positive difference above it. With both the simulated and observed data having very similar medians for each value of h, this pattern is indicative of lower variance in the distributions plotted on the vertical axis, compared to the distributions plotted on the horizontal axis.

The lower variance of the Δ_h-occupancy distributions generated by the PA data are likely to be a consequence of the maximum capacity which exists on the real ward. This upper bound has the effect of curtailing occupancy when the demand for beds exceeds the ward's capacity, creating the negative-skew distribution shown in Figure 4.12b. The simulation on the other hand, employs an uncapacitated modelling approach which leaves midnight occupancy free to vary above the real maximum, contributing to the heavier right tail in the distribution of simulated midnights. The lighter left tail of the distribution of observed midnight occupancies may be the result of interventions made by the hospital to accommodate outliers from other wards, thereby decreasing the likelihood that the real ward will be found at lower occupancy levels. Such interventions are not made in the simulation in its current form, increasing the likelihood of simulating lower midnight occupancy levels relative to the observed data.

In contrast to Ward 5D, Figure 4.13a shows good agreement for the Emergency Department when comparing the empirical distributions of Δ_h-occupancy across a six-day planning horizon. Figure 4.13b shows that midnight occupancy on this ward is *positively-skewed*, and therefore less likely to be near, or at its maximum capacity than the other two wards. This means the existence of an upper bound on the number of occupied beds is likely to have little impact on the distribution of *midnight* occupancy for the levels of bed demand that ED typically experiences. However, emergency departments are well-known to be one of the busiest in the hospital, therefore the time of census collection (midnight) might play a part in this result. Also, having a busy emergency department might not necessarily correspond to high bed occupancy, since walk-in patients are accommodated in waiting rooms, depending on the seriousness of their condition.

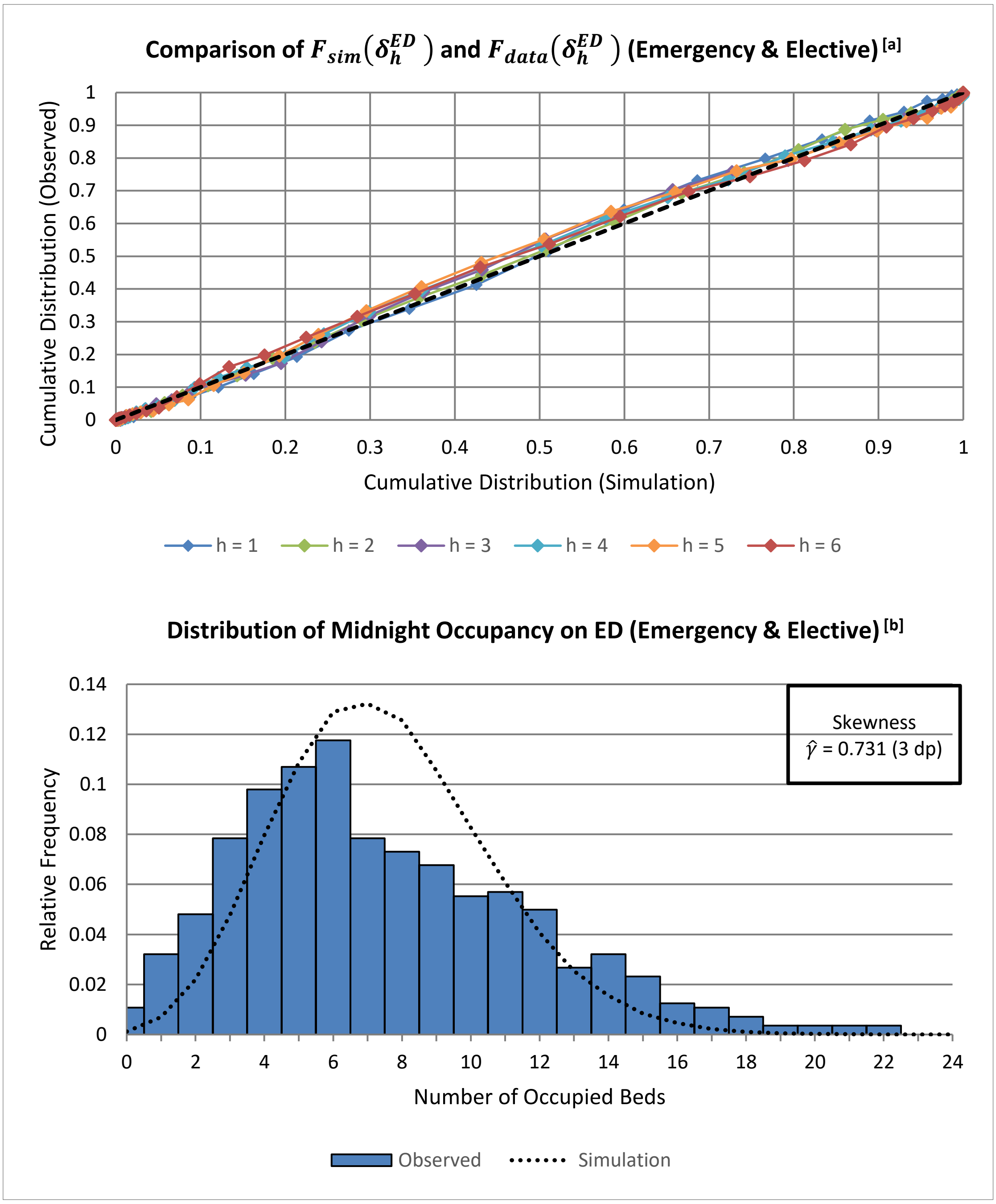


Figure 4.13: [a] The cumulative distributions of Δ_h-occupancy observed in the historic data, plotted against the cumulative distributions generated by simulation outputs for the ED at each time from initialisation (h). [b] Histogram of midnight occupancies recorded on the ED during the 560-day observation period, overlaid with the estimated p.m.f generated by the simulation (ignoring time-dependence).

4.7.6 Online Validation Summary

This section has focused on the development of an online simulation validation technique in which the time-dependence of the simulation outputs is accounted for, rather than using offline validation techniques in which only the "long-run" performance characteristics of the model are considered. By defining the Δ_h-occupancy random variable, the observed midnight occupancies can be pooled in such a way that comparisons can be made with the simulated midnight occupancies, whose distribution evolves with time-from-initialisation. Since Δ_h-occupancy is analysed via a comparison of the entire empirical distribution function, differences in trend, variability, or cycling behaviour which may occur over time, can all be detected.

Within each ward, the P-P plots presented in this section (and in Appendix A), show very similar patterns for each value of h, suggesting that where differences occur between the simulated and observed distributions, they are consistent as a function of the elapsed time from initialisation. These differences are caused (in part) by the conceptual modelling decision (Section 4.3.3) to treat each simulated ward as an uncapacitated node (heavier right tails), and the increased likelihood of finding the simulated wards at low midnight occupancy (heavier left tails), which could be attributable to the lack of any policy for distributing patient load between wards. Modelling the wards using an uncapacitated approach facilitates straightforward estimation of the degree to which a ward is over-subscribed (given historical levels of demand), although the patient diversions when this happens will not be modelled.

However, modelling non-ideal patient placements is likely to be important for operational planning purposes, and based on the histograms presented in this section, interventions such as these could be taking place in the real system when attempting to admit patients to wards which are frequently found near their maximum capacity, such as Ward 5D. The same can be said of Ward 4D and Ward 5B (Figure A.2.1b and Figure A.2.2b) whose histograms of observed midnight occupancy are clearly negative-skew, and to a lesser extent Intensive Care, Ward 5A and Ward 6D (Figures A.2.3b, A.2.4b and A.2.5b), which display some negative-skewness, although not to the same degree. These wards in the real hospital can also act as outlier wards for each other; absorbing patient placements when they cannot be accommodated at their "first-choice" ward. This behaviour contributes to the lighter left tails in the observed distributions, compared to their simulated counterparts, where it does not exist.

On the other hand, not all simulated wards over-estimate the variance of midnight occupancy. As Figure 4.13 shows, the Emergency Department is one ward in which the fit between the simulated and observed distributions is good for all times from initialisation, indicated by only small departures from the dashed line $y = x$. Similarly, Ward 4K, Northside, and the aggregate Other ward, also display good agreement for all times from initialisation, shown by Figures A.2.6a, A.2.7a and A.2.8a in Appendix A. Interestingly, the samples of observed midnight occupancy for these wards are *positively* skewed (Figures A.2.6b and A.2.8b), or fairly symmetric in the case of Northside (Figure A.2.7b), indicating that these wards are not near their maximum capacity as often as other wards, and are therefore less likely to turn patients away. This allows the agreement between the simulated and observed empirical distributions to be quite good, without the existence of rules for diverting patients during peak demand.

4.8 Discussion and Conclusions

To answer Research Question 1, that is; "*How can an on-line simulation, which provides estimates of bed demand, be developed for the operational management of hospital beds at the ward level?",* six stages of development have been described within this chapter, including a novel method for conducting "black-box" type validation of an online model, prior to its "real-world" implementation, and before any connection to the real system is established.

In the first stage of development, the requirements of an online simulation are discussed, to assess the feasibility of online modelling in the hospital context. At the second stage, a conceptual model of the hospital is developed at a level of detail which is capable answering questions associated with ward-level bed management, such as the likelihood of reaching a given ward's maximum capacity. The conceptual model developed at this stage is not dissimilar from a conceptual model developed for an offline or non-terminating simulation, although special attention is paid to *temporal* level-of-detail modelling decisions, in addition to structural level-of-detail modelling decisions, since the outputs of the eventual online model are necessarily time-dependent. This includes decisions regarding the frequency at which results are to be collected from the model, along with the time-scale of the possible decision variables which could be used to run alternative scenarios.

In the third stage of development, the conceptual model is implemented in the Micro Saint Sharp simulation package, and the PA data used to parameterise the model is analysed for the existence of time-dependent patterns which might affect the midnight census at the finest temporal level-of-detail (one day). Statistically significant relationships are found between emergency patient arrival rates and day-of-the-week, along with patient length-of-stay and weekday of admission when both patient types are pooled, therefore these relationships are included in the offline model.

The parameterised offline model is then run for the entire observation period, and the summary statistics generated by the realisations of midnight occupancy are compared with those observed in the PA data, after discarding the results from the warm-up period. These checks of offline model validity represent the fourth stage of model development which are intended to provide an indication that the model is performing as expected, by analysing simulation outputs generated by greater run lengths than the online model is likely to use. No statistically significant differences are found when comparing the mean midnight occupancy generated by the offline model with the PA data for each admission type (emergency/elective) across all wards. The variability of the model also seems comparable with the time-series of observed midnight occupancy for each admission type, with 88.3% of emergency patient censes falling within their corresponding 90% prediction interval, and 90.9% of elective patient censes doing the same (Section 4.5.3).

The penultimate stage of development sees the offline model augmented with the ability to be loaded with observed system states at initialisation. For the model to be used in practice, **Requirement 2** calls for these states to be loaded via an online connection to the real system, however, for the purpose validating the *online* model, a connection is made to a *historic* set of system states instead. The ability to load these states necessitates the inclusion of conditional length of stay distributions from which realisations of *remaining* length of stay can be drawn.

Finally, methods for black-box validation of the online model are developed by assuming a sensible length for the planning horizon, and re-initialising the online model using system states observed in the PA data at the beginning of each planning horizon period. These methods contribute towards clarifying **Requirement 1** in terms of obtaining a *validated* model for online use and offer techniques for comparisons of the distribution of the performance indicators, rather than comparing summary statistics. This type of online validation is distinct from the "auto-validation" sometimes associated with online models in the literature, since it allows for the assessment of the model's performance in an "online way" prior to its connection with the real system, while it is still possible to make complex adjustments to the model, if needed.

As mentioned in Section 4.7.3, the development of the Δ_h-occupancy random variable used to compare the distributions of midnight occupancy as they evolve over time, is a by-product of the discrete nature of the performance indicator (midnight census). This situation is not unique to the hospital setting; therefore, this method could be generalised to any online simulation in which the performance measure can be thought of as a discrete quantity. Although this method pools observations generated under different initial conditions (on which each distribution of simulated midnight occupancy is known to depend), it goes further than typical validation techniques, by evaluating the fit of the model with the observed data over time. For models with *continuous* performance indicators, it is possible to pool the simulated and observed data in such a way the initial conditions *are* accounted for, by first normalising the data, using the percentiles method described in Section 4.7.3. The validation method in the continuous-case assesses the agreement in distribution (as does Δ_h-occupancy), rather than simply assessing the similarity of summary statistics.

With Δ_h-occupancy defined, the quality of the fit of the distribution of the performance measure (midnight census) from the simulation is assessed against that of the data, for each time h (in days) from initialisation. While the quality of the fit does not change significantly with h, differences worthy of consideration are found to exist on wards which are more likely to be found at high levels of midnight bed occupancy, relative to their maximum capacity. The midnight occupancies generated for these wards by the online simulation are found to have higher sample variances than their PA data counterparts – a likely consequence of using uncapacitated simulation nodes to model wards in which the maximum capacity is more regularly encountered, coupled with the inability of the simulation to distribute patient load among free beds on other wards.

Research Question 1 has been answered in this chapter, through the development and validation of an online simulation for ward-level bed management. However, the development of additional model components remains to be discussed. Specifically, a method for modelling patient diversions during times of peak bed demand is proposed and evaluated in the next chapter. While these components could be viewed as part of the model development process, they also contribute to answering Research Questions 2 and 3 and are therefore discussed in the chapters which follow.

Chapter 5

Model Development 2: The Effect of Busyness on Patient Routing

5.1 Introduction

As part of the model development process described in Chapter 4, assumptions were made regarding the nature of patient-to-ward placement decisions. These assumptions included the choice to model hospital wards as uncapacitated nodes in the simulated ward network, along with the choice to assign patients to wards stochastically, based on so-called Static Transition Matrices (STMs). While these assumptions have merit in many modelling contexts (for example, STMs are likely to play an important role in systems which can be modelled using the theory of Markov processes), their combined usage does not fully address the dependencies which occur in real hospitals when one or more wards reach full capacity. In practice, many hospitals will divert patients to non-ideal wards when beds are unavailable on an ideal ward, however uncapacitated nodes in a simulation network will always accept new admissions.

Nevertheless, the ability to capture patient diversions in an operational bed management model is likely to play an important role in the estimation of ward-level bed demand over the course of a planning horizon in which the hospital experiences busy periods. A survey of *midday* bed censuses carried out by the Audit Commission (2003, p.7) found that on average, 7.5% of surgical beds were occupied by medical patients in NHS Trusts across England and Wales, and similar situations are not unique to the UK. Staff from the Australian General Hospital participating in this study also indicated (during early discussions) that the prevalence of so-called “outlier” patients on their wards posed several significant challenges for staff, such as relocating necessary equipment and forcing nurses to work outside their areas of expertise. Additionally, Harper and Shahani (2002) argue that outlier patients effectively represent unexpected emergency demand for the wards on which they are placed, and that such placements cause distress to the patient as well as having consequences for elective waiting lists.

Given the impact that periods of high bed demand or busyness are likely to have on patient routing policies through the inpatient ward network, the aim of this chapter is to address Research Question 2, that is; “*Can the effect of hospital busyness on patient-to-ward placement decisions be detected in patient administrative data, and can this be incorporated in a simulation model? If so, what effect does it have?”*

Research Question 2 is addressed in this chapter in the following way. First, the methods employed for routing patients in response to high occupancy within bed management models found in the literature are briefly discussed in Section 5.2. A novel patient routing method is then proposed in Section 5.3, and this provides a framework for the statistical analysis of the relationship between ward-level busyness and patient routing. In Section 5.4, this new method (parameterised by the PA data) is implemented in the online model, and the effect is examined via a comparison with the previous online model, in which

Static Transition Matrices govern patient routing. Section 5.5 concludes this chapter with a discussion of the results and answers to the research question.

5.2 Background

While outlier patients are likely to have a negative impact on the wards on which they are placed, modelling the entire decision process undertaken by a hospital which results in placement on a ward is not possible. As has been mentioned in Chapter 4, patient placement decisions are influenced by several factors, including the sex of the patient, their condition, nursing requirements and estimated length of stay, therefore some simplification of this process is inevitable for modelling purposes.

Inpatient models in the simulation literature employ a variety of different methods and simplifying assumptions for modelling patient routing through a network of wards, although in general, the assignment of patients to wards can be viewed as a two-step process in which a ward is chosen for a given patient, then the patient is admitted if some admission criteria are met. These criteria often relate to the availability of ward beds, but could be related to other resources limitations, such as nursing time. Models which include admission criteria can capture between-ward dependencies, since patients which are turned away must be placed elsewhere in the ward network.

In Günal (2008), this is achieved by deleting the ward which is at maximum bed occupancy from the relevant row in the transition matrix, and rescaling the remaining probabilities to sum to unity. The next ward is then sampled from this updated transition matrix. While this is a pragmatic solution to modelling a complex process, it assumes that the most likely ward (other than the ward which the patient has been turned away from), remains the most likely alternative ward in the event first choice ward is full. However, there is no reason why some other ward could not become the most likely choice, once it is observed that the first-choice ward is full.

Harper and Shahani (2002) asked bed managers to provide information relating to patient priorities, and these come into effect when attempts are made to place a simulated patient on a full ward. Assuming this information is representative of how the priorities are applied in practice, this might be a good way of determining where to place outlier patients in the simulation. However, this method requires detailed information to be provided by hospital staff, and the process of updating priorities obtained in this way cannot be automated if they were to change over time.

For uncapacitated models, such as the one developed in Chapter 4, each ward's admission criteria will always be met, therefore modelling the interaction between wards through the occurrence of patient turn-aways is not possible. This means ward admission criteria must be accounted for in the first step of the patient placement process, when the ward is being chosen by the patient placement algorithm, rather than responding to turn-aways when they occur. For the model described in Chapter 4, this means modifying the transition matrices which govern patient placements.

In contrast to the Static Transition Matrices (STMs) which contain fixed probabilities for each simulation run, Dynamic Transition Matrices (DTMs) are developed and evaluated in this chapter. DTMs contain functions rather than fixed probabilities and are designed to route patients according to occupied bed numbers on each ward; assessed at the time of transfer.

5.3 Transition Probability as a Function of Occupancy

At one end of the data-requirements spectrum, Günal (2008) makes assumptions about the nature of patient diversions by re-drawing from a set of possible wards when the first selected ward is not available. An algorithm such as this is relatively straightforward to implement in the simulation, however it is not informed by the diversion behaviour taking place in the real hospital. At the other end of this spectrum, Harper and Shahani (2002) collect detailed patient priority information from hospital experts. While this information can potentially improve the approximation of the routing behaviour within the simulation, its use rests on the assumption that the hospital is following the processes described by the expert in question, and that a different expert would not provide a different set of priorities. Additionally, there can be no automated process for collecting this data, should it change over time.

It may be possible to balance the trade-offs of both methods and strike some middle ground in terms of data requirements. Specifically, if the relationship between patient routing behaviour and ward level occupancy can be inferred from the administrative data, then doing so has the potential to improve the model's quality, without the need to rely on expert opinion. Some hospitals even flag the occurrence of outlier ward stays in their PA databases, which could also be used to inform a patient routing model. However, many do not, including the

AGH whose data was supplied for this research, although this information would be straightforward to incorporate within the statistical framework used in the next section.

5.3.1 Multinomial Logistic Regression

The online model developed in Chapter 4 contains 10 modelled wards, along with Entry and Exit nodes which are visited each time a simulated patient is admitted or discharged respectively. These nodes can be viewed as categories, and therefore techniques for the statistical analysis of categorical data can be applied to investigate the relationship between transition probability and ward-level busyness. Multinomial Logistic Regression (MLR) is a statistical technique for analysing the probability of categorical outcomes which may be influenced by other factors, such as ward occupancy. MLR is a generalisation of the more widely known Logistic Regression model, in that more than two outcomes can be considered. In this application, the outcomes are the destinations of patients after leaving their current ward or exit (discharge). The influencing factors (explanatory variables) are the occupancies of the 10 modelled wards. The general form of the MLR equation for this application is as follows:

$$ln\left(\frac{\pi_{i,j}(\boldsymbol{X})}{\pi_{i,w+1}(\boldsymbol{X})}\right) = \boldsymbol{\beta}_{i,j}^{\mathrm{T}}\,\boldsymbol{X} \text{ for } i,j \in \{1,\dots,w\} \qquad \text{(Eq. 5.1)}$$

where w is the number of modelled wards, and $\boldsymbol{X} = [1\; X_1 \dots X_n]$ is a vector of explanatory variables, namely ward occupancies and products of the occupancies of pairs of wards (i.e. two-factor interactions). Hence $\boldsymbol{\beta}_{i,j} = [\beta_{0,i,j}\; \beta_{1,i,j} \dots \beta_{n,i,j}]$ is the vector of regression coefficients associated with transitions from ward i to ward j which can be estimated using statistical software. $\pi_{i,j}(\boldsymbol{X})$ represents the probability of transitioning from ward i to ward j as a function of these coefficients and ward occupancies. To apply the MLR approach, a reference outcome must be chosen, which is used to fit the log-odds of observing any other outcome in Equation 5.1. The Exit node (ward $w+1$) has been used as the reference outcome for convenience, although the choice is arbitrary.

Equation 5.1 is the form of the model that is fitted in most statistical software packages (such as R or SAS), however some rearrangement is needed to provide the probability of transition $\pi_{i,j}(\boldsymbol{X})$. By making use of the requirement that the probabilities of the outcomes must sum to unity, Equation 5.1 can be written in terms of $\pi_{i,j}(\boldsymbol{X})$ so that it can be used in the simulation. This gives the probability of transition as a function of the regression coefficients and ward occupancies (including products).

$$\hat{\pi}_{i,j}(\boldsymbol{X}) = \frac{e^{\hat{\boldsymbol{\beta}}_{i,j}{}^{\mathrm{T}}\boldsymbol{X}}}{\sum_{k=1}^{w+1} e^{\hat{\boldsymbol{\beta}}_{i,k}{}^{\mathrm{T}}\boldsymbol{X}}} \qquad \text{(Eq. 5.2)}$$

For each source ward i, Equation 5.2 can be thought of as providing the 11 probabilities associated with the $j = 1$ to 11 possible destinations (10 modelled wards plus exit). Since there are 10 modelled wards, plus the dummy Entry node from which transitions can occur, i also ranges from 1 to 11, and 11 MLR models must be fitted to fully describe the DTM for each patient type (emergency/elective). As with STMs, reflexive transfers are only permissible from the ‘Other’ ward, therefore $\pi_{i,i}(\boldsymbol{X}) = 0$ identically, for all wards except ‘Other’.

With this framework for modelling the relationship between transition probabilities and ward occupancies, statistical software can be used to estimate the regression coefficients which are the basis of the MLR models. In the next section, the model fitting process is explained, using the Emergency Department as an example source ward.

5.3.2 Fitting the MLR Models

In order to estimate the regression coefficients and fit the MLR models, the data must be transformed in such a way that each observed patient transition is mapped to a set of ward occupancies. A dataset of this type can be constructed by querying the occupancy levels in the PA data at a time *just before* each transition occurs. This timing is important, as the more natural post-transition observation time causes the occupancy levels to be confounded by the transition itself.

Figure 5.1 shows an example of a dataset constructed in this way, which contains the details of ten transitions away from the Emergency Department. The field “Next_LocationID” is the dependent variable and stores the location of subsequent ward stays. The ten numeric fields (ED, IC,…,Other) contain the occupancy levels one second before the recorded transition time. These fields form the set of potential explanatory variables.

	LocationID	Next_LocationID	TransitionTime	ED	IC	Ward 4D	Ward 4K	Ward 5A	Ward 5B	Ward 5D	Ward 6D	Northside	Other
1	ED	Exit	22MAR10:00:14	8	1	25	12	25	20	30	28	11	17
2	ED	IC	22MAR10:06:09	8	1	25	12	25	20	30	28	11	17
3	ED	Ward 4D	22MAR10:10:50	12	2	22	13	26	23	30	28	11	20
4	ED	Ward 4D	22MAR10:10:51	11	2	23	13	26	23	30	28	11	20
5	ED	Ward 4D	22MAR10:12:16	13	2	24	13	25	21	30	28	10	20
6	ED	Ward 6D	22MAR10:13:29	13	2	24	12	24	18	30	27	12	20
7	ED	Ward 4K	22MAR10:13:42	12	2	25	12	24	18	30	28	12	19
8	ED	Ward 5B	22MAR10:15:40	14	3	22	12	24	18	28	27	11	18
9	ED	Ward 6D	22MAR10:15:59	14	3	22	12	23	19	28	27	11	18
10	ED	Ward 5B	22MAR10:16:17	13	2	22	13	24	19	28	28	11	18

Figure 5.1: An example of ten transitions from the dataset used the fit the MLR for the Emergency Department.

SAS software is used to fit the MLR models. The fitting procedure conducts a stepwise search over the set of possible explanatory variables (ED, IC,…,Other), and selects the combination of variables which minimise the Akaike Information Criterion (AIC). AIC is used as the model selection criteria, since the goal is to include all variables which increase the predictive capability of the models, whilst penalising over-fitting. However, by default, SAS uses p-values as the selection criteria while conducting the stepwise search, rather than the AIC. Therefore, some additional processing is required. The method of Shtatland et al. (2003) modifies the default SAS procedure to use the AIC as the selection criteria via a two-stage process. In stage one, selection is carried out using the default stepwise procedure, but with high p-values (such as p = 0.5) to eliminate the least significant variables from further consideration. In stage two, the model which minimises the AIC is chosen from the models which were encountered during the search in stage one. An alternative to this process is to evaluate the AIC for all possible combinations of the explanatory variables, and then select the minimum. However, total enumeration over all possible combinations becomes less feasible as the number of candidate variables increases. An interested reader can find the SAS code for carrying out the two-stage selection process which is used to fit all of the MLR models, in Appendix D.

The model which achieves the minimum AIC for each ward of departure is used to estimate the regression coefficients. Figure 5.2 shows the set of coefficients which are estimated for the Emergency Department, for the MLR model which minimises the AIC. This model has 18 effects, 10 of which are two-factor interaction terms. Note that the reference outcome “Exit” is not seen in the “Response” field because the coefficients for the reference outcome are set to zero by the fitting procedure.

Response	Intercept	WARD5D	OTHER	WARD4D	WARD5B
LGH - IC	9.3342	-0.2542	-0.0705	-0.2915	0.6459
LGH - Ward 4D	20.9576	-0.3790	-0.2206	-0.2097	0.2967
LGH - Ward 4K	23.5242	-0.2879	0.0702	-0.5232	0.1753
LGH - Ward 5A	36.5854	-0.6878	0.0975	-0.7564	0.1786
LGH - Ward 5B	15.6085	-0.3905	-0.1693	-0.6054	0.4031
LGH - Ward 5D	21.4872	-0.2098	-0.4093	-0.00375	0.0471
LGH - Ward 6D	13.6554	0.0778	-0.3028	-0.4661	0.2107
LGH - Ward Northside	-112.1	2.0455	0.6539	4.5601	1.9038
Other	16.5340	-0.2193	-0.00022	-0.4120	0.1271

Response	WARD5A	ED	WARD6D	IC	WARD4D_IC
LGH - IC	-0.2878	0.0122	0.3688	-1.8121	0.0607
LGH - Ward 4D	-0.6219	-0.0346	0.1574	-0.9896	0.0324
LGH - Ward 4K	-0.5369	-0.0319	0.2024	-1.3928	0.0490
LGH - Ward 5A	-0.9159	-0.0770	0.1982	-1.4821	0.0528
LGH - Ward 5B	-0.3530	-0.0779	0.5610	-1.0994	0.0384
LGH - Ward 5D	-0.7331	-0.1485	0.3275	-0.9867	0.0344
LGH - Ward 6D	-0.1989	-0.1290	0.2890	-1.1415	0.0368
LGH - Ward Northside	-0.5062	-0.2237	0.4332	0.7710	-0.0226
Other	-0.4875	-0.0757	0.2532	-0.8757	0.0294

Response	WARD4K	WARD4D_WARD5B	WARD5D_WARD4D	WARD5B_WARD5A	WARD5D_WARD5A
LGH - IC	-0.0666	-0.0111	0.00110	-0.00105	0.00834
LGH - Ward 4D	0.3454	-0.00990	0.00164	0.00509	0.0137
LGH - Ward 4K	0.0584	0.00549	-0.00342	-0.00373	0.0182
LGH - Ward 5A	0.1057	0.00608	0.00554	-0.00510	0.0287
LGH - Ward 5B	0.2871	0.00575	0.00396	-0.00675	0.0149
LGH - Ward 5D	0.3966	0.00590	-0.0182	0.00471	0.0177
LGH - Ward 6D	0.5320	0.00113	0.000421	0.00403	-0.00180
LGH - Ward Northside	-0.0567	-0.0827	-0.0691	0.0305	-0.0233
Other	0.2889	-0.00091	0.00619	0.00538	0.00821

Response	WARD5D_OTHER	WARD5D_WARD4K	OTHER_ED	WARD5B_WARD6D
LGH - IC	-0.00101	0.00252	-0.00343	-0.0128
LGH - Ward 4D	0.00591	-0.0132	-0.00209	-0.00589
LGH - Ward 4K	-0.00652	-0.00249	-0.00211	-0.00728
LGH - Ward 5A	-0.00747	-0.00383	-0.00100	-0.00683
LGH - Ward 5B	0.00196	-0.00975	-0.00019	-0.0192
LGH - Ward 5D	0.00885	-0.0141	0.00426	-0.0117
LGH - Ward 6D	0.00609	-0.0186	0.00471	-0.0139
LGH - Ward Northside	-0.0303	0.0108	0.00765	-0.0206
Other	-0.00317	-0.0102	0.00284	-0.00990

Figure 5.2: Regression coefficients for each of the 18 effects (plus intercepts) in the final MLR model for the Emergency Department and the emergency patient type.

After obtaining the regression coefficients of the selected model, the final step is to implement the MLR equations in the simulation. Since Equation 5.2 is written in terms of the transition probabilities $\pi_{i,j}(\boldsymbol{X})$, this form is used in the ODES. As an example, the form of the equation which governs the probability of transitioning from the Emergency Department to the Intensive Care Unit (for the emergency patients) can be written with fitted regression coefficients as follows:

$$\hat{\pi}_{ED,IC}(\boldsymbol{X}) = \frac{1}{\kappa} exp(9.334 + 0.012X_{ED} - 1.812X_{IC}$$
$$- 0.292X_{4D} + 0.061X_{4D}X_{IC} - 0.011X_{4D}X_{5B}$$
$$- 0.067X_{4K} - 0.288X_{5A} + 0.646X_{5B}$$
$$- 0.001X_{5B}X_{5A} - 0.013X_{5B}X_{6D} - 0.254X_{5D}$$
$$+ 0.001X_{5D}X_{4D} + 0.003X_{5D}X_{4K}$$
$$+ 0.008X_{5D}X_{5A} - 0.001X_{5D}X_{Other}$$
$$+ 0.369X_{6D} - 0.071X_{Other}$$
$$- 0.003X_{Other}X_{ED}) \qquad \text{(Eq. 5.3)}$$

The denominator κ is used in Equation 5.3 for brevity, although it represents the sum of the numerator in this expression, and the numerators in the equivalent expressions for $\hat{\pi}_{ED,j}(\boldsymbol{X})$. In a more general sense, κ is equivalent to the expression in the denominator of Equation 5.2. In Micro Saint Sharp, the MLR code follows the form of Equation 5.3 closely, and the precise syntax is listed in Appendix D. The appendix includes the equations for the Emergency Department, in which the coefficients are consistent with the values presented in Figure 5.2, and the equations for all other wards and patient types.

5.3.3 Summary of the MLR Models

As with any regression model, it is possible to make inferences about the relationship between the dependent variable (in this case the likelihood of transitioning to a particular ward) and the explanatory variables (ward occupancies), however there are a number of ways in which the inference procedure is imperfect. There may well be occasions when there is a real relationship, but there is not enough evidence of it in the data for it to be detected. On the other hand, whilst the use of AIC is designed to prevent over-fitting, this is not guaranteed, and it is possible that relationships 'detected' in the data are spurious. Tables 5.1 and 5.2 summarise the MLR models for each ward and patient type, including the wards whose occupancies, and products of occupancies, were determined to be predictors by the MLR fitting process.

	Explanatory variables (midnight ward occupancies) in each MLR model
Ward MLR	**Emergency Patients' MLR**
Entry	ED, ED*5B, IC, IC*5B, IC*Other, 4D, 4D*Northside, 4K, 4K*5A, 4K*5B, 5A, 5B, 5B*5D, 5D, 5D*6D, 5D*Northside, 6D, Northside, Other
Emergency Department (ED)	ED, IC, 4D, 4D*IC, 4D*5B, 4K, 5A, 5B, 5B*5A, 5B*6D, 5D, 5D*4D, 5D*4K, 5D*5A, 5D*Other, 6D, Other, Other*ED
Intensive Care (IC)	IC, 4D, 5A, 5B, 5D
Ward 4D	IC, 4D, 5A, 5B, 5D, 5D*5A, 6D, Other, Other*5B
Ward 4K	Other
Ward 5A	ED, ED*6D, ED*Northside, IC, IC*Northside, 4D, 4D*5B, 4K, 4K*5A, 5A, 5A*6D, 5B, 5B*6D, 5B*Northside, 5D, 5D*Northside, 5D*Other, 6D, Northside, Northside*Other, Other
Ward 5B	IC, 4D, 5B, 5B*6D, 5D, 6D, 6D*4D, Other
Ward 5D	IC, 5A, 5D, Other
Ward 6D	6D, Other
Northside	IC, IC*6D, IC*Northside, IC*Other, 4D, 4D*5A, 4D*6D, 5A, 5B, 5B*Northside, 5D, 6D, Northside, Other
Other	ED, IC, 4D, 4K, 5D, 5D*ED, Other

Table 5.1: Summary of the effects which minimise the AIC in each of the 11 MLR models for the emergency patients.

	Explanatory variables (midnight ward occupancies) in each MLR model
Ward MLR	**Elective Patients' MLR**
Entry	N/A
Emergency Department (ED)	NULL
Intensive Care (IC)	5A, 5B, Other
Ward 4D	5A, 5B, 5D
Ward 4K	5D, Northside
Ward 5A	IC, 5D, Other
Ward 5B	IC, 5B, 5D, Other
Ward 5D	4K, Northside
Ward 6D	ED, 5A, 6D
Northside	NULL
Other	IC, 4D, 4K, 5A, 5A*5B, 5B, 5B*IC, 5B*Northside, 5D, 6D, Northside, Other, Other*5B

Table 5.2: Summary of the effects which minimise the AIC in each of the 11 MLR models for the elective patients.

The "N/A" entry for the elective patients' Entry model (Table 5.2) is to signify that the first location to which elective patients arrive is part of the elective admissions schedule, which is treated as a decision variable rather than a random variable. The two "NULL" entries in Table 5.2 identify two wards for the elective patients for which the null model (intercept only) achieved the lowest AIC. These are the Emergency Department and the Northside ward. In fact, these combinations also have the two smallest samples of transitions in the PA data (108 transitions and 9 transitions respectively). Hence, the inability to find any significant explanatory variables for these wards could well be due to a lack of statistical power. The remaining non-null models indicate that it is possible to detect some relationship between ward-level busyness and transition probability, although there may be some unavoidable examples of overfitting.

5.4 The Effect of Implementing Dynamic Transition Matrices

Regardless of whether the transition matrices are static or dynamic, the probabilistic routing structure used in DES is always “forward-looking”, i.e. transition probabilities govern where a simulated patient is *sent to*, rather than where they are *received from.* Since DTMs modify the way in which patients are *sent* to wards, they must be implemented across all wards at once to gauge their impact. In this section, the results generated by the simulation including DTMs are analysed via the Δ_h-occupancy random variable defined in Chapter 4, and the impact of their inclusion is assessed against the results generated by the model in which STMs are used.

With each of the 21 MLR models fitted, each vector of estimated regression coefficients $\hat{\boldsymbol{\beta}}_{ij}$ can be taken from SAS and used to create the functions $\hat{\pi}_{i,j}(\boldsymbol{X})$ in Micro Saint Sharp. At the end of each ward-stay for each simulated patient, the next ward (or discharge, through the Exit node) is drawn from the set of probabilities $\{\hat{\pi}_{i,1}(\boldsymbol{X}), \dots, \hat{\pi}_{i,w+1}(\boldsymbol{X})\}$ where $\boldsymbol{X}$ is a vector of ward occupancies collected at the time the simulated patient completes their stay on the current ward. Although the simulation runs in discrete time, the arrivals within each arrival batch are staggered with a negligible amount of random simulation time so that subsequent patient transitions do not occur in unison. This allows the ward occupancies, and therefore the transition probabilities, to update with each transition.

5.4.1 Run Configuration

With the DTMs implemented in the online model, a new set of midnight occupancies can be generated and used for analysis. Table 5.2 summarises the configuration of the simulation experiment, from which another 100 replications of the 560-day observation period are run.

	Treatment	
Online Modelling Element	*Emergency*	*Elective*
Arrivals/Admissions	Empirical Distributions (Stochastic)	Observed Schedule (Deterministic)
Ward Length of Stay	Cond. Emp. Distributions (Stochastic)	Cond. Emp. Distributions (Stochastic)
Ward Transitions	Dynamic Transitions (Stochastic)	Dynamic Transitions (Stochastic)

Table 5.3: Treatment of each of the major modelling elements in the online simulation, grouped by admission type.

5.4.2 DTMs vs STMs: Empirical Results

To assess the effect of implementing DTMs in the online model, P-P plots are used to display comparisons of the Δ_h-occupancy distributions with those of the PA data, for the static and dynamic models. These are stacked to provide a visual comparison of how the distributions from both simulations compare to the data, for all times from initialisation. As in Chapter 4, histograms of midnight occupancy are also provided which present the distributions (irrespective of time) in a more familiar format. The wards whose figures are included in this chapter are chosen in such a way that their results are representative of the remaining wards, for brevity. However, summary statistics for all wards are tabulated in this chapter for an overall comparison of the two transition models. The figures associated with the remaining wards are included in Appendix B.

Figure 5.3 shows the P-P plots generated by the online model with DTMs, and the online model with STMs respectively, for both patient types on Ward 5D. An analysis of this ward was also presented in Chapter 4, since it has the largest number of occupied beds at midnight (on average) of the ten modelled wards.

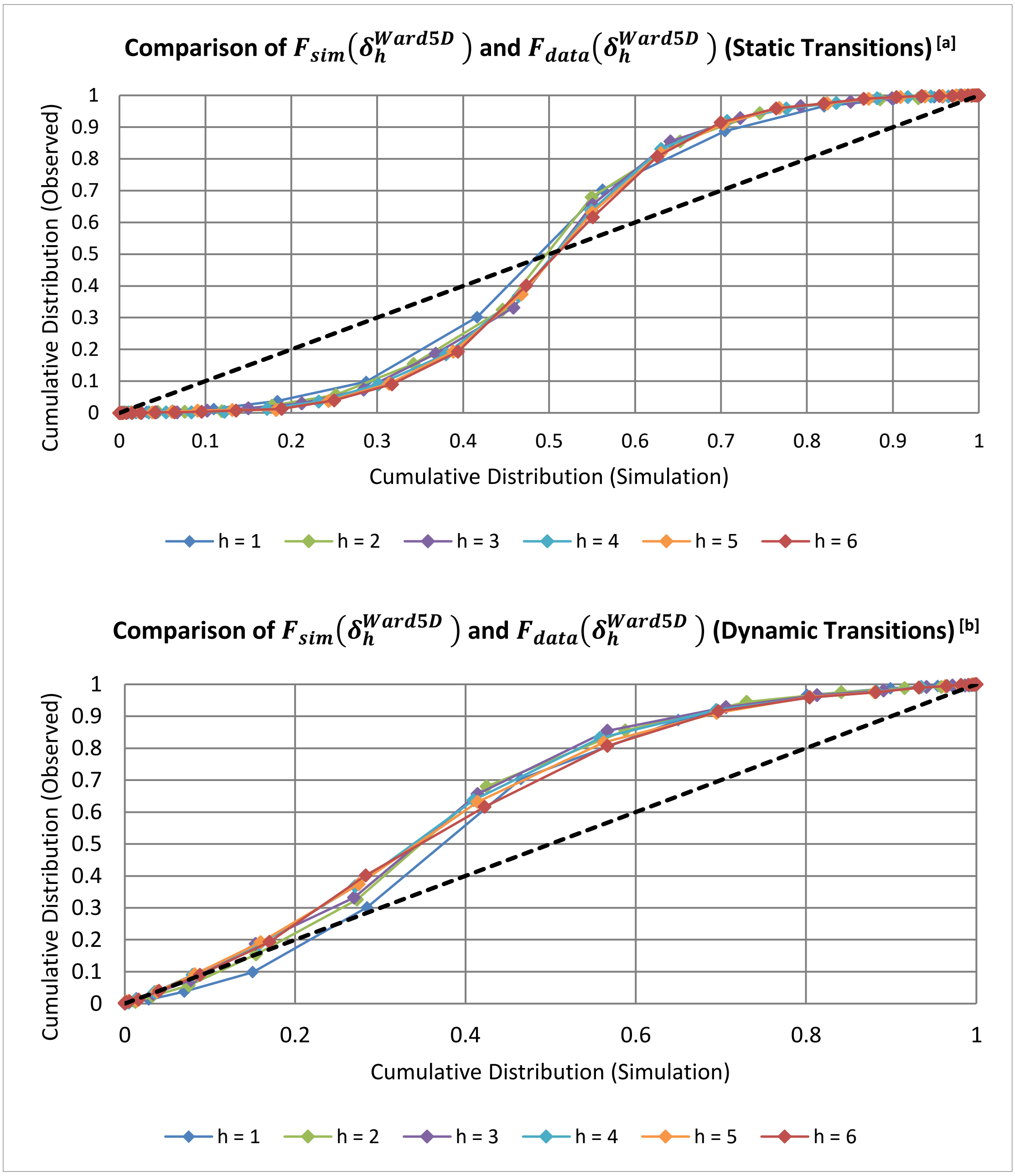


Figure 5.3: P-P plots comparing cumulative distributions of Δ_h-occupancy generated by the historic data and the simulation outputs for Ward 5D, under [a] static and [b] dynamic patient transition policies.

Comparing the P-P plots, there is a noticeable shift in median Δ_h-occupancy under DTMs, which is now found near the upper-quartile of the corresponding distributions derived from the PA data. Under STMs, the simulated and observed distributions share a very similar median for all times from initialisation. Also of note is the improved agreement in the left tails of the distributions under DTMs. The right tails however, are heavy for STMs and DTMs when compared to the Δ_h-occupancy distributions coming from the data – a pattern which is consistent for all times from initialisation.

While the P-P plots are useful for identifying differences across the planning horizon, it is not immediately clear from these plots which set of simulated distributions fit the observed data the best. Figure 5.4 pools the occupancy data for all values of h to generate a histogram of the observed midnight censuses on Ward 5D, and the estimated probability mass functions from each of the two simulations. In this format, the large variance under STMs, and the increased median under DTMs (identified by the P-P plots) can be seen, along with a reduction in variance when DTMs are used. This reduction noticeably improves the fit with the PA data and exists because the model can divert and accept patients to and from Ward 5D in response to its occupancy; thereby reducing the likelihood that the midnight census is found at extreme levels. Although this effect is most readily seen in Figure 5.4, the standard deviation for each Δ_h-occupancy distribution is calculated in Table 5.4b at the end of this section, which confirms a reduction in variability for all h, relative to the outputs generated by the model in which STMs are used.

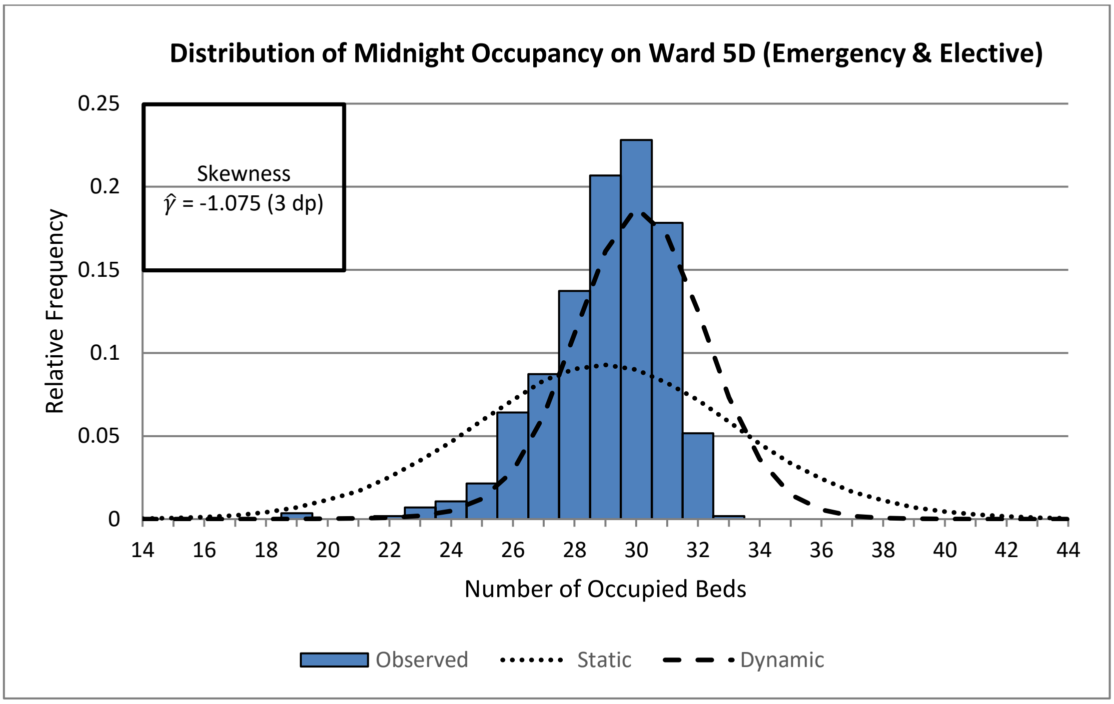


Figure 5.4: Histogram of midnight occupancies recorded on Ward 5D during the 560-day observation period, overlaid with the estimated p.m.f generated by the online simulations using STMs and DTMs (ignoring time-dependence).

Although the use of DTMs appears to have a desirable effect, the P-P plots in Figure 5.3b highlight differences in the cumulative distributions which warrant explanation. These differences increase as Δ_h-occupancy increases (for most of its range) and are largely the result of comparing a symmetric distribution (simulated census) with one which is negatively skewed (observed census). While it has been shown that the DTMs reduce the likelihood that simulated occupancy will be found at levels beyond its real maximum, the effect is not as pronounced as the effect that a *fixed capacity* has on the real ward, which causes the observed distribution to become noticeably skewed. When these two distributions are compared via P-P plots, the left-hand side of the distributions are similar, although the skewed real data means that the cumulative distributions increase faster than that of the simulation, resulting in the patterns seen in Figure 5.3b.

The inability of DTMs to completely reproduce this skewness also explains the shift in median, identified by the P-P plots. Although the likelihood of generating high occupancies has been reduced, the simulation with DTMs still creates midnight occupancies beyond what is possible in the real system, which in-turn increases the median, relative to the sample of observed midnights. While this is also true of the model using STMs (in fact, it is more likely to occur when STMs are used), the high occupancies are offset by low occupancies; which occur in the simulation more frequently than hospital staff would allow in practice (in the presence of busy wards elsewhere in the hospital). The result is a comparable median, and symmetric P-P plots, but significantly overestimated variance in midnight occupancy, when STMs are used.

In addition to the improved fit offered by DTMs for Ward 5D, there are also significant improvements for Ward 4D and Ward 5B; most readily seen in Figures B.2 and B.4 in Appendix B. Together, the figures for these three wards show the most noticeable improvements in fit compared to the STM-based model. Interestingly, the observed midnight occupancies obtained over the 560-day observation period for these three wards are the three highest, on average, while also being the most negative-skew. These results indicate that DTMs seem to have the greatest impact on wards which frequently experience high bed utilisation and are therefore the most likely to be influencing the use of outlier beds.

In Figure 5.5, the same analysis is presented for the Intensive Care ward, whose observed midnight occupancies are also negative-skew, although not to the extent of the three wards discussed so far. The most noticeable feature of Figure 5.5b, is that although the variance of the Δ_h-occupancy distributions appears to be quite similar under DTMs and STMs, median Δ_h-occupancy is increasing with time from initialisation, causing a greater disparity between the simulated and observed distributions than if STMs are used. To investigate this, ten simulations of the 560-day observation period were run in which each patient-pathway was collected as a string of visited wards. No re-initialisation of the system was carried out, since this would necessarily interrupt the collection of the pathways. This data revealed that averaged over the ten simulation runs, the number of visits to the Intensive Care ward increased by approximately 10% by moving from STMs to DTMs, which represents the second largest increase of all modelled wards and corresponds to an extra 73 visits to the ward per simulated year. This rate of extra patient accrual is around 1.4 per week using DTMs and is broadly in-line with the increase in mean Δ_h-occupancy by weeks-end (h = 6) of 1.12, when compared to mean Δ_h-occupancy using STMs (Table 5.4a). This suggests that the added visits are the source of the trending behaviour.

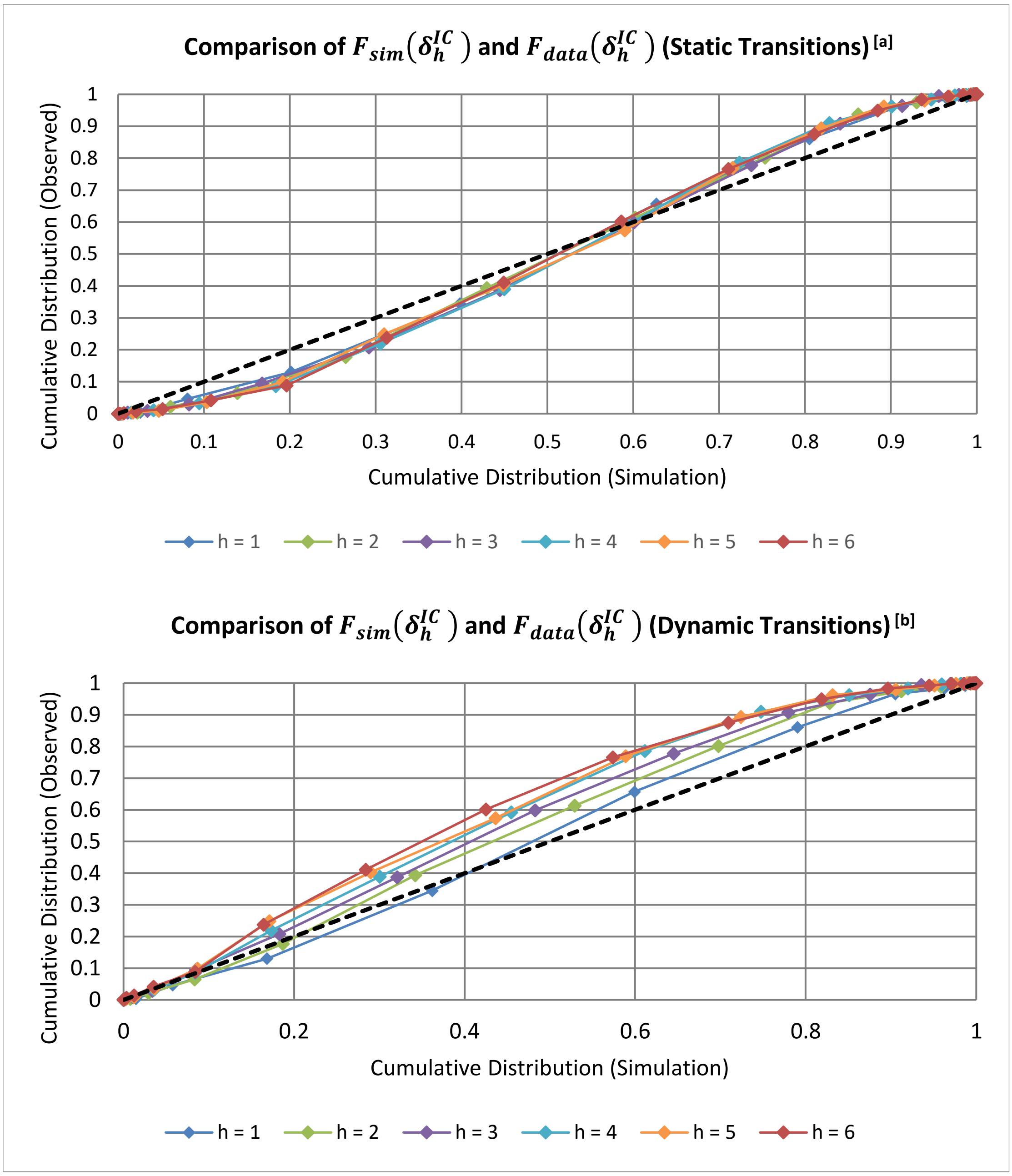


Figure 5.5: P-P plots comparing cumulative distributions of Δ_h-occupancy generated by the historic data and the simulation outputs for Intensive Care, under [a] static and [b] dynamic patient transition policies.

One explanation for this effect, is that although it might be possible for Intensive Care to accommodate outliers from other wards (contributing to the reduced likelihood of the ward being found at low occupancy, however expensive), patients are unlikely to be turned-away from Intensive Care once sufficient clinical need is established. The logistic regression models from which the DTMs are derived use only ward-level occupancy as explanatory variables, thereby ignoring what might be the most important factor affecting admission to Intensive Care; the severity of a patient's condition. While the distribution of observed ward occupancy is negatively skewed, indicating that bed availability is likely to play some part in admission considerations, the inability of DTMs to meaningfully reduce the likelihood of occupancy beyond the real maximum (illustrated by the heavier right tail in Figure 5.6), compared to STMs, suggests some information relating to the management of beds in Intensive Care is missing.

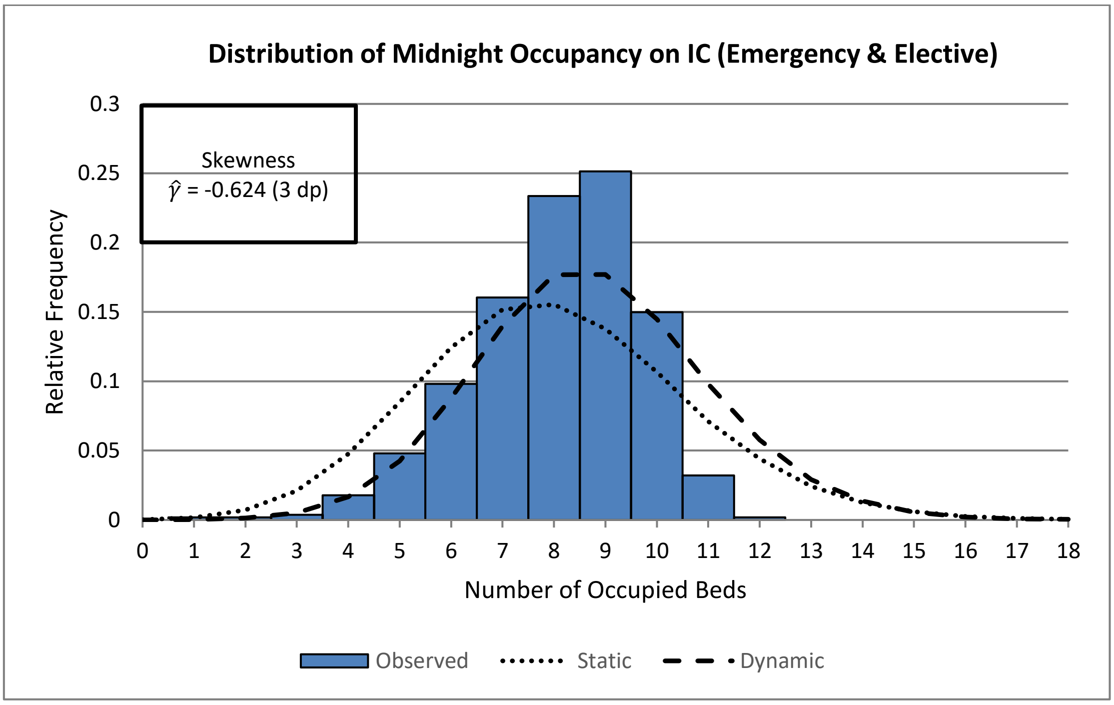


Figure 5.6: Histogram of midnight occupancies recorded in Intensive Care during the 560-day observation period, overlaid with the estimated p.m.f generated by the online simulations using STMs and DTMs (ignoring time-dependence).

Other wards which exhibit moderate negative-skewness include Ward 5A and Ward 6D (Figures B.6 and B.8). While the skewness of observed midnight occupancy is negative, these wards both feature distributions of a peculiar shape, making for a challenging modelling task, and possibly signalling bed management rules specific to these wards, an unusual patient case-mix, or some combination of the two. The shapes of the P-P plots for Ward 5A (Figure B.5) are similar to those of Ward 5D, in which DTMs appear to offer an improved fit with the observed data, however the histograms and overlaid empirical distributions for this ward (Figure B.6) offer less certainty. The DTMs make some attempt to reduce the variance of the simulated distributions, although the right-tail of the distribution remains too heavy, even with respect to the model in which STMs are used. The STM model on the other hand, performs poorly in terms of its variance, although it approximates the right-tail of the observed distribution well because of its strange shape, which could be useful for modelling the likelihood of exceeding the ward's capacity. Ward 6D on the other hand, while not experiencing a large improvement under DTMs, sees a notable improvement in fit for all midnight occupancy distributions (compared to their STM counterparts), owing to a reduction in variance of the simulated midnights. This is most readily seen in Figure B.8 in Appendix B, with variance estimates calculated in Table 5.4b, at the end of this section.

Finally, Figure 5.7 contains the P-P plots associated with the Emergency Department. In Chapter 4, the Δ_h-occupancy distributions for the ED (generated by the STM-based model), were shown to fit their observed counterparts well, which is also true of the model in which DTMs are used.

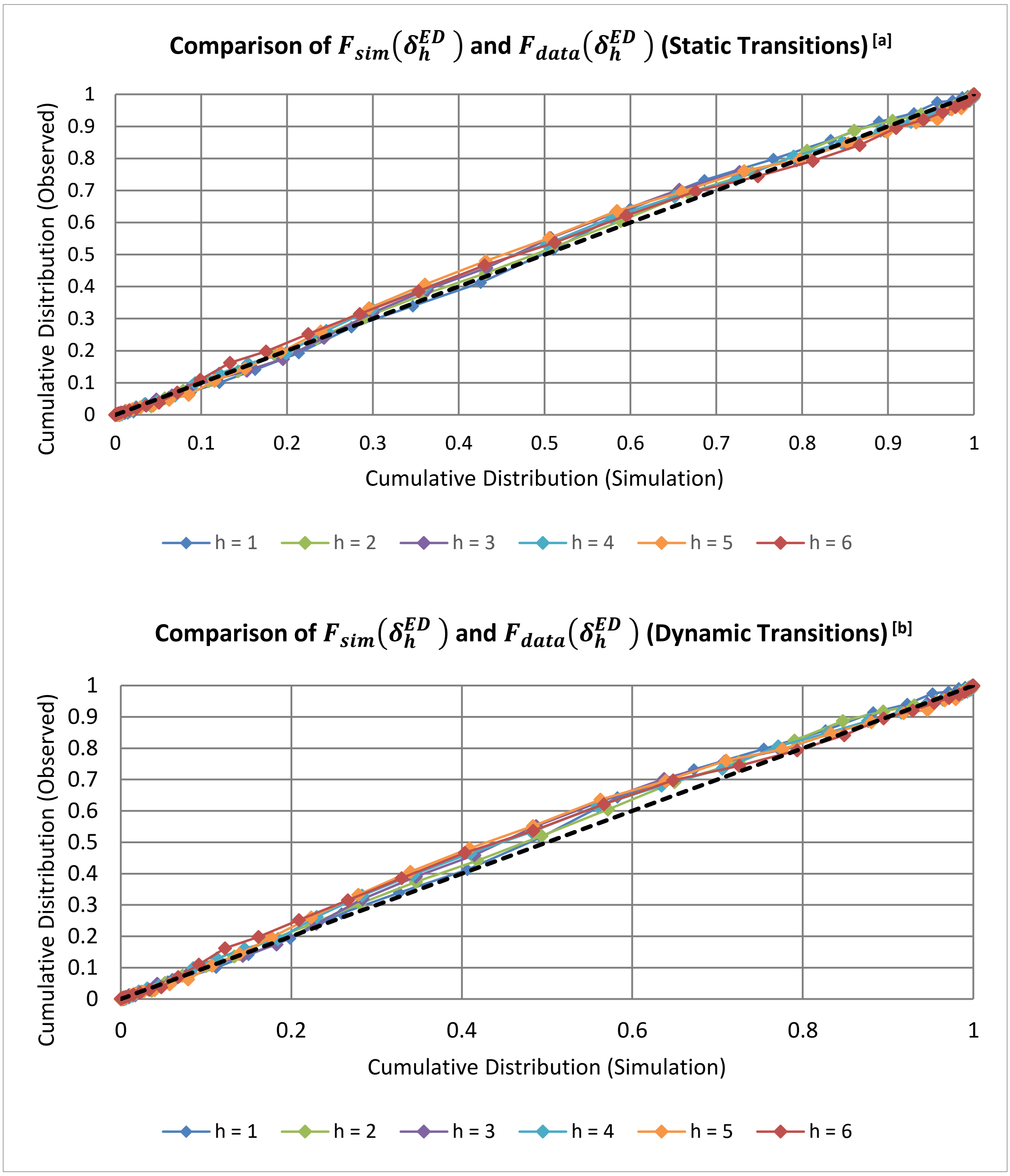


Figure 5.7: P-P plots comparing cumulative distributions of Δ_h-occupancy generated by the historic data and the simulation outputs for the Emergency Department, under [a] dynamic and [b] static patient transition policies.

Figure 5.8 shows that distribution of observed midnights is positively skewed, indicating that for the Emergency Department, the likelihood of encountering capacity issues at midnight is small. Consequently, the use of dynamic transitions has very little effect on the way the ward is modelled by the simulation, since patients are almost always able to be accommodated. A similar effect is seen on Ward 4K and Northside. Ward 4K is positively skewed (Figure B.10), while Northside is fairly symmetric (Figure B.12). Consequently, both wards also show good agreement between the simulation and the PA data, for all h, under both STMs and DTMs (Figures B.9 and B.11).

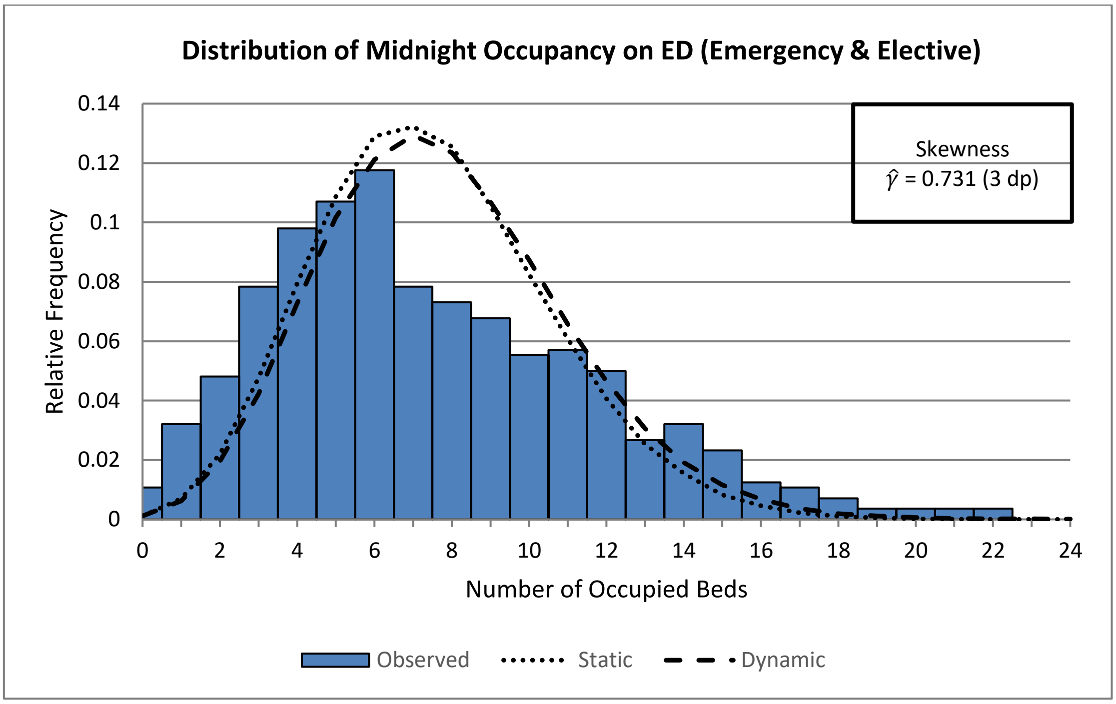


Figure 5.8: Histogram of midnight occupancies recorded in the Emergency Department during the 560-day observation period, overlaid with the estimated p.m.f generated by the online simulations using STMs and DTMs (ignoring time-dependence).

The sample of midnight occupancies from the Other ward (the pseudo-ward which represents a group of low occupancy wards on which the emergency and elective patient types stay) also exhibits slight positive skewness and good agreement with the simulated distributions (STMs or DTMs), if the time-dependence of the simulated distributions is ignored (Figure B.14). However, the P-P plots for this ward (Figure B.13) suggest that under DTMs, on average, the simulation tends to under-estimate occupancy early in the week, and overestimate it towards the end. This trending behaviour is similar to that of the Intensive Care Unit after DTMs are implemented, and like Intensive Care, this ward also sees an increase in simulated visits. In fact, Other and Intensive Care are subject to the two largest increases in ward visits under DTMs. For the Other ward, this is likely to be caused by lost information, due to the aggregate nature of the ward i.e. any relationship between the likelihood of transition to one of the sub-wards within Other, and the occupancy of that sub-ward, will always be confounded by the summation of occupancy levels on other sub-wards. However, in an operational setting, the performance of this ward may prove immaterial, since its creation was primarily to maintain adequate sample sizes for parameterising the model, and along with completeness of the ward network.

Part 1: Ward-Level Summary of Δ_h-occupancy							
Ward	***h***	***Mean***			***Std. Deviation***		
		Obs.	*STM*	*DTM*	*Obs.*	*STM*	*DTM*
ED	*1*	-0.01	0.05	0.29	4.31	4.71	4.66
	2	-0.02	0.06	0.30	5.10	5.16	5.16
	3	-0.03	0.04	0.35	5.44	5.37	5.43
	4	0.00	0.08	0.38	5.64	5.43	5.46
	5	0.02	0.08	0.41	5.44	5.26	5.32
	6	0.03	0.10	0.43	5.35	4.99	5.06
IC	*1*	0.01	-0.01	0.16	1.45	1.86	1.79
	2	0.03	-0.02	0.45	1.75	2.35	2.23
	3	0.03	-0.05	0.70	1.92	2.58	2.46
	4	0.03	-0.04	0.88	1.95	2.72	2.57
	5	0.03	-0.02	1.01	2.05	2.82	2.66
	6	0.03	-0.01	1.11	2.11	2.89	2.73
Ward 4D	*1*	0.01	0.04	-0.04	1.88	3.20	2.69
	2	0.02	0.09	0.23	2.52	4.14	3.27
	3	0.02	0.15	0.40	2.88	4.57	3.47
	4	0.02	0.21	0.52	3.11	4.82	3.56
	5	0.02	0.25	0.61	3.19	4.96	3.63
	6	0.03	0.31	0.72	3.19	5.05	3.69
Ward 4K	*1*	0.00	-0.01	-0.17	3.03	2.98	3.03
	2	0.00	-0.01	0.08	3.72	3.70	3.85
	3	0.00	0.00	0.33	4.08	4.05	4.33
	4	0.01	0.01	0.54	4.18	4.31	4.67
	5	0.02	0.00	0.70	4.25	4.47	4.92
	6	0.03	0.01	0.83	4.26	4.54	5.04
Ward 5A	*1*	0.00	0.04	0.77	2.09	3.21	2.75
	2	0.00	0.05	1.09	2.40	4.10	3.28
	3	0.00	0.06	1.35	2.65	4.61	3.63
	4	0.01	0.09	1.55	2.85	4.93	3.88
	5	0.01	0.09	1.70	2.92	5.07	3.99
	6	0.02	0.14	1.79	2.94	5.20	4.09

Table 5.4a: Summary statistics of observed Δ_h-occupancy, along with their STM and DTM counterparts for five wards. Light blue cells denote the simulated value closest to its observed counterpart.

Part 2: Ward-Level Summary of Δ_h-occupancy							
Ward	h	Mean			Std. Deviation		
		Obs.	STM	DTM	Obs.	STM	DTM
Ward 5B	1	0.02	0.02	-0.09	2.40	3.69	3.00
	2	0.03	0.03	0.16	2.90	4.56	3.49
	3	0.03	0.01	0.45	3.16	5.04	3.81
	4	0.05	0.05	0.69	3.49	5.35	4.03
	5	0.05	0.09	0.90	3.62	5.57	4.20
	6	0.06	0.10	1.04	3.75	5.67	4.28
Ward 5D	1	0.00	-0.01	0.71	1.40	2.84	2.22
	2	-0.01	0.00	1.02	1.74	3.80	2.53
	3	-0.01	-0.03	1.16	1.89	4.35	2.69
	4	-0.01	-0.04	1.22	2.01	4.69	2.77
	5	0.00	-0.05	1.21	2.06	4.88	2.81
	6	0.00	-0.05	1.16	2.05	5.00	2.84
Ward 6D	1	0.00	0.00	0.23	1.31	2.34	2.22
	2	-0.01	0.00	0.36	1.68	3.16	2.84
	3	-0.01	0.02	0.40	1.87	3.66	3.20
	4	-0.01	0.01	0.38	1.99	3.97	3.41
	5	-0.01	0.00	0.32	2.05	4.18	3.58
	6	-0.02	-0.02	0.27	2.08	4.33	3.71
Northside	1	0.01	0.00	-0.01	1.52	1.55	1.58
	2	0.01	0.02	0.04	2.06	2.10	2.21
	3	0.03	0.04	0.07	2.38	2.43	2.63
	4	0.04	0.04	0.06	2.62	2.68	2.96
	5	0.04	0.05	0.05	2.80	2.88	3.20
	6	0.04	0.05	0.03	2.94	3.02	3.40
Other	1	-0.01	0.06	-0.49	1.94	2.28	2.18
	2	-0.01	0.13	-0.40	2.63	2.92	2.89
	3	-0.01	0.16	-0.16	3.04	3.29	3.32
	4	-0.01	0.20	0.16	3.22	3.52	3.64
	5	-0.02	0.22	0.51	3.41	3.65	3.90
	6	-0.03	0.23	0.84	3.53	3.77	4.12

Table 5.4b: Summary statistics of observed Δ_h-occupancy, along with their STM and DTM counterparts for five wards. Light blue cells denote the simulated value closest to its observed counterpart.

5.4.3 DTMs vs STMS: Overall Evaluation

The results of this section have shown that Dynamic Transition Matrices provide the greatest improvement in distribution-fit for the wards which are most likely to cause outliers. Given that fixed capacities are known to exist in the real system, these wards are the ones which exhibit the largest degree of negative skewness. The primary contribution that DTMs make in improving the fit of the simulated distributions with the data is a reduction in variance; reducing the likelihood that the simulated realisations of midnight occupancy take very low or very high values. Indeed, the three most negative-skew wards are also those which show the biggest improvement in distribution-fit (via visual inspection) with the data gathered from the AGH.

As negative skewness reduces (approaches symmetry), so too will the frequency with which the hospital will be required to turn away patients from the ward, thereby decreasing the impact of using DTMs. Three wards were identified as being moderately skewed; Intensive Care, Ward 5A and Ward 6D. For Intensive Care, the trend identified in the outputs generated by the simulation meant that DTMs offered no improvement in fit, over the use of STMs. For Ward 5A, the peculiarity of the empirical distribution from the PA data meant that neither STMs nor DTMs performed particularly well, suggesting other factors, not included in the simulation, influence the distribution of midnight occupancy on this ward. For Ward 6A, the DTMs offer a clear improvement in the fit of the midnight occupancy distributions.

For the wards on which the distribution of midnight occupancy is positive-skew or symmetric, the performance of the two routing policies is so similar that the modeller is likely to be indifferent. These wards include the Emergency Department, Ward 4K and Northside. The aggregate Other ward *is* positively skewed, although the increase in average occupancy over time means that DTMs offer no improvement in terms of fit with the PA data, relative to the use of STMs.

In summary, if a modeller is *only* interested in predicting measures of central tendency over the course of the planning horizon, then given the choice between STMs and DTMs, and based on the estimates of mean Δ_h-occupancy shown in Table 5.4, STMs should be the routing policy of choice. They outperform DTMs in terms of absolute error of the simulation mean relative to the observed mean and require no statistical modelling to derive. However, if more information is required about the distribution of midnight occupancy then DTMs should be used. By using a routing model which is informed by occupancy, DTMs provide a better representation of the patient diversion strategy employed by many hospitals during peaks in bed demand, thereby improving the overall fit with the PA data, for the wards most frequently causing such diversions. Although most wards see a small overestimation in the value of mean/median midnight occupancy under DTMs (due to comparing skewed and symmetric distributions, in most cases), this is offset by improved estimation of the variability of the occupancy distributions. This in-turn improves estimates of the likelihood that a ward is found above a given capacity threshold during the planning horizon, on the assumption that the busyness-dependent routing behaviour estimated from the PA data will continue in a similar way.

5.5 Discussion and Conclusions

To answer the first part of Research Question 2, which is; "*Can the effect of hospital busyness on patient-to-ward placement decisions be detected in patient administrative data, and can this be incorporated in a simulation model?",* a statistical model (Multinomial Logistic Regression) was chosen which allowed the extent of the relationship between occupancy and transition probability to be quantified, for each ward. The use of a generalized linear model (along with appropriate statistical analysis software) means the *detection* of a relationship between the explanatory and response variables becomes part of the model fitting process. In this application, a relationship between ward-level occupancy and transition probability is detected by using a stepwise search, with AIC as the model selection criteria.

Tables 5.1 and 5.2 show that for all but two of the ward and patient type combinations, it is possible to improve the predictive capability of the *transition model* by incorporating occupancy information collected at a time just prior to the occurrence of a transition between wards. Therefore, it *is* possible to statistically detect the effect of hospital busyness on patient placement, when busyness is gauged by the number of occupied beds on each ward just prior to ward transition, and the transitions themselves are framed in terms of their probability of occurring.

As with any generalized linear model, once the regression coefficients have been estimated, the modeller is left with a functional relationship between the explanatory and response variables which can be used for prediction. Such an equation can be used in any simulation package which is flexible enough to define transition rules in terms of a mathematical equation. Therefore, the effect of hospital busyness on patient-to-ward placement decisions can always be incorporated in a simulation model provided the effects can be approximated by a set of formulae (such as Dynamic Transition Matrices) and the chosen simulation software offers sufficient flexibility when defining the routing policy.

The second part of Research Questions 2 asks, "*If so, what effect does it have?*". In terms of simulation run-time, using DTMs increases the real time taken to run 100 parallel 560-day simulations (re-initialising once per simulated week) by approximately 22%, adding around 2.5 minutes to a 12-minute simulation experiment in which STMs are used. However, the full 560-day run is solely for validation purposes, and in practice, the simulation would only be run for the length of the planning horizon. Therefore, the impact of the additional run time will be inconsequential for a simulation running on time-scales such as this.

In terms of the simulation outputs, the empirical results from this chapter show that for the wards whose simulated distributions provide a better fit to the data, the improvement is largely the result of a reduction in variability once DTMs are implemented, meaning the wards are less likely to be found at both very low and very high occupancies. Reduced probability in the right-tail of the distributions may be an expected result since one of the primary reasons for using DTMs is to redistribute arrivals and transfers at times when the ward is experiencing busyness. However, there is also a reduction in probability in the left-tails of the simulated distributions (compared to the model in which STMs are used), indicating that wards experience fewer days at low occupancy once the routing policy allows for outlier patients to occupy alternative beds, along with those which would have spent time on the ward regardless of busyness.

In addition to the model's improved ability to represent the impact of outlier patients for the largest wards, the statistical framework used to produce the DTMs (Multinomial Logistic Regression) allows patient diversion behaviour to be estimated directly from the hospital's PA database, rather than being derived from separately collected information, or assumed, for modelling convenience. A data-driven approach also allows the model to be easily recalibrated in the presence of new data, or if the routing procedures are believed to have changed appreciably after the model development phase. A data-driven approach to approximating the routing behaviour also means that a DTM recalibration procedure could theoretically be included within a so-called "auto-validation" module which automatically updates the simulation parameters as necessary.

Chapter 6
Case Studies

6.1 Introduction

In this chapter, two case studies are presented which demonstrate how the ODES developed in earlier chapters, could be used in practice. Motivated in part by the needs of the AGH participating in this study, the first case study demonstrates how the model could be used as an early-warning system to anticipate days in the planning horizon when the demand for beds is at risk of exceeding the maximum capacity of the wards. A particularly busy week is chosen from the PA database, and the ODES is used to assess the likelihood of demand exceeding capacity for the observed elective schedule. Since the elective admissions schedule is the decision variable for this model, a set of alternative schedules are developed (based on the initial results) and tested to illustrate how the model might be used to reduce the likelihood of excessive bed demand; thereby balancing emergency and elective workloads.

The second case study aims to answer Research Question 3 by investigating the potential for improving the results generated by the ODES by making use of clinicians' discharge date estimates. However, the subjective and potentially changeable nature of these estimates, compounded by informal collection methods (such as staff whiteboards or hand-written patient notes) means this information is often overlooked from a modelling perspective. While no survey has been carried out to collect clinicians' estimates from the AGH participating in this study, a post-hoc analysis can be conducted using patients' actual length-of-stay to assess the value of including information like this in an ODES model.

For both case studies, the ODES uses DTMs to model patient transitions. In practice, hospital management could choose either routing model (DTMs or STMs) according to whether they want predictions which incorporate historic dynamic behaviour, or alternatively want predictions which ignore possibly undesirable patient diversions.

6.2 Case Study 1: High Risk Plans

In this case study, the ODES is loaded with the initial conditions and elective schedule of a busy week during the 560-day observation period. The purpose of loading a particularly busy week is to demonstrate that the ODES can identify wards and days in the planning horizon which are likely to experience high occupancy, ahead of time. The second aim is to show how the ODES can be used to evaluate plausible changes to the elective schedule, and therefore reduce the likelihood of demand exceeding capacity for the wards which are most at risk.

The busy week chosen for analysis using the ODES, occurs 47 weeks into the observation period. This week was identified by counting the number of midnights each ward spends above its 90% occupancy threshold (listed in the second column of Table 6.1). The remaining columns of Table 6.1 show the observed midnight occupancies for each ward during Week 47, with those exceeding the 90% threshold shaded red.

Busy Week Midnight Occupancies								
Ward	90%	Mon	Tue	Wed	Thu	Fri	Sat	Sun
ED	19.8	14	13	15	17	4	9	8
IC	10.8	11	9	10	8	9	10	10
Ward 4D	27.9	29	28	27	30	31	30	30
Ward 4K	21.6	5	7	14	7	10	11	9
Ward 5A	28.8	25	25	28	30	31	30	30
Ward 5B	29.7	32	32	32	31	31	32	33
Ward 5D	29.7	32	31	31	31	31	31	31
Ward 6D	28.8	27	27	27	27	26	26	27
Northside	18.0	14	11	11	13	12	13	14
Other	20.7	11	9	10	12	12	14	12

Table 6.1: Midnight occupancy during Week 47 of the observation period. Red cells denote midnights which exceed 90% of maximum ward occupancy.

The corresponding elective schedule for this week was extracted from the PA data by counting the number of elective arrivals on each of the modelled wards for each day of the week. The number of arrivals per day, per ward, is shown in Table 6.2.

Observed Elective Schedule							
Ward	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
ED	-	-	1	-	1	-	-
Ward4D	-	-	-	-	1	-	-
Ward4K	2	3	-	1	-	-	-
Ward5A	-	1	-	1	-	-	1
Ward5D	-	-	-	2	-	-	-
Other	6	8	6	12	4	-	-

Table 6.2: Number of elective arrivals by ward and day-of-the-week for Week 47.

It is worth noting that although Table 6.1 shows Ward 4D, 5A, 5B and 5D to be the most highly occupied (in percentage terms) during the week, these wards have relatively few elective arrivals. Instead, Table 6.2 shows that the Other ward is responsible for the largest number of elective arrivals during Week 47. During the conceptual development of the model, the Day Procedure Unit (see Table 4.1) was grouped with several wards which also display low average midnight occupancy, to form the Other ward in the simulation. Since the Day Procedure Unit (DPU) sees many of the elective arrivals, but displays relatively low midnight occupancy, the DPU arrivals occur on the Other ward under this conceptual model.

6.2.1 Run Configuration

To assess the likelihood of bed demand exceeding the capacity on each ward during the week, the ODES is loaded with the state of the hospital at 00:00am on Monday of Week 47. On each of the seven days which follow, elective patients arrive onto the simulated wards based on the elective schedule shown in Table 6.2. Since the length of each simulation run is only required to be seven days, the number of replications can be significantly increased compared to the experiments conducted in Chapters 4 & 5. For the simulation experiments reported in this section, 400 seven-day replications were run, with the configuration outlined in Table 6.3.

	Treatment	
Online Modelling Element	*Emergency*	*Elective*
Arrivals/Admissions	Empirical Distributions (Stochastic)	Observed Schedule (Deterministic)
Ward Length of Stay	Cond. Emp. Distributions (Stochastic)	Cond. Emp. Distributions (Stochastic)
Ward Transitions	Dynamic Transitions (Stochastic)	Dynamic Transitions (Stochastic)

Table 6.3: Treatment of each of the major modelling elements in the online simulation, grouped by admission type.

6.2.2 The Observed Elective Schedule

By transforming the simulated realisations of midnight occupancy into histograms for each ward and day-of-the-week, the model can be used to assess the likelihood of demand for beds exceeding each ward's maximum capacity, or some other threshold of interest.

Figure 6.1 shows the histograms for each day during Week 47 for two of the modelled wards. For brevity, only Wards 5B and 5D have been included since their observed midnight occupancies sit above their respective 90% capacity thresholds for every day during the week; making them suitable for demonstrating the simulation's use as an early warning system. In this example, the dashed red line represents the 90% occupancy threshold, while the solid red line represents the maximum occupancy of the ward. The solid blue cells indicate the actual level of midnight occupancy which the ward experienced. For both wards, the distributions derived from the simulation outputs indicate that midnight occupancy is more likely to be above the 90% occupancy threshold, rather than below, for most days of the week. Therefore, the ODES could have been used to warn hospital staff of the high probability of high midnight occupancy.

In addition to indicating the days when midnight occupancy is likely to be above the 90% threshold, Figure 6.1 also shows that the ODES can be used to anticipate days where the demand for beds might be *more* than the number of beds which can be offered, due to the uncapacitated nature of the model. While it clearly isn't possible for wards to exceed their own capacity in practice (short of using hallway beds which are assigned to the ward), a high probability of over-occupancy in the simulation can indicate times during the week when the hospital is more likely to need to take some sort of preventative action, such as devising an alternative elective schedule.

Ward	Occupancy	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
Ward5B	39				0.25%	0.75%		
	38		0.25%	0.25%	0.50%	0.50%	0.75%	
	37		1.25%	0.75%	1.00%	2.25%		0.25%
	36		2.00%	2.00%	1.25%	3.50%	1.25%	1.25%
	35		4.50%	5.75%	2.50%	5.00%	2.00%	1.75%
	34		7.25%	10.25%	5.00%	11.00%	2.50%	5.00%
	33		13.75%	13.00%	7.75%	13.75%	7.00%	5.25%
	32	100.00%	15.25%	16.75%	14.00%	12.00%	11.75%	7.50%
	31		20.50%	14.75%	14.25%	16.75%	15.25%	7.75%
	30		13.75%	14.75%	12.75%	13.50%	11.00%	15.00%
	29		10.75%	9.00%	15.75%	8.25%	13.75%	12.75%
	28		5.75%	6.75%	9.75%	6.25%	12.00%	11.75%
	27		4.25%	3.75%	6.25%	3.25%	8.50%	11.25%
	26		0.75%	2.00%	3.75%	1.75%	5.75%	8.00%
	25			0.25%	3.00%	0.75%	4.00%	7.25%
	24				1.25%	0.25%	1.75%	2.75%
	23				0.50%	0.25%	1.75%	1.00%
	22				0.50%	0.25%	0.75%	0.25%
	21							1.00%
	20							
	19						0.25%	0.25%
Ward5D	37		0.25%	0.25%		0.25%		
	36		0.25%	0.75%		1.00%		
	35		0.50%	1.50%	2.50%	3.25%	0.50%	0.50%
	34		2.75%	2.25%	2.25%	5.75%	1.25%	2.75%
	33		6.50%	5.75%	6.25%	9.00%	2.75%	6.50%
	32	100.00%	15.50%	14.75%	12.25%	17.00%	9.75%	10.75%
	31		22.50%	16.25%	17.75%	17.50%	14.00%	17.25%
	30		22.50%	22.00%	17.00%	20.75%	17.00%	20.75%
	29		15.75%	20.25%	19.00%	10.75%	19.50%	16.00%
	28		9.00%	9.25%	13.25%	9.00%	14.75%	15.00%
	27		3.50%	4.75%	6.50%	2.50%	11.00%	6.25%
	26		1.00%	2.00%	2.25%	2.75%	3.75%	2.75%
	25				0.75%		4.25%	1.50%
	24			0.25%	0.25%	0.50%	1.25%	
	23						0.25%	

Figure 6.1: Distributions of midnight occupancy generated by the ODES for Week 47, on Ward 5B and 5D. The dashed red lines indicate the 90% capacity thresholds. The solid red lines indicate the wards' maximum capacity.

Although charts like Figure 6.1 are useful for visualising the midnight occupancy distributions for a small number of wards, the likelihood of demand exceeding capacity should be assessed for every ward, in a holistic way. This is especially true when gauging the merits of competing schedules, in the knowledge that dependencies between wards exist. Additionally, a method for quantifying any improvement offered by an elective schedule, across all wards, is also desirable.

Figure 6.2 charts the probability of bed demand exceeding the maximum capacity on *every* simulated ward, for the observed elective schedule shown in Table 6.2. Since the real hospital cannot exceed their own maximum capacities, the probabilities on Monday (when the simulation is initialised) are identically zero and have therefore been excluded. Also added to the chart, are the *bed-midnights over capacity;* a metric based on Chow et al. (2011), who computed *bed-days over capacity* by summing the number of beds in excess of a user-specified bed capacity for each day in the planning horizon[2]. Since bed-midnights over capacity (BMOC) are realisations of a random variable for each replication, the values presented hereafter are the mean of a sample of 400 runs.

---

[2] Midnights are distinguished from days, since the former is a point in time, while the latter is a duration.

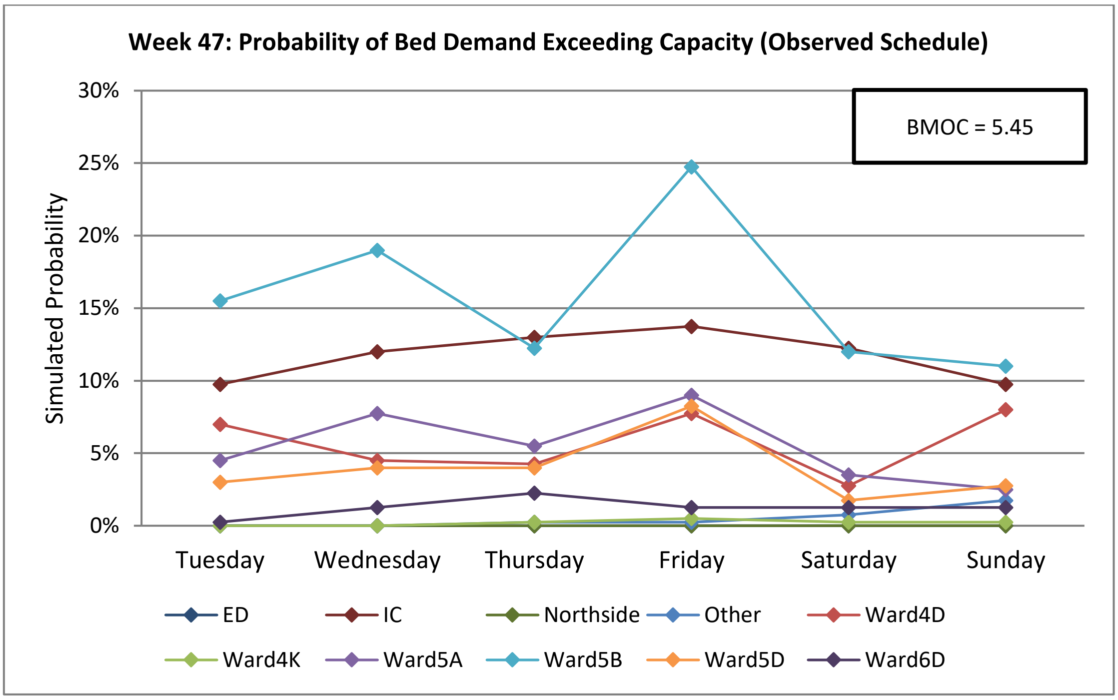


Figure 6.2: ODES estimates of the probability of the demand for beds exceeding total capacity on each of the ten modelled wards during Week 47.

Figure 6.2 clearly shows that during Week 47 (according to the ODES), Ward 5B is the most likely ward to encounter capacity issues for four of the six days of the week and is a very close second to the ICU on the other two. Whilst the next most likely ward is the Intensive Care Unit, the model validation analyses conducted in Chapter 5 showed that the use of Dynamic Transition Matrices tends to slightly overestimate midnight occupancy on the ICU, despite offering significant improvements on other wards. Therefore, the probability of running into capacity issues on the ICU is likely to be lower than is presented.

It is important to note that although the estimated BMOC for this week is 5.45, the real hospital cannot exceed the maximum capacities on each ward. Therefore, interventions (such as early discharges or unusual ward placements) could have been made by the hospital to cope with the prevailing bed demand, which are not modelled by the ODES. Nevertheless, charts such as Figure 6.2, in conjunction with estimated BMOC, serve as indicators for evaluating the quality of an elective schedule for each planning horizon, and may be related to the number of interventions needed to manage bed capacity during the week.

6.2.3 Alternative Elective Schedules

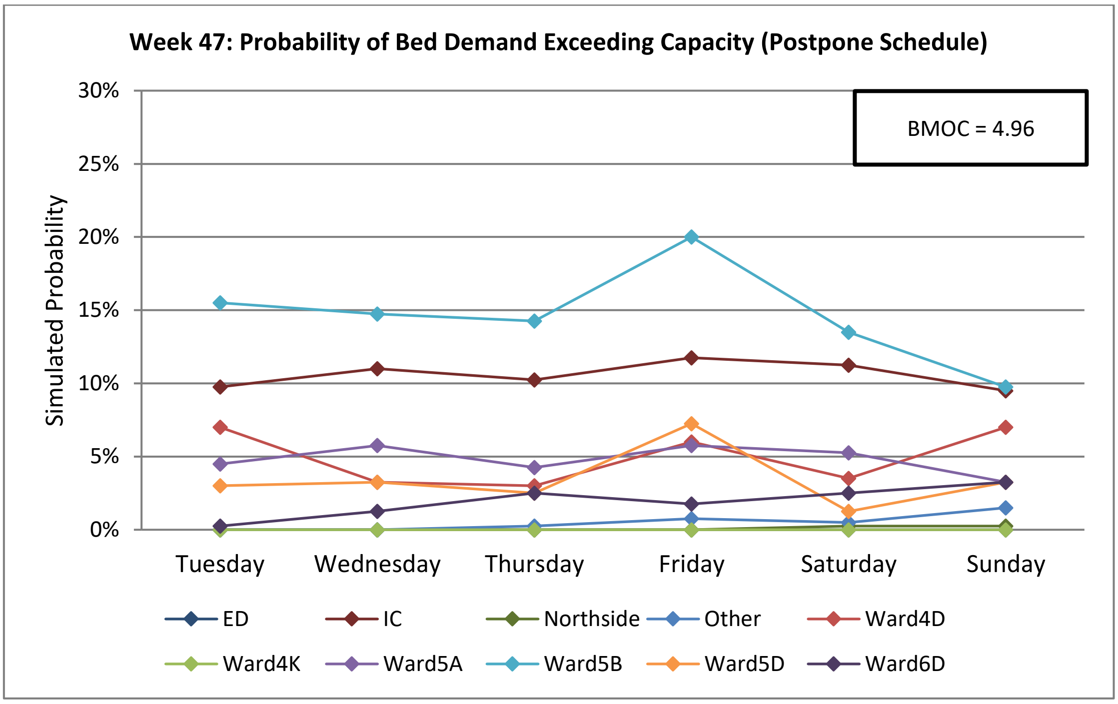
In this section, the ODES is used to investigate the effect of modifying the elective schedule for Week 47, to illustrate how the model might be used in practice. By using the observed elective schedule as a starting point, a set of alternative schedules are developed and tested in an iterative way, based on the results of the previous run.

For the sake of maintaining patient throughput and efficiency, the hospital might consider postponements until later in the week to be preferable to cancellations. Therefore, the first schedules investigated in this section will focus on *postponements* within Week 47.

Based on the results in Figure 6.2 from running the observed elective schedule, Ward 5B is the ward most likely to benefit from help in the form of schedule modifications, despite having no direct admissions to the ward during the week. However, almost 50% of patients who arrive at the Other ward (via direct admission or transfer) transition to Ward 5B (see table 4.3). Since the Other ward sees the largest number of elective admissions during Week 47, modifications to the elective schedule for the Other ward are likely to have an impact on Ward 5B.

The observed elective schedule in Table 6.2 shows that the days with the highest number of admissions to the Other ward are Tuesday (8 admissions) and Thursday (12 admissions). Similarly, Figure 6.2 shows spikes in the probability of exceeding capacity on Ward 5B on Wednesday (19%) and Friday (25%). Therefore, it is reasonable to expect that the peaks in the elective schedule, and the peaks in probability, are related, albeit one day later.

Figure 6.3 shows the probability of demand exceeding capacity, for all wards, by running a possible Postponement Schedule. In this schedule, three postponements are made which aim to reduce the predicted peaks in probability on Wednesday and Friday, in comparison to the observed schedule. The first postponement reschedules one Tuesday admission to Wednesday, and the next two postponements reschedule Thursday admissions to Friday. Two postponements are made on Thursday to reduce the higher Friday peak in probability, seen in Figure 6.2. Rescheduling the arrivals to the day on which peak probability is predicted to occur might seem counter-intuitive, however, Figure 6.2 and the *observed* elective schedule suggest that patients may not arrive on Ward 5B until the following day.



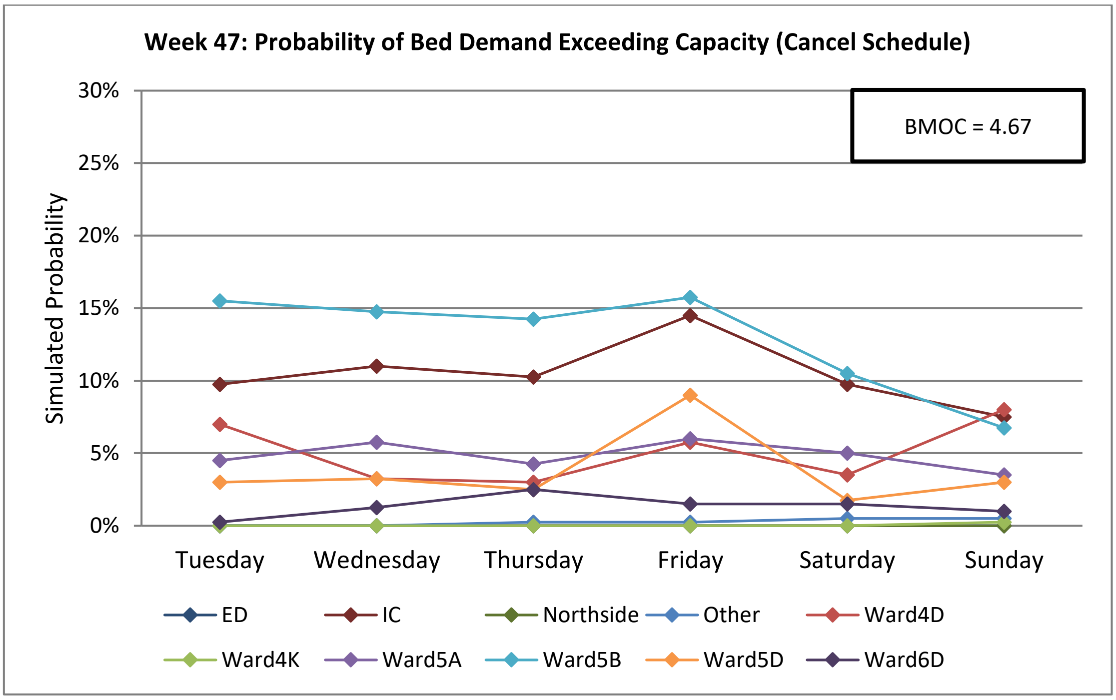
Postponement Schedule							
Ward	***Monday***	***Tuesday***	***Wednesday***	***Thursday***	***Friday***	***Saturday***	***Sunday***
ED	-	-	1	-	1	-	-
Ward4D	-	-	-	-	1	-	-
Ward4K	2	3	-	1	-	-	-
Ward5A	-	1	-	1	-	-	1
Ward5D	-	-	-	2	-	-	-
Other	6	8-1=7	6+1=7	12-2=10	4+2=6	-	-

Figure 6.3: ODES estimates of the probability of demand exceeding maximum capacity for Week 47, and the accompanying elective schedule.

The results for the Postponement Schedule show that the peaks in probability on both Wednesday and Friday for Ward 5B have been noticeably reduced. Since the schedule has been modified by postponing, rather than cancelling, an *increase* in probability is seen on Thursday and Saturday, however these two days are in a better position to accommodate additional patients than Wednesday and Friday. With the Wednesday and Friday peaks reduced, estimated BMOC also sees a decrease of 0.49 bed-midnights, suggesting a net improvement across all wards using this schedule for Week 47.

Although it may be possible to postpone more of Thursday's admissions to reduce the Friday peak on Ward 5B, further increases in probability over the weekend might be unacceptable if staff numbers are reduced over Saturday and Sunday. To further reduce the Friday peak on Ward 5B, without continuing to increase the probability over the weekend, the hospital might consider *cancellation* instead of postponement. Cancellation in this setting means that the patient will not be treated in the current planning horizon, however in practice, this is more likely to represent a *postponement* to a later planning horizon.

Figure 6.4 shows the effect of cancelling one admission to the Other ward during Week 47, in addition to the postponements which have already been made. The cancellation has the effect of further reducing the probability of Ward 5B encountering capacity issues by approximately 5%, and BMOC by 0.29. At first glance, a larger reduction in BMOC might be expected, given the removal of one patient from the admissions schedule. However, cancelling admissions to Other has consequences for all the wards, not just Ward 5B, because of the other routes the cancelled patient might have taken. The Intensive Care ward also sees a small increase in probability on Friday, which is likely to be caused by new sets of occupancies (based on the patient cancellation) which influence the transition probabilities generated by the DTMs.



Cancellation Schedule							
Ward	***Monday***	***Tuesday***	***Wednesday***	***Thursday***	***Friday***	***Saturday***	***Sunday***
ED	-	-	1	-	1	-	-
Ward4D	-	-	-	-	1	-	-
Ward4K	2	3	-	1	-	-	-
Ward5A	-	1	-	1	-	-	1
Ward5D	-	-	-	2	-	-	-
Other	6	8-1=7	6+1=7	12-3=9	4+2=6	-	-

Figure 6.4: ODES estimates of the probability of demand exceeding maximum capacity for Week 47, and the accompanying elective schedule.

6.2.4 Case Study 1: Summary

The alternative elective schedules presented in this case study demonstrate how the ODES could be used in practice to help hospital planners decide on potential elective admissions schedules in a more informed way. This is achieved by identifying the wards which are most at risk of running short of beds, along with peaks and troughs in the probability of demand exceeding capacity throughout the week. By modifying an initial elective schedule in an iterative way, the examples in this section show that reductions in the probability of encountering capacity issues can be achieved for an example busy week with minimal changes to the total number of elective admissions.

Although only two types of schedule modifications have been considered for creating alternative schedules, the hospital might have a range of possible actions that they could consider. Some of these actions can be framed as postponements or cancellations, and can be investigated directly using the ODES, while some are likely to be subtler. More subtle actions, such as changing the ward of first admission, have not been explored in the examples in this section, since they are likely to involve the use of extra information or expert judgement about specific patients. However, if this information is available, the hospital stands an even greater chance of lowering the risk of running out of beds, without reducing the planned workload, and the ODES can again be used to investigate the impact of such actions.

Finally, the Week 47 example is necessarily an artificial one, since the observed elective schedule is a record of actual patient arrivals during the week. Therefore, it is possible that hospital planners already made modifications to a *planned* elective schedule, which is unobservable in the data. Nevertheless, the example clearly demonstrates how the ODES could be used in practice.

6.3 Case Study 2: Additional Information

In the second case study, the ODES is used to answer Research Question 3, which is; "*How can additional patient information, made available at simulation* *run-time, affect the estimates of bed demand from an online simulation?"* In this case study, the additional information being considered is the length-of-stay of patients on the elective admissions schedule, and the remaining length-of-stay of any patient (emergency or elective) who has a bed when the ODES is initialised. This information is used as a proxy for the Estimated Date of Discharge (EDD) which NHS Improvement (2016b) emphasises is an essential care coordination tool within the UK.

While other types of patient information could be considered "additional" with respect to the data already used in the model, the EDD (and thus, estimated length-of-stay) aligns with the ODES method particularly well. New system state data is already read into the model at regular intervals, and this can easily be augmented with information about a patient's condition (i.e. how long they are expected to stay) as it unfolds.

For the incoming elective patients, clinicians will have approximate EDDs in mind to help manage hospital resources, and to inform prospective patients of the time they can expect to spend in hospital. Additionally, NHS Improvement (2016b) recommends that an EDD should be set at the first consultant review, and set no later than the first consultant ward round the following morning. Therefore, estimates length-of-stay should be available for the scheduled elective arrivals, and remaining length-of-stay estimates should be available for most, if not all patients occupying a bed.

However, clinician's assignment of an EDD is by no means a guarantee that the corresponding patient will be discharged on their estimated date. Factors such as variation in individual recovery times, and complications associated with treatment, can contribute to differences between the EDD and the actual date of discharge. Therefore, as part of assessing the value of using discharge date estimates in an ODES, it is also important to consider how accurate they might be.

Although estimates of LOS/EDD were not explicitly provided by the AGH participating in this study, the *actual* length-of-stay can be loaded from the PA data, retrospectively. However, using actual observations would represent a scenario in which clinicians were able to perfectly predict LOS. This level of accuracy is clearly not attainable; therefore, modelling different levels of prediction error is necessary to test the impact of using length-of-stay estimates in a more realistic way.

Since the EDD should be assigned by clinicians before, or shortly after admission, the estimated LOS which is derived from the EDD is a *total* length-of-stay (TLOS) i.e. the duration of a patient's stay in hospital from admission to discharge. The structure of the ODES is such that each simulated patient's TLOS is the sum of individual *ward* lengths-of-stay (WLOS), and each WLOS is a random draw from an empirical LOS distribution. For the sake of simplicity, the PA data which is loaded into the model (to emulate clinician's estimates) are observations of WLOS, rather than TLOS. This data is loaded for the *first ward-stay of the incoming elective patients* (to reflect clinicians' prior knowledge about scheduled arrivals) and the *current ward for patients which occupy a bed at initialisation* (to reflect clinician's knowledge about patients which are already admitted).

Adding first/current WLOS information to the ODES is simpler than adding TLOS, since there is no need to divide TLOS among a set of stochastically generated ward-stays. However, if patients are not discharged from the first/current ward, this method only adds a portion of the available TLOS information to the simulation. Nevertheless, the results of running a model in which partial information is added, can serve as a lower bound on the types of improvements one might see in a more complex model in which TLOS estimates (via EDD) can be applied, across multiple wards.

In the sections which follow, simulation experiments are conducted which aim to show the effect of additional LOS information on the distributions of midnight occupancy, thereby answering Research Question 3. In the first experiment (see subsections 6.3.1 to 6.3.3), the observed elective schedule for Week 47 is revisited, to gain an understanding of the effect the additional information can have in a given planning horizon. The elective schedule modifications made earlier for this week, in Section 6.2.3, are re-tested to see if they would still be deemed useful or if a different set of modifications should be considered. In the second experiment (see subsection 6.3.4) the full 560-day observation period is run for the purpose of estimating how much the additional WLOS information can reduce the variation seen in the simulated distributions of midnight occupancy.

6.3.1 Run Configuration

The ODES is again loaded with the state of the hospital at 00:00am on Monday of Week 47, along with the accompanying observed elective schedule. The state data is augmented with the remaining WLOS for each loaded patient (emergency and elective), and the elective schedule is augmented with the observed WLOS for the ward of admission. Table 6.4 outlines the configuration of the ODES, and the treatment of length-of-stay for patients which have a bed at initialisation (Initialised), those which are admitted during the planning horizon (New Admission), and their subsequent ward stays (Subsequent Wards). To emulate the errors that could be made by clinicians when setting EDDs, the additional WLOS information is not included deterministically. Instead it is used with a fixed probability; the details of which are described in the next section. For this reason, the term 'semi-stochastic' is used to describe its implementation. The lengths-of-stay for subsequent ward stays are drawn from empirical LOS distributions using the same treatment used in Chapters 4 & 5. The ODES is run for 400 replications of Week 47.

	Treatment	
Online Modelling Element	*Emergency*	*Elective*
Arrivals/Admissions	Empirical Distributions (Stochastic)	Observed Schedule (Deterministic)
Ward Length of Stay	Initialised: Additional WLOS Information (Semi-stochastic)	Initialised: Additional WLOS Information (Semi-stochastic)
	New Admission: Empirical Distributions (Stochastic)	New Admission: Additional WLOS Information (Semi-stochastic)
	Subsequent Wards: Empirical Distributions (Stochastic)	Subsequent Wards: Empirical Distributions (Stochastic)
Ward Transitions	Dynamic Transitions (Stochastic)	Dynamic Transitions (Stochastic)

Table 6.4: Treatment of each of the major modelling elements in the online simulation, grouped by admission type.

6.3.2 Additional Information and Observed Schedules

To model the uncertainty associated with clinicians' LOS estimates, the simulation adds random variation to the WLOS observations taken from the PA data. To add this variation, a patient's length-of-stay comes from either the WLOS observations, with probability d, or it is drawn from the empirical LOS distributions, with probability $1 - d$; resulting in 'semi-stochastic' LOS realisations. By modelling variation in this way, different levels of WLOS accuracy can be set before each simulation run. The interpretation of modelling prediction error in this way, is that the clinicians' estimates are correct $(d \times 100)\%$ of the time, and no better than guesses for the remainder.

Figure 6.5 shows the probability of the demand for beds exceeding maximum capacity on each ward for Week 47 (under the observed elective schedule) for d=0.5. Noticeable features of Figure 6.5 (compared to Figure 6.2 in which no additional information is included) are the increases in probability on Tuesday for Ward 5D, and on Wednesday for Ward 5B and the ICU. For wards such as Ward 5A, Ward 5D and the ICU, the probabilities appear to either stay the same or increase for most days of the week. As might be expected from visual inspection, BMOC also sees an increase, from 5.45 to 6.56.

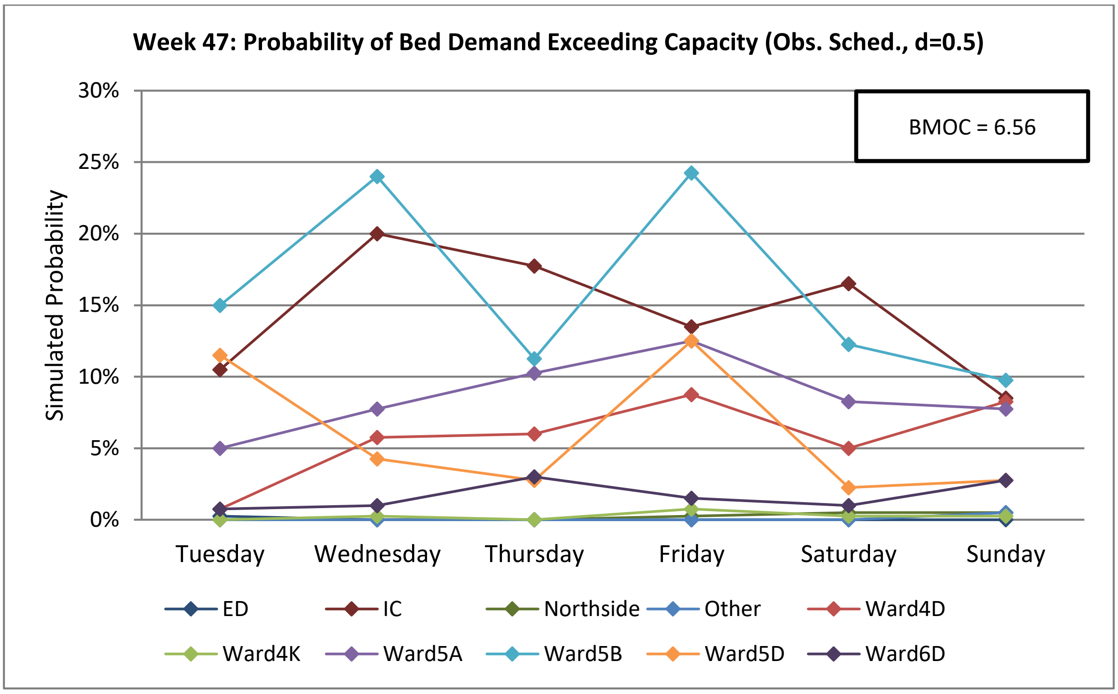


Figure 6.5: ODES estimates of the probability of the demand for beds exceeding maximum capacity during Week 47. Actual WLOS for the first/current ward for each patient is used with probability 0.5.

Similar features to Figure 6.5 are also seen in Figure 6.6 when the accuracy of the WLOS data is increased to 75%. Additionally, the probability on Wednesday for Ward 5B has surpassed the second peak on Friday; signalling that these days may require equal attention in terms of adjustments to the elective schedule. Ward 5D on Tuesday sees a further increase in the probability of exceeding capacity, making it the ward and day most affected by the inclusion of WLOS data, with 3% when d=0 (no addition information), 12% when d=0.5, and finally 19% when d=0.75. The rise in probability as a function of WLOS accuracy suggests the existence of at least one simulated patient who was (according the empirical LOS distributions) very likely to have been discharged or transferred from Ward 5D by midnight on Tuesday, but whose actual length of stay on Ward 5D was longer than the average. The early-week increases on these wards contribute to a further 0.12 increase in average BMOC, compared to results generated when d=0.5.

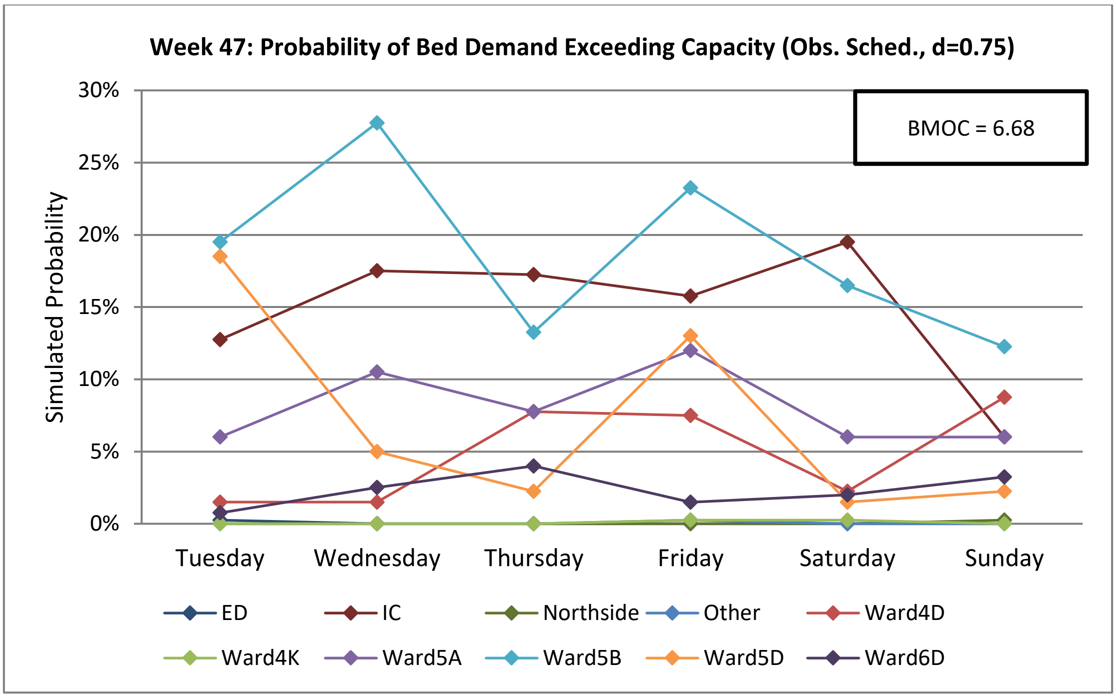


Figure 6.6: ODES estimates of the probability of the demand for beds exceeding maximum capacity during Week 47. Actual WLOS for the first/current ward for each patient is used with probability 0.75.

Since WLOS on the first/current ward is being taken from the true value in the PA data with increasing frequency, it stands to reason that the distributions of midnight occupancy should see a reduction in variance. However, the probability of demand exceeding capacity is shown to increase in some cases, indicating an upwards shift in mean midnight occupancy as well. With all other factors being held equal, an increase in mean midnight occupancy suggests that several incoming elective patients, or patients loaded at initialisation, have a greater actual LOS than the average LOS which would be drawn from the empirical distributions in the ODES, during Week 47.

Although the additional information added during Week 47 suggests that the observed schedule poses a greater risk to exceeding ward capacity than initially expected, the ODES will also be able to show situations in which the clinicians' estimates might *decrease* the likelihood of demand exceeding capacity. This would occur when the LOS information added to the model is *less* than the average LOS, for a group of patients.

By way of contrast, Figure 6.7 charts the probability of demand exceeding capacity across all wards, for the observed elective schedule in Week 53, rather than Week 47. For this run of 400 replications, no additional LOS information has been added to the model, hence d=0. During Week 53, the ODES is predicting that the probability of bed demand exceeding maximum capacity on Ward 5B is almost 35% on Wednesday; higher than any day observed in Week 47. However, this week would not be considered as busy as Week 47 overall, judging by the BMOC and the levels of actual midnight occupancy seen in the PA data.

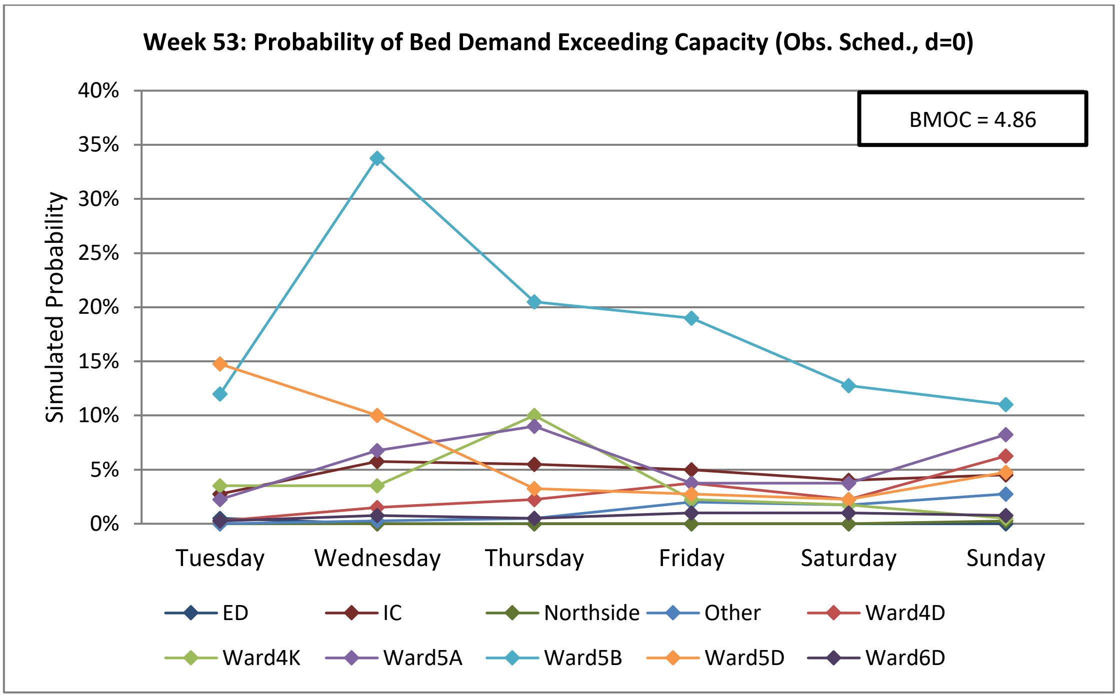


Figure 6.7: ODES estimates of the probability of the demand for beds exceeding maximum capacity for Week 53. Actual WLOS for the first/current ward is not used.

Figures 6.8 and 6.9 show the results of running the observed elective schedule for Week 53 again, this time with d=0.5 and d=0.75, respectively. Figure 6.8 shows that with 50% accuracy of the additional WLOS information, the peak in probability on Wednesday can be reduced from 34% to 26%. However, if higher levels of prediction accuracy can be achieved, Figure 6.9 shows that the probability of demand exceeding capacity is again reduced; to 21% on Wednesday. A reduction of the same size is seen on Thursday, with a 13% difference between the d=0 and d=0.75 runs. With no changes to the simulation, other than the value of d, the results of Figures 6.8 and 6.9 suggest that several patients have an actual WLOS which is *less* than the average WLOS which would have been drawn from the empirical distributions in the ODES, causing the simulated beds to become available earlier in the week.

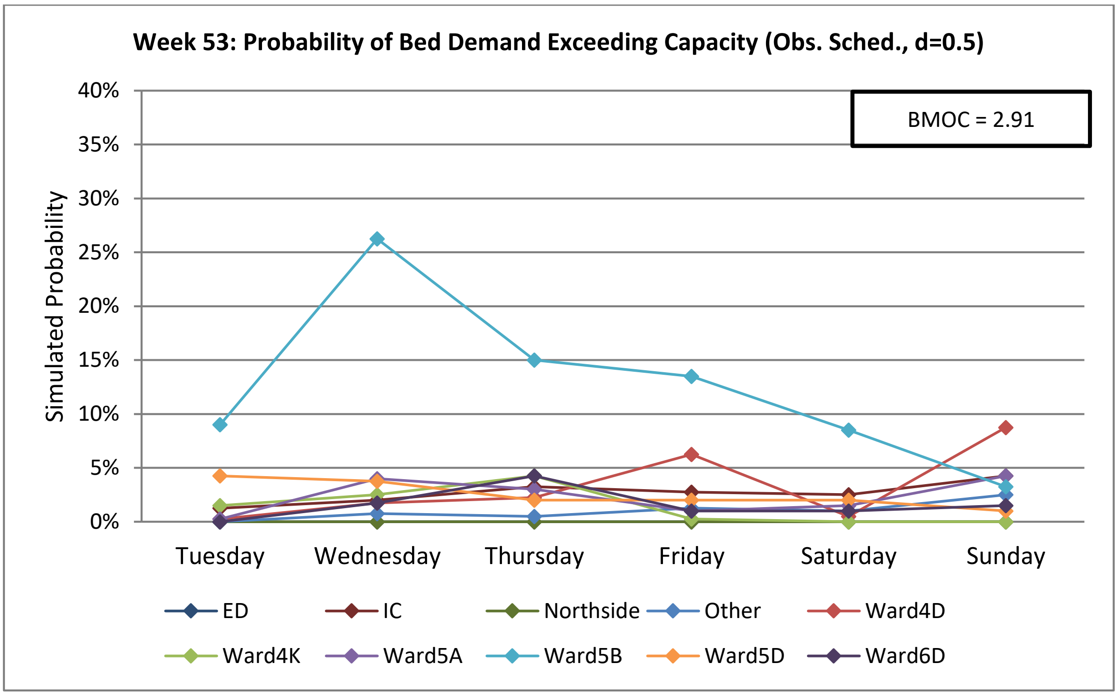


Figure 6.8: ODES estimates of the probability of the demand for beds exceeding maximum capacity for Week 53. Actual WLOS for the first/current ward for each patient is used with probability 0.5.

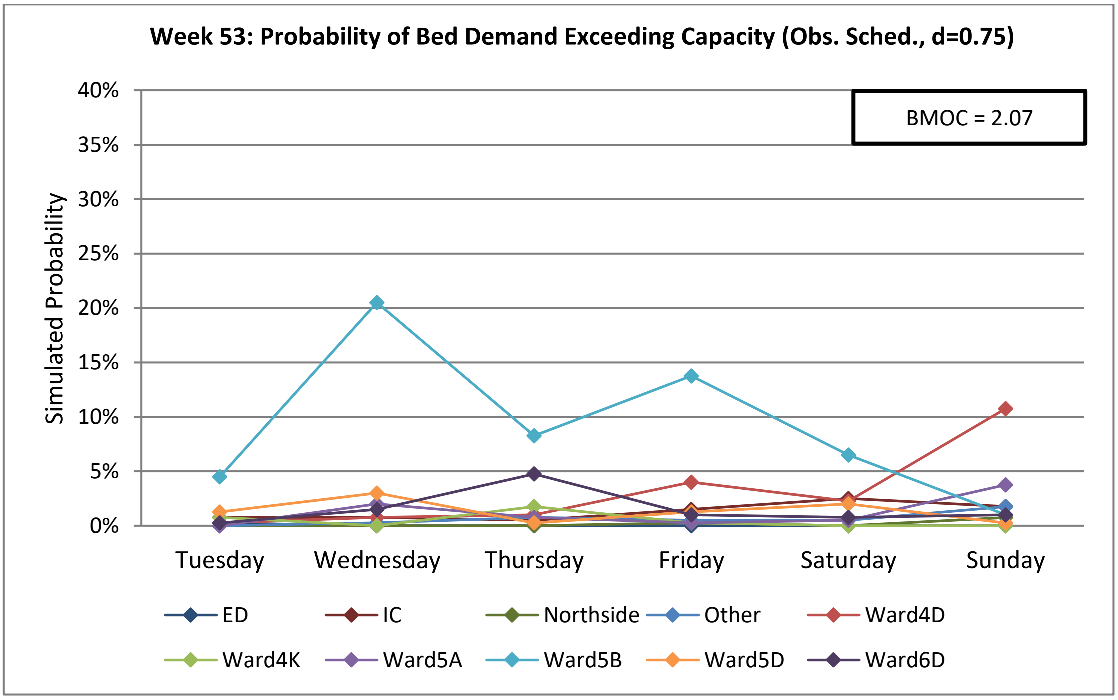


Figure 6.9: ODES estimates of the probability of the demand for beds exceeding maximum capacity during Week 53. Actual WLOS for the first/current ward for each patient is used with probability 0.75.

The results presented in this section for Week 47 and Week 53 show that by including additional LOS information, it is possible for an ODES to generate noticeably different results for a given elective schedule than if LOS is drawn at random from LOS distributions. Although it might be expected that including additional LOS information would simply reduce the variance of the distributions of midnight occupancy, the results show that the mean can also be impacted, resulting in shifts of the distributions which necessarily effect the probability of exceeding capacity.

In the examples presented, the values of d have been chosen arbitrarily, however, realistic values could be calibrated for a hospital (or even for each ward) by collecting clinician's LOS estimates and comparing those to the actual LOS once the patient is discharged. For a hospital willing to explore the use of an ODES model to help with bed management decisions, this might be particularly important for weeks like Week 53, in which several patients have unusually short stays. If LOS estimates are available, and d is calibrated accordingly, the reduced levels of risk reflected in the ODES outputs can mitigate the need for deviations from the planned elective schedule, or other types of interventions.

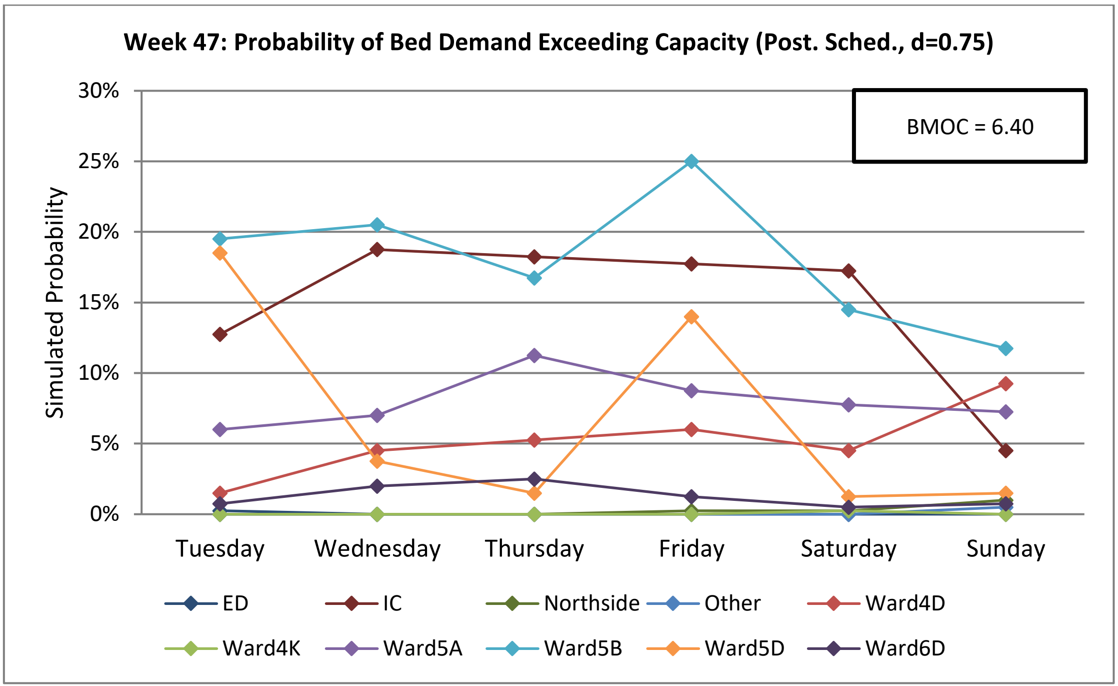
Although the inclusion of additional information for Week 53 resulted in a reduction in probability for Ward 5B, Week 47 saw the probability increase for several wards. In the next section, the modified elective schedules from Section 6.2.3 are re-run to demonstrate their effect when combined with the additional WLOS information.

6.3.3 Additional Information and Alternative Schedules

In Section 6.2.3 (Alternative Elective Schedules), two types of modifications were made to the elective admissions schedule (postponement and cancellation) to demonstrate the way in which the ODES could be used to reduce the probability of bed demand surpassing maximum capacity. However, for Week 47, the results generated by including the additional length-of-stay information indicates that the risk of running into capacity issues might be higher than is predicted by the model in which empirical LOS distributions are used exclusively. This section aims to investigate whether the elective schedule modifications used in Section 6.2.3 have similar benefits, given the additional LOS information, or if a different set of schedule modifications might be preferable.

During Week 47, the ODES model in which no additional WLOS information was read from the PA data (d=0) identified Ward 5B to be the most at risk of running into capacity issues, particularly on Wednesday and most markedly Friday. Three postponements were made; one to reduce the Wednesday peak, and two to reduce the higher peak on Friday. However, the results of running the observed schedule in the ODES model with d=0.75, indicated that the risk of excessive bed demand on Wednesday and Friday were similar, therefore the distribution of postponements throughout the week warrants revisiting.

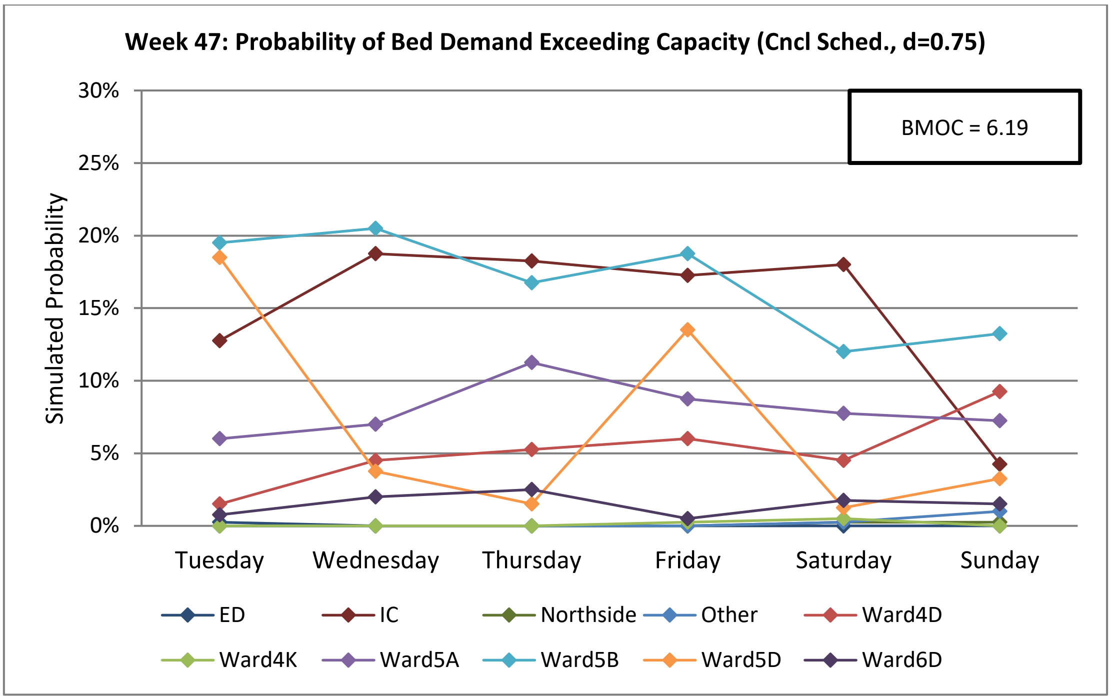
Figure 6.10 charts the probability of bed demand exceeding maximum capacity, by running the Postponement Schedule used earlier, in Section 6.2.3. The results show that with one postponement on Tuesday, it remains possible to reduce the Wednesday peak to levels seen the previous day; an outcome which is consistent with the d=0 model. However, the probability over these two days remains approximately 5% higher for the d=0.75 model than the d=0 model, due to patients with greater WLOS than the average. The two postponements which were made on Thursday in the d=0 model had the effect of reducing the Friday peak in probability on Ward 5B from 25% to 20%. However, after including these two postponements in the d=0.75 model, there appears to be little change in the probability on Friday for Ward 5B. Although these postponements appear to have little effect, they offset the Tuesday postponement which would otherwise increase the probability on Friday. They also have the effect of decreasing the Friday probability for Ward 4D and Ward 5A, resulting in an overall decrease in BMOC.



Postponement Schedule							
Ward	***Monday***	***Tuesday***	***Wednesday***	***Thursday***	***Friday***	***Saturday***	***Sunday***
ED	-	-	1	-	1	-	-
Ward4D	-	-	-	-	1	-	-
Ward4K	2	3	-	1	-	-	-
Ward5A	-	1	-	1	-	-	1
Ward5D	-	-	-	2	-	-	-
Other	6	8-1=7	6+1=7	12-2=10	4+2=6	-	-

Figure 6.10: ODES estimates of the probability of demand exceeding maximum capacity for Week 47, and the accompanying elective schedule. Actual WLOS for the first/current ward for each patient is used with probability 0.75.

Since the Friday peak for Ward 5B remains present in the results generated by running the Postponement Schedule using the d=0.75 model, the Thursday cancellation is again applied to the elective admissions schedule. Figure 6.11 charts the probabilities generated by the ODES during Week 47 by running the Cancellation Schedule, as per Section 6.2.3, in the d=0.75 model. The results show that the Thursday cancellation reduces the peak by approximately the same amount as the d=0 model and brings the Friday probability broadly in-line with previous days of the week. However, the average probability across Week 47 for Ward 5B remains noticeably higher in the d=0.75 model than the d=0 model with the same elective schedule applied.



Cancellation Schedule							
Ward	***Monday***	***Tuesday***	***Wednesday***	***Thursday***	***Friday***	***Saturday***	***Sunday***
ED	-	-	1	-	1	-	-
Ward4D	-	-	-	-	1	-	-
Ward4K	2	3	-	1	-	-	-
Ward5A	-	1	-	1	-	-	1
Ward5D	-	-	-	2	-	-	-
Other	6	8-1=7	6+1=7	12-3=9	4+2=6	-	-

Figure 6.11: ODES estimates of the probability of demand exceeding maximum capacity for Week 47, and the accompanying elective schedule. Actual WLOS for the first/current ward for each patient is used with probability 0.75.

The results presented in this section show that the schedule modifications used in Section 6.2.3 to reduce the Wednesday and Friday peaks for Ward 5B, also reduce the peaks in probability which result from running the d=0.75 model.

However, the inclusion of LOS information also indicates that the risk of encountering capacity issues on Ward 5B, with the schedule modifications applied, is higher than initially expected. Therefore, it is possible that further modifications to the elective schedule might be desirable. For example, the probability of exceeding capacity for Ward 5D on Tuesday was estimated at only 3% by the d=0 model, although it increases to 19% when d=0.75. Clearly, the inclusion of additional LOS information enables the consideration of further actions which could reduce this probability.

The changes in probability which occur as a result of including the additional LOS information (for Week 47 and Week 53) have been largely driven by shifts in the midnight occupancy distributions i.e. changes in the mean. However, including additional LOS information also has the potential to reduce the variance of the midnight occupancy distributions. Therefore, the degree to which this might occur is investigated in the next section.

6.3.4 Variance Reduction

If clinicians' LOS estimates are more informative than draws from LOS distributions, then using this type of information should also reduce the variance of the midnight occupancy distributions (in addition to potentially shifting the mean, as seen previously), resulting in more precise inferences about the planning horizon in question. While any reduction in variance is clearly dependent on the accuracy of the additional LOS information, it is also likely to be related to the elapsed time from initialisation due to the gradual discharge of initialised patients for whom additional LOS information is available.

For example, Figure 6.12a charts the average variance of the distributions of midnight occupancy for Ward 5B, for one-week planning horizons initialised on Monday. The average variance is estimated by simulating the 560-day observation period 100 times, then averaging the daily variance estimates by day-of-the-week. Figure 6.12b charts the percentage reduction in variance, relative to the results generated by the d=0 model.

Figure 6.12a shows that the additional information has the effect of reducing the variance of midnight occupancy over the average planning horizon, and Figure 6.12b shows that the information has the greatest relative effect immediately after the simulation is initialised. The diminishing effect throughout the week occurs as the initialised patients (who all have WLOS information) are discharged from hospital. While some of these discharges are replaced by elective arrivals who have LOS estimates for the first ward they are admitted to, most of the arrivals during the week are emergency patients for whom no additional information is available at run-time.

The results for the d=1 model set an upper limit on how much the variance can be reduced if additional LOS information is included. For this model, Figure 6.12b shows that the diminishing benefits of extra information are quite marked, and that a similar (but weaker) pattern exists for the d=0.75 model. For d=0.5, the diminishing effect is not apparent, and for d=0.25 the effect is also very small. Together, Figures 6.12a and 6.12b suggest that when the accuracy of the WLOS information is 25% or less ($d \leq 0.25$), the additional information has little impact on the variance of midnight occupancy.

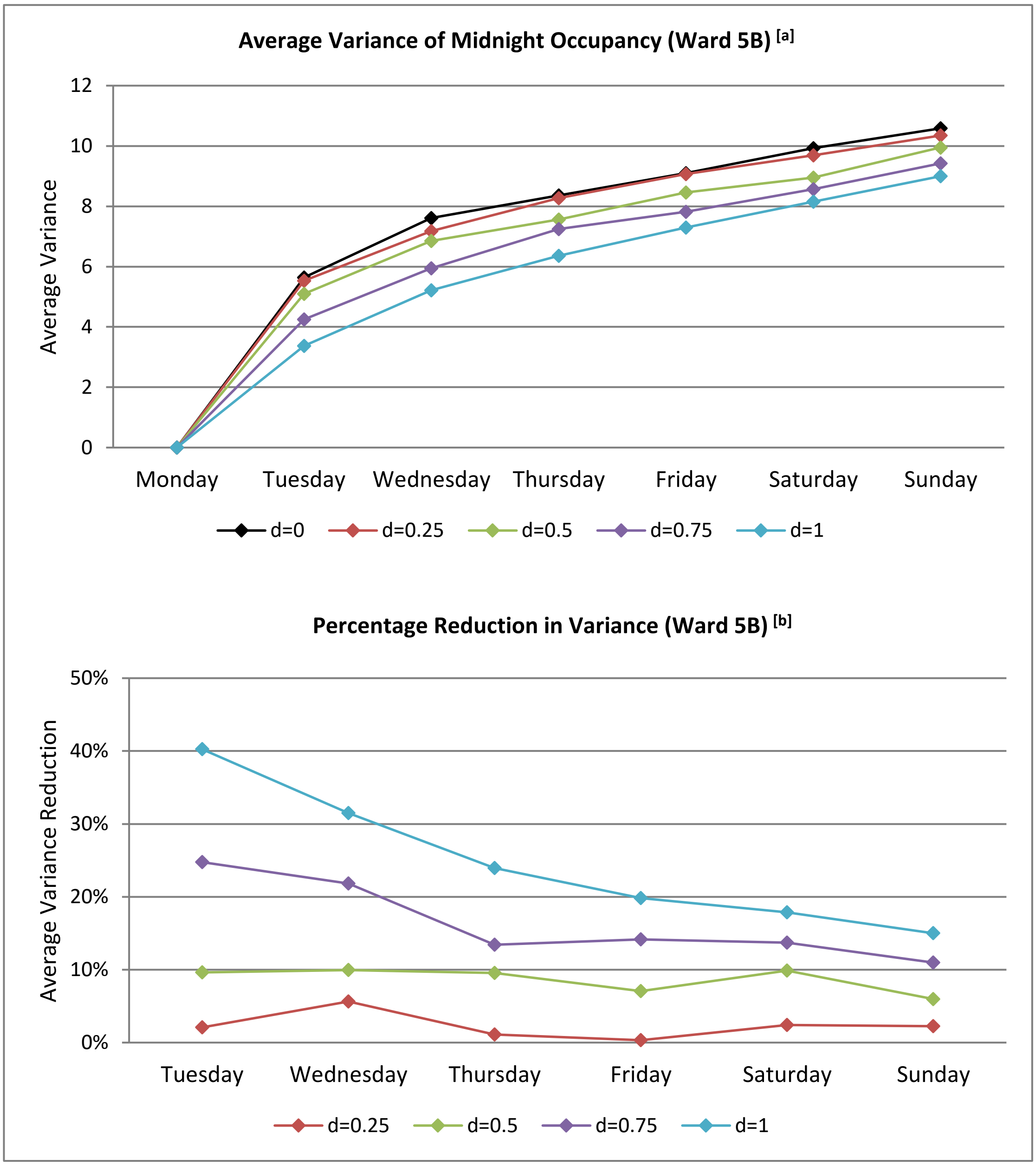


Figure 6.12: [a] The estimated variance of simulated midnight occupancy for each day in the observation period is averaged over the day of the week on which it occurs. [b] The percentage reduction in variance, relative to the results generated by the d=0 model.

For the sake of brevity, charts of the reduction in variance have not been included for all wards, since in most cases, they resemble those generated for Ward 5B. However, estimates of the reduction in variance for all wards, for an average week (initialised on Monday) are provided in Table 6.5. Summary statistics across the average week are included, rather than tabulating the estimated variance reductions for each day of the week.

Reduction in Variance for a 1-Week Planning Horizon (Average)						
Ward	**Avg. LOS**	**Stat**	**d=0.25**	**d=0.5**	**d=0.75**	**d=1**
ED	0.4	Max	-1%	4%	1%	5%
		Mean	-2%	0%	1%	1%
		Min	-3%	-2%	-1%	-2%
IC	4.0	Max	3%	10%	20%	35%
		Mean	1%	6%	13%	21%
		Min	-2%	3%	8%	14%
Ward 4D	5.3	Max	5%	11%	19%	38%
		Mean	3%	8%	13%	25%
		Min	1%	5%	8%	16%
Ward 4K	2.7	Max	3%	12%	26%	40%
		Mean	2%	9%	17%	27%
		Min	1%	7%	10%	20%
Ward 5A	4.4	Max	3%	10%	21%	40%
		Mean	1%	6%	14%	27%
		Min	-2%	2%	9%	19%
Ward 5B	3.8	Max	6%	10%	25%	40%
		Mean	2%	9%	16%	25%
		Min	0%	6%	11%	15%
Ward 5D	7.0	Max	3%	13%	23%	42%
		Mean	2%	8%	16%	28%
		Min	-1%	3%	10%	18%
Ward 6D	9.1	Max	3%	14%	25%	42%
		Mean	0%	13%	21%	35%
		Min	-1%	12%	19%	25%
Northside	10.9	Max	4%	14%	29%	53%
		Mean	3%	13%	20%	36%
		Min	2%	12%	16%	28%
Other	1.4	Max	4%	14%	28%	49%
		Mean	2%	11%	19%	35%
		Min	-2%	7%	12%	24%

Table 6.5: Average LOS for each ward (in days) computed from the PA data, along with summary statistics of the percentage reduction in variance for an average week, relative to the d=0 model.

As with Ward 5B, Table 6.5 shows that for the other modelled wards, the reduction in variance of the midnight occupancy distributions is very small when d=0.25. In fact, for this level of WLOS accuracy, some of the estimated reductions are negative. However, these instances are more likely to be the result of comparing distinct sets of simulation runs i.e. sampling error; rather than a material increase in variance.

While an appreciable reduction in variance can be seen for most wards as additional WLOS information accuracy increases, the Emergency Department is a clear exception, and appears to be relatively unaffected by the values of d in Table 6.5. Figure 6.13 expands on these results and charts the variance over an average week for the Emergency Department, again showing little difference in the variance of midnight occupancy over the course of an average simulated week.

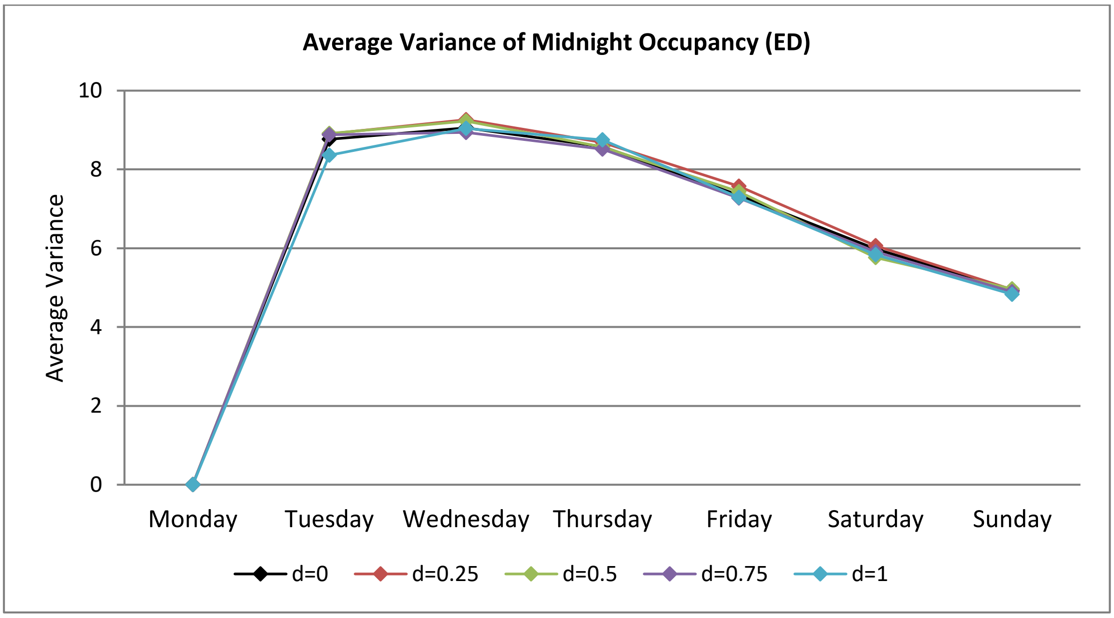


Figure 6.13: The estimated variance of simulated midnight occupancy for each day in the observation period is averaged over the day of the week on which it occurs.

Since the average LOS in the Emergency Department is short compared to the other wards, and patient turnover is high, patients who are initialised at the ED with additional information quickly move on to other wards or are discharged.

Therefore, the simulated ED reaches a state consisting of randomly generated emergency patients (for whom no additional information is included) shortly after initialisation. For these reasons, the results in Figure 6.13 suggest that clinicians' EDD/TLOS estimates do not need to include time spent in the Emergency Department, to be used in an ODES of this type.

In contrast, the summary statistics for the d=1 model in Table 6.5 indicate that the Northside ward has the greatest potential for reducing the variance of the midnight occupancy distributions. Table 6.5 also shows that the Northside ward also has the longest average LOS of any of the ten modelled wards, closely followed by Ward 6D, for which the reductions in variance are also high compared to the other wards. It is reasonable to expect that wards with greater average LOS would see greater reductions in variance, since a greater proportion of patients present at initialisation (for whom WLOS is estimable), will also be present throughout the planning horizon. On the other hand, Table 6.5 shows the Other ward to be the exception to this trend, with the second shortest average LOS, but with one of the largest reductions in variance for the values of d which have been simulated. This is caused by the contribution of the Day Case Unit; which is one of a group of wards aggregated to form Other as part of the conceptual modelling process. While the Day Case Unit has a short average LOS, it is the first ward-stay for many of the elective arrivals, therefore the additional WLOS information is used with probability d. Since many of the simulated patients with additional WLOS information arrive at Other throughout the week, it experiences a more significant reduction in variance than the other modelled wards, despite its relatively short average LOS.

6.3.5 Case Study 2: Summary

Research Question 3 has been answered in this case study, by specifying a type of additional patient information which could become available with relatively short-notice, and then testing the effect this information has on the simulation outputs, at different levels of accuracy. Sections 6.3.2 and 6.3.3 show that including additional LOS information for just the first/current ward, can result in upwards or downwards shifts in the predicted distributions of midnight occupancy, depending on how its net effect (across all resident/incoming patients) compares to the average LOS which otherwise would have been assumed. If clinician's LOS estimates are for a set of patients with greater-than-average LOS, incorporating this into the ODES could notify hospital planners that further action must be taken to reduce busy wards to an acceptable level of risk. Conversely, the same type of information for a set of patients with lower-than-average LOS could also be used to satisfy hospital planners that an otherwise high-risk schedule has risks which are within an acceptable range.

In addition to the shifts in the midnight occupancy distributions which can occur, Section 6.3.4 also showed that the variance of the distributions can be reduced, depending on the accuracy of clinicians' estimates, the ward's average LOS, and the number of elective arrivals for whom additional LOS information is available. For wards with very short average LOS, such the Emergency Department, even 100% WLOS accuracy has little effect on the on the variance of the midnight occupancy distributions, since any LOS information which is added quickly dissipates when the initialised patients are transferred or discharged. On the other hand, assuming the same level of prediction accuracy is possible, wards with higher LOS tend to see greater reductions in variance due to a larger proportion of initialised patients being present throughout the planning horizon. The number of elective admissions also plays a role in reducing the variance, as shown by the Other ward. Other has one of the highest potential variance reductions, despite a relatively short average LOS, due to the large number of elective arrivals for whom WLOS is estimable.

In terms of additional information accuracy, the results showed that if WLOS could not be estimated correctly more than 25% of the time, then the impact it has on the variance of midnight occupancy is negligible. However, if WLOS estimates can be 50% accurate, the mean reduction in variance over six days (not including the day of initialisation) for an average week ranges from 0% (ED) to 13% (Ward 6D and Northside). At 75% accuracy, the range is extended from 1% (ED) to 21% (Ward 6D).

Although the value of d has been set globally for the scenarios considered in this section, d could be calibrated for different wards and patient types. For example, there is likely to be greater uncertainty in estimating the LOS for wards where patients typically stay for longer periods. Similarly, clinicians can probably provide more accurate estimates for elective patients, than they can for those requiring emergency care. Therefore, if a hospital wanted to include information such as EDD, in an operational ODES model for bed management, it would also make sense to conduct a small study to estimate the rate at which clinicians' estimates are expected to be correct; disaggregated to an appropriate level of detail for the model. As demonstrated in this section, it is then possible to assess whether this information, with d calibrated based on evidence, has a large enough effect to become a permanent part of the operational ODES.

6.4 Discussion and Conclusions

In this chapter, two case studies have been presented which demonstrate how the ODES developed in Chapters 4 & 5 could be used in practice. In Case Study 1, a particularly busy week is chosen from the PA data, whose corresponding system state (on Monday) and observed elective schedule is loaded and run by the ODES. In the first instance, the simulation outputs are used to produce estimates of the probability of bed demand exceeding ward capacity. Based on these results, modified elective schedules are developed and tested, to demonstrate how the model could be used by hospital planners to help balance emergency and elective workloads.

Using Week 47 as an example, it is worth noting that although the ODES predicts fairly high probabilities for Ward 5B and an average BMOC of 5.45, the table of observed occupancies (Figure 6.1) shows that this ward had between one and two unoccupied beds at midnight for most days of the week. Therefore, an important part of using the ODES in practice, is to determine the probability thresholds beyond which the hospital should take preventative action; such as modifying the elective schedule. One way to estimate such a threshold for a given ward, is to run a set of historical weeks in which the ward experienced high occupancy. While this would preferably be done in consultation with hospital staff to determine what “high occupancy” looks like, it could be carried out independently by making some basic assumptions. The probabilities generated by the ODES under these conditions could then be used as indicators of high risk for future planning horizons. A similar exercise could also be carried out to calibrate an acceptable BMOC threshold, and use this as a summary metric for the risk associated with a proposed elective schedule. However, implementation details such as this, are left as further work.

In Case Study 2, additional patient level information (in the form of ward length-of-stay data) is added to the ODES as a proxy for the EDD assessments regularly made by clinicians. Research Question 3 is answered by investigating the effect that this information has on the midnight occupancy distributions. The results show that including this sort of information has the potential to shift the predicted distributions of midnight occupancy up or down, depending on how the added WLOS information compares to the average WLOS that would otherwise be assumed by the model. As a result, days in the planning horizon can be identified as being potentially problematic for one or more wards, which might otherwise have gone undetected. Conversely, the predicted probability of running into capacity issues could be reassuring if several patients are expected to have a shorter-than-average WLOS. Case Study 2 also showed that clinicians' estimates of LOS tend to reduce the variance of the predicted distributions of midnight occupancy. However, the results suggest a minimum accuracy of 25% is required (when WLOS information is incorporated for the first/current ward) before the variance is noticeably reduced.

While the structure of the ODES lends itself well to including ward-level or patient-level information, information relating to multiple wards is more challenging to incorporate. An EDD (or its equivalent; TLOS) falls into the latter category, since the total time spent in hospital potentially spans multiple wards; whereas Case Study 2 considers only first/current ward LOS information. One way to incorporate fuller information in the ODES would be for clinicians to provide an expected pathway through the hospital, as well as an EDD. In this case the ODES would need to be modified to include classes of patients whose visited wards follow a deterministic sequence.

Chapter 7
Conclusions

7.1 Summary

This research was motivated by the need for a model which could help hospital staff maintain an effective balance between emergency and elective patient bed demand. While there are several examples in the literature of models which could help to achieve this balance *on average*, models designed for use on an operational basis are scarcer. Further, the breadth of patient types within scope of this research (emergency and elective patients) requires multiple wards within a hospital to be modelled to avoid treating beds as a homogenous resource. After considering these requirements in tandem, Online Discrete Event Simulation was identified as method which could offer the flexibility to model a network of wards (and any potential interactions), while also accounting for recent events — one of the key features of operational decision making. The relative modernity of ODES as a field of research (at least compared to classical DES) also necessitated the development of novel modelling and validation techniques. In this section, the earlier chapters of this thesis are summarised, along with the answers to the research questions for each chapter, where applicable.

In Chapter 2, the hospital modelling literature is focused by defining the levels at which planning decisions in hospitals can be made. At the operational level, decision-making is impacted by the number of patients currently occupying a bed, or those who are scheduled to arrive in the short-term. Therefore, the literature review first focuses on reported models which aim to manage bed occupancy via the elective or surgical schedule, before discussing models designed to inform operational planning by taking the current bed-state into account. The review concludes that models which account for both emergency and elective patient workloads, while keeping the whole ward network in scope, are relatively few. Fewer still, are models within this class which are designed to inform operational decision-making, and an online simulation of this type is not known to exist. The chapter closes by setting out the following three research questions which form the basis of the work reported in Chapters 4, 5 and 6:

RQ1: *How can an on-line simulation, which provides estimates of bed demand, be developed for the operational management of hospital beds at the ward level?*

RQ2: *Can the effect of hospital busyness on patient-to-ward placement decisions be detected in patient administrative data, and can this be incorporated in a simulation model? If so, what effect does it have?*

RQ3: *How can additional patient information, made available at run-time, affect the estimates of bed demand from an online simulation?*

Chapter 3 discusses the participation of the Australian General Hospital who supplied the anonymised data for this research. The assumptions and data cleaning steps which ready the database are described, along with the process for deriving the real midnight occupancy time series from a database of individual patient stays. When disaggregated by ward and patient type, these series form the benchmark which the ODES is validated against in subsequent chapters. Additionally, an analysis of the cleaned data is carried out to get a better understanding of each ward, based on statistics which would typically inform the development of a simulation model.

In Chapter 4, Research Question 1 is answered by describing the development and validation process for the core ODES model. The chapter begins with a discussion of the three requirements for developing an online simulation set out by Hanisch et al. (2005), and how they could be met in the context of hospital modelling. Requirement 1 posits the need for a validated model, in which the level of detail is equivalent to the structures in the real system. While validation techniques for classical "offline" simulations are well-researched, the validation of models for *online* use appears to be an open area in the literature. This leads to a two-stage validation process. In the first stage, the validity of the offline model is assessed via comparisons of mean midnight occupancy for each day of the week. Since there is little statistical evidence to suggest there are problems with the model's parameterisation, the offline model is brought online with the ability to load specific bed-states at initialisation. Conditional LOS distributions are also added, which model patients' remaining time on the ward, given the length-of-stay already accrued at the time of initialisation. The second stage of validation is conducted by assuming a frequency at which the ODES will be run, and by using the bed-states observed in the PA database as initial conditions. Since the probability of the demand for beds exceeding maximum capacity is a key output of the model, the online validation focuses on comparisons of the *distribution* of the midnight census, rather than summary statistics alone.

Comparing the simulated and observed empirical distributions requires some aggregation of the observed data. However, the distribution of midnight occupancy on any given day is dependent on the bed-state of the preceding days. To better respect these dependencies, the Δ-occupancy random variable is defined (for use with discrete performance measures) and the observations of midnight occupancy are pooled for each day in the planning horizon, relative to the day on which the ODES is initialised. P-P plots compare the simulated and observed distributions of Δ-occupancy for each day-of-the-week, along with histograms which chart the occupancy distributions irrespective of time. While the P-P plots do not suggest any *time-dependent* differences in the Δ-occupancy distributions, the variance of simulated midnight occupancy is noticeably higher than that of that of the data for the wards which are more likely to be found near their maximum capacity (negative-skew). This is partly due to the decision to model each ward as an uncapacitated node in the simulation, and partly due to the lack of any mechanism for distributing patient load among the free beds on other wards. While the discrepancy in variance leaves room for further improvements, the steps described in this chapter formalise the process for the development and validation of ODES models. The result is an ODES for operational bed management at the ward-level, and the answering of Research Question 1.

Chapter 5 addresses Research Question 2 by investigating the relationship between ward occupancy and patient transitions between wards. Since transitions can be viewed a result of ten possible outcomes (ten wards other than the current ward, including Exit), a categorical data analysis technique (Multinomial Logistic Regression) is applied to the PA data. The fitted models for each ward and patient type show that for most cases (except ED and Northside wards for elective patients), there is a detectable relationship between transition probability and ward-level bed occupancy. Together, the systems of fitted MLR equations form the Dynamic Transition Matrices which are implemented in the simulation. The validation techniques developed in Chapter 4 are re-applied to compare the results generated by STMs and DTMs respectively, with the results showing that for eight of the ten modelled wards, the use of DTMs offers a comparable or improved fit with the empirical distributions of midnight occupancy from the PA data. DTMs are especially effective when modelling wards which are likely to be found near their maximum capacity (negative-skew occupancy distributions) which could require more frequent interventions from hospital staff. The chapter concludes that if the modeller is only interested in predicting measures of central tendency over the course of the planning horizon (such as mean occupancy on each day of the week) then STMs should be used. However, if more information is required about the distribution of midnight occupancy, such as the probability of demand exceeding capacity on each ward, then the model using DTMs provides a better overall fit with the empirical distributions of midnight occupancy from the PA data. For this reason, all simulation experiments conducted in the remainder of the thesis use DTMs as the default routing mechanism.

In Chapter 6, two case studies are presented which demonstrate how the ODES could be used in practice. In the first case study, the ODES is used to assess the impact of an observed elective schedule (and modifications to it) on the likelihood of bed demand exceeding each ward's capacity during a particularly busy week. The results of the simulation suggest that Ward 5B is the most likely ward to experience capacity problems for that week. An iterative process of modifying and re-testing elective schedules is demonstrated, and the results show that by postponing three elective patient admissions to other days during the week, peaks in the predicted probability of exceeding capacity are noticeably reduced (along with estimated BMOC) while maintaining total patient throughput. The effect of cancelling a patient is also investigated, which further reduces the peak in probability seen on Friday while maintaining low probabilities over the weekend. Other types of schedule modifications, such as changing the admitting ward for a set of patients, could also be tested using the ODES, although this sort of modification would probably require input from hospital staff to determine the most appropriate alternative. The examples in this case study demonstrate the potential for ODES models to predict the time and location of short-term problems (or other events) which might arise in complex systems such as hospitals, with the ability to test the impact of actions which could prevent them.

The second case study answers Research Question 3 by testing the effect of additional patient-level information on the predicted midnight occupancy distributions. One piece of information which aligns particularly well with the ODES method, is the Estimated Date of Discharge (EDD), which is equivalent to Total Length of Stay (TLOS) from the date of admission. EDDs are routinely set for planned and recently admitted patients, and depending on their accuracy, they have the potential to improve the bed demand estimates generated by an ODES model. In the second case study, additional LOS information is added for the first ward for the incoming elective patients, and the current ward for patients who occupy a bed at initialisation. While this approach only uses a portion of the available ward-stay information, it serves as a lower bound for the effects which might be seen in a more complex model in which EDD/TLOS is applied across multiple wards.

To represent potential inaccuracies in clinicians' estimates of the discharge day, the simulation draws from the actual ward LOS with probability d, or from the assumed (empirical) LOS distributions with probability $1 - d$. The results show that by including this information, it is possible to detect days and wards within a planning horizon which pose a greater risk than would otherwise be detected with the empirical LOS distributions alone ($d = 0$). Conversely, it is also possible that the predicted risk could be reduced by using this information, e.g. when a cohort of patients are expected to stay for a shorter-than-average amount of time. In this scenario, hospital planners could be satisfied that an otherwise high-risk schedule results in acceptable risks, or even choose to admit additional patients from short-notice waiting lists. The results also show that the inclusion of additional information can reduce the variance of the simulated midnight occupancy distributions, thereby improving the accuracy of the model's predictions. With these potential benefits, the ability to make use of patient-level information as it unfolds should be an important consideration for any ODES developed for operational bed management. In cases where this information may be subjective, such as Estimated Date of Discharge, an ODES can also be used to simulate the impact of inaccuracies in the data (as demonstrated in Case Study 2) to determine the case for its continued use.

7.2 Discussion

During the early stages of model development, Chapter 4 discusses the three requirements set out by Hanisch et al. (2005), which help to define the nature of an online simulation. The first of the three requirements plays an important role in this thesis, in part by prompting further discussion about appropriate methods for validating online models. However, the literature relating to validation methods for online simulation is scarce. In the cases where validation methods are discussed, the authors are usually referring to auto-validation methods in which the simulation parameters are periodically updated with new values based on new data. Even still, these methods are sometimes only described in passing, or with simple diagrams which show the data flows between the real system and the model.

Two publications do offer more detail. Hill et al. (2001) provide a description of the auto-validation procedure, in which deviations between the model's predictions and real world events trigger the simulation parameters to be updated, using new data to re-fit the appropriate probability distributions. However, diagnostic information about how often the simulation makes correct predictions, for example, or how the quality of the predictions change over time, is not offered.

On the other hand, Hoot et al. (2008) fully describe their online validation procedure and results, which are based on two separate analyses. The first method uses Pearson's r coefficient of correlation at 2, 4, 6 and 8-hour forecasts to indicate how much of the variation in the testing data can be explained by the simulation model. Each realisation of r is benchmarked against the autocorrelation at the same intervals from the testing data alone. For the second method, the authors conduct a residual analysis to check that the forecasts are unbiased, which averages the predictions from each replication and compares them to the real observations at the same time point and hence takes some account for the time-dependence of the simulation outputs. However, the use of Pearson's r coefficient as a summary statistic could make it difficult to diagnose problems with a simulation during early development, since it lacks any "physical" interpretation. Therefore, this approach is probably best suited to providing a simple check on a simulation that works well, rather than for diagnosing problems when it does not. To that end, the authors supplement this approach with a residual analysis, which again takes some account of time dependence. However, the results from each replication are averaged and then compared to the real data, therefore any information about the variability of the simulation outputs, for example, is lost.

In contrast, the Δ-Occupancy method, pools the simulated results and the real observations (separately) at the same time from initialisation. Therefore, it is possible to compare the variation using this method, and both empirical distribution functions, if desired. However, it is worth noting that the Δ-Occupancy method (as described in Chapter 4) currently relies on visual inspection of the P-P plots to make inferences about the quality of the fit between the simulation and the real data. If this approach was to be operationalised as part of an auto-validation component, a method for quantifying the quality of the fit at each time point would be required, possibly via standard statistical tests such as Pearson's chi-squared test, or Kolmogorov-Smirnov tests.

Although the first part of Requirement 1 (Hanisch et al.) has triggered the validation developments in this thesis, the second part of this requirement (which specifies that the level of detail of the model must be equivalent to the structures in the real system) seems to be more questionable. A literal interpretation of this would mean modelling the minutiae of ward-level processes, which is neither pragmatic nor feasible given the data available. Instead a more practical approach to determining a model's structure has been adopted, which recognises that all models are abstractions at some level of detail. In this way the, level of detail is informed by the questions the model aims to answer, rather than attempting model every possible subprocess. For the ODES reported in this thesis, that means modelling the largest individual wards (structural level of detail) and observing occupancy levels once each day at midnight (temporal level of detail). So, while Requirement 1 is a useful starting point for thinking about ODES development, a more lenient interpretation of structural equivalence is recommended for most, if not all applications.

In Chapter 5, Multinomial Logistic Regression is used to model the relationship between ward occupancy and transition probability, and subsequently derive Dynamic Transition Matrices. While MLR has been shown in Chapter 5 to be a useful method for modelling ward transitions, other methods exist for addressing the so-called Multiclass Classification problem, which ward placement is surely an instance of. These alternative methods predominantly reside within the field of Machine Learning/ Data Mining. One of the benefits of using MLR, is that if none of the explanatory variables are significant predictors of transition probability, the model reduces to a set of fixed probabilities which sum to unity i.e. the i^{th} row of the Static Transition Matrix. This is important, because even if no relationship is detected, average patient flows will be maintained, and continue to be modelled in a stochastic way. For this reason, MLR is preferable to machine learning methods which classify outcomes deterministically.

There are also machine learning methods which classify outcomes probabilistically, such as Multinomial naïve Bayes. However, a naïve Bayes formulation requires that every combination of the dependent and explanatory variables ("classes" and "features" respectively, in machine learning parlance) for which probabilities are estimated, occurs at least once in the data used to fit (or train) the model. This is unlikely to occur in practice, since ward occupancies are integer-valued. MLR on the other hand, will simply interpolate between missing values of the ward occupancies, with no extra adjustment to the fitting procedure.

Some more advanced machine learning methods, such as Multiclass Multilayer Perceptrons, will also be able to probabilistically assign patients to wards based on occupancy data. However, implementing this type of technique is likely to require a specialist user who can code it "from scratch", since at present, the overlap between discrete event simulation software and data mining/machine learning software is limited (although open source packages for both are available in Python). In contrast, an MLR implementation is essentially a system of linear equations, which is straightforward to code, and is generally understood by a modeller with a reasonable grasp on applied statistics.

After fitting the appropriate MLR models and implementing the DTMs, Chapter 5 uses the Δ-Occupancy method to comment on the accuracy of the results generated by the DTM model, and to compare them with the results generated by the STM model. Although DTMs are shown to perform well for most of the wards, the Intensive Care Unit highlights some limitations that remain with this approach. The widening disagreement between the simulated and observed occupancy distributions under DTMs suggest that ward occupancies alone are not enough to help explain the distribution of midnight occupancy seen in the data. For cases like this, methods which include closer consultation with bed managers, such as the priority lists developed in Harper and Shahani (2002), could reveal bed placement procedures which are difficult to detect without any input from hospital staff. Therefore, a mixed approach, in which data-driven methods (e.g. DTMs) are used alongside special-case routing procedures informed by expert opinion, might provide a better fit with the data for some aspects of the model.

7.3 Conclusions and Further Work

This thesis has reported the development of a proof-of-concept ODES which aims to help hospitals balance emergency and elective bed demand within operational planning horizons. Both the conceptual model and the parameterisation of the model’s components are data-driven; meaning a comparable model could be developed for any hospital collecting similar types of data. Additionally, the data requirements are thought to be straightforward to satisfy, as the ODES is developed using information which is likely to be collected in most modern hospitals.

As a result of developing the model, this thesis lays out an overall approach for ODES development and validation. Much of this approach is a natural extension of DES development, however model validation is a more substantial challenge. While validation techniques for classical or "offline" simulations are well-established, research concerning the validation of *online* models is much less commonplace. This thesis contributes to the theory of online simulation by developing a so-called "black-box" method which can be used to validate simulations in an online way, where the simulation outputs are treated as time-dependent. While this is not the first online validation method to do so (Hoot et al. (2008) also account for time), this method goes a significant step further by carrying out the validation based on empirical distribution functions, rather than reducing the comparison to summary statistics. If for example, the modeller's primary interest is in measures of central tendency over time, the difference may not be important. However, if more detailed features of the simulated distributions are important, as in this application, the Δ-Occupancy method (or Δ-Metric method more generally) can help to achieve this.

This thesis contributes a second technique to the discrete event simulation "tool-kit" by demonstrating how Multinomial Logistic Regression can be used to detect and model the relationship between server busyness and transition probabilities. As Chapter 5 demonstrates, the ward-level distributions of midnight occupancy are noticeably improved under DTMs compared to STMs (for most wards), in terms of their agreement with the historic data. Additionally, the wards which benefit most from this approach are the wards most frequently found near their maximum capacity (i.e. with negative skew real occupancy distributions). However, the relationship between busyness and entity routing is not unique to hospitals, and parallels can be drawn with systems where customers can switch servers if they think they might be served sooner (also known as "jockeying"). In more simplistic cases where the servers are homogenous resources, or the deferral rules are known, modelling jockeying behaviour may not necessitate an MLR-based analysis. However, in more complex cases, such as hospitals, simple deferral to the emptiest server (or ward) is not always appropriate, and alternative routing rules can be unclear. In cases of this type, a data-driven way of obtaining this information reduces the modeller's reliance on subjective sources such as expert opinion (which may also be prone to obsolescence), or assumptions in the absence of anything else.

Although Chapter 5 showed that DTMs can be fitted and implemented with positive results, it is also worth noting that their use can affect the traceability of patient transitions. For example, under STMs, it is a straightforward task to understand where the largest patient flows to a particular ward are coming from. However, when the probabilities are no longer fixed, this task becomes more challenging. Littig and Isken (2007) convey a similar sentiment, noting that their predictive occupancy model (POM) is essentially a "black-box" due to aggregating the results of many statistical sub-models. This situation can be difficult to work with, especially in the context of debugging. One useful method is to add tracing variables to the simulation which can help tackle parameterisation or debugging issues. For example, a "PreviousWard" entity attribute was used in the ODES during early development to better understand differences in patient flows between STMs and DTMs, in instances where the model was thought to behave unexpectedly. This helped to understand where the new sources of bed demand on each ward were coming from. Hence, if a dynamic routing procedure is deemed to be an advantageous modelling approach, then liberal collection of diagnostic information should also be considered.

The second case study in Chapter 6 answers the third research question by investigating the impact of additional patient information, such as Estimated Date of Discharge (EDD), on the estimates generated by the ODES. Since real estimates from hospital staff are not present in the patient administrative data, patients' actual length-of-stay, and an accuracy parameter d, are used together to create proxy LOS estimates, whose reliability can be adjusted to reflect the subjective nature of setting EDDs. The results show that clinicians' estimates could play an important role in improving the estimates generated by an online model, by allowing the detection of days and wards which pose a greater risk than would otherwise be detected using the LOS distributions fitted from historical data. Conversely, this information can also reduce the predicted risk of encountering capacity issues, if for example, hospital staff correctly estimate a shorter-than-average LOS for a cohort of patients.

Chapter 6 also examines the relationship between the accuracy of the additional LOS/EDD information and the estimates generated by the ODES. In a series of simulation experiments, the accuracy parameter d is increased from 0% to 100% in 25% increments, to represent the reliability of the LOS/EDDs provided by hospital staff. The accuracy of the simulation outputs is assessed for each value of d, by calculating the variance of the realisations of midnight occupancy. The results show a clear relationship between average LOS and variation, with higher LOS wards seeing greater variance reductions, for the same value of d. This is due to a larger proportion of initialised patients (for whom additional LOS information is available) being present throughout the planning horizon, due to their longer stays. The results also suggest that if hospital staff cannot correctly estimate LOS/EDD more than 25% of the time (approximately), then using this information has a negligible effect on the variance of the midnight occupancy distributions, and hence the model's accuracy. Interestingly, this observation appears to hold true for all modelled wards, regardless of average LOS.

If real estimates of LOS/EDD from hospital staff are to be incorporated in an operational ODES model, it makes sense that realistic values of d should be chosen to model the instances where staff make incorrect predictions. To better understand the accuracy of EDD assignment, Ou et al. (2011) compare EDDs with ADDs (Actual Date of Discharge), using data collected at a tertiary referral centre in Australia. Although differences in scope (and obviously collection site) mean their results are not directly comparable with this research, the authors find that on general wards, 46.5% of patients are discharged on the date of their original EDD (the analysis does not include EDD revisions). For stays of up to seven days, EDDs are reported to be correct 63.2% of the time, which is an encouraging statistic when compared to the minimum required accuracy of 25% reported in Chapter 6. However, as might be expected, EDD accuracy sharply declines as LOS increases, which suggests that in practice, calibrating d at the ward level should be considered if average LOS is known to differ by ward.

7.3.1 Further Work: Implementation

While the ODES model reported in this thesis has the potential to help clinicians maintain an effective balance between emergency and elective workloads, it remains a proof-of-concept which is implemented in specialist discrete event simulation software. To be used in an operational environment by non-specialist users, steps would need to be taken to improve user-friendliness. An important first step would be the inclusion of a graphical user interface for entering the details of a proposed elective admissions schedule and any other patient-level details, such as EDD. Additionally, a system for collating the midnight occupancy realisations into an easily interpretable results dashboard is also required, and the contents of such a dashboard would need to be decided in consultation with hospital staff. The model might also require re-coding in open-source software to facilitate wider adoption.

Another area which would need to be addressed, is the nature of the connection with the “live” hospital database which allows the initial conditions of the ODES to be loaded at run-time. While a direct software connection would certainly facilitate ease-of-use, this could require the development of software so that the patient database and the simulation can communicate with each other. Different hospitals might also use different brands of database software, potentially adding to the challenge. However, with the time-scales on which the simulations run (days/weeks), the ability to instantaneously query the hospital database might not be as important as it is in other applications. One solution could be to read the current bed-state from a formatted text file. While this requires an extra step by the user (exporting the file), the ability to generate a delimited text file from a saved query is a common feature in many types of database software.

Although the ODES has been *parameterised* to a specific hospital, further work is required to *calibrate* the model for operational use. For instance, the Week 47 example in Chapter 6 indicated that the demand for beds could exceed available capacity by 5.45 bed-midnights (the average over 400 simulation runs) over the course of the week, using the elective schedule observed in the PA data. This is possible due to the "soft" maximum capacities applied to the uncapacitated wards, but in practice these patients would become outliers. The Dynamic Transition Matrices contribute towards a more realistic representation of outlier patient placement; although it is not possible to completely model a process of this complexity. Therefore, a comparison of high occupancy in the simulation versus high occupancy in the real hospital should be carried out (preferably in consultation with hospital staff) to calibrate the thresholds at which preventative action should be taken. In addition to calibrating occupancy-based thresholds, the accuracy of any additional (and potentially subjective) information should also be calibrated.

7.3.2 Further Work: Research and Development

In terms of the technical aspects of the ODES, there are a few areas which could benefit from further work and development.

In addition to the "black-box" validation methods which have been used to investigate the model's performance, Pidd (2009) also suggests the use of "open-box" validation, where the structure and components of the model are discussed in partnership with real-system experts, to confirm it can address the issues they are concerned with. The structure of the ward network, described in Chapter 4 as part of the conceptual modelling phase, could potentially benefit from this type of validation. An entry criterion (highest average occupancy) was used to select a set of modelled wards, which is a pragmatic solution based on the contribution of each ward to the performance indicator of interest (midnight occupancy). However, some verification of the resulting structure should also take place with hospital staff before the model becomes operational. Wards such as the Day Case Unit, which have high patient throughput but low average occupancy, will not be selected for individual modelling by an occupancy-based entry criterion, although they could be relevant to hospital staff for other reasons. Further consultation as part of an open-box validation process could ensure a mixture of high occupancy wards and wards of special significance, are included in the final ODES.

Although auto-validation methods have not been the focus of this research, an appropriate auto-validation component (of the type described by Davis (1998)) would be required in operational version of the ODES, to periodically update the simulation parameters. Although models exist in the literature which reportedly contain such components, such as Hill et al. (2001), the theoretical aspect of their development remains a relatively open area of research. Open issues include determining when simulation parameters should be updated, and how these "update rules" relate to the various components of a discrete event simulation; such as arrival patterns, service times and transitions between nodes. The functionality of an auto-validation component is also likely to depend on how the ODES connects with the real-system, so these aspects should be considered in parallel.

The patient type classification (emergency/elective) in the model is intentionally coarse; focusing on the detail of patients' locations, rather than their speciality or treatment group. While this is a useful starting point for illustrating how emergency/elective workloads could be managed, there remains scope for further disaggregation into finer patient groupings. By employing the type of classification analysis (CART) proposed by Harper (2002), or another system of classification already used by the hospital, there is potential to reduce the variance of the simulated midnight census even further. However, this type of disaggregation would only be meaningful for the patients on the elective admissions schedule, or patients who occupy a bed at run-time. For these patients, the benefits of disaggregation could be realised through more accurate length-of-stay estimation, or by narrowing the choice of subsequent wards, or both. However, it is important to note that disaggregating the simulated emergency arrivals is not expected to offer any improvement over the current approach, since these patients cannot be classified to a patient type at initialisation.

There are authors (Fetter and Thompson, 1969, Chow et al., 2011) who argue that modelling patient transitions based on the current ward oversimplifies the patient transition process and does not preserve within-patient correlations between patient type, length of stay and patient pathway. Both authors adopt a so-called "trace-driven" approach, which samples an entire patient pathway from historical data; thereby maintaining these relationships. While the authors' criticism may be justified, the trace-driven alternative cannot model conditional events (such as the occurrence of outlier patients) since the sequence of wards is fixed for each simulated patient. However, it may be possible to strike some middle ground by adding different types of explanatory variables to the DTMs, in addition to ward occupancy. Categorical variables, such as previously visited wards, could be added to see if they have any significant relationship with transition probability; allowing more complex within-patient dependence structures to be modelled. This usage highlights another strength of the MLR-based approach, since within-patient and between-patient effects on transition probability can be modelled alongside each other.

In Chapter 6, additional LOS information is only applied to the first/current ward because patient pathways through the hospital are stochastically generated. This make it difficult to split an estimate of total LOS or date of discharge across a sequence of wards which is only partially known when the estimate becomes available. However, clinicians might also have an idea of the wards each patient can expect to visit during their stay, to which an EDD can be applied. While this level of detail has not been assumed to exist for the simulation experiments in Chapter 6, further consultation with hospital staff could be carried out to assess the feasibility of collecting this information as well. As with EDDs, indications of ward placement are not expected to be perfectly accurate, and the ODES could again be used to model the impact of these inaccuracies.

Finally, it has been that the impact of alternative schedules could be investigated in an iterative way by hospital staff, similar to the process described by Vanberkel et al. (2011). However, the structure of the schedules also lend themselves to being formulated as mathematical programmes. If constraints on the feasible postponements, cancellations or other types of patient swaps could be determined in consultation which hospital staff, an optimisation component could be added to the ODES to expedite the search for a schedule which minimises BMOC or the probability of demand exceeding maximum capacity. Admissions planners would naturally make the final decision as to whether the “optimised” schedule should be followed verbatim, or if further adjustments should be made, however a component of this type could speed up the search. Additionally, since online simulation and simulation-optimisation are both relatively new fields of study, considering the intersection of the two might yield new areas for research.

Appendix A
Appendix to Chapter 4

Figure A.1.1: Emergency Department

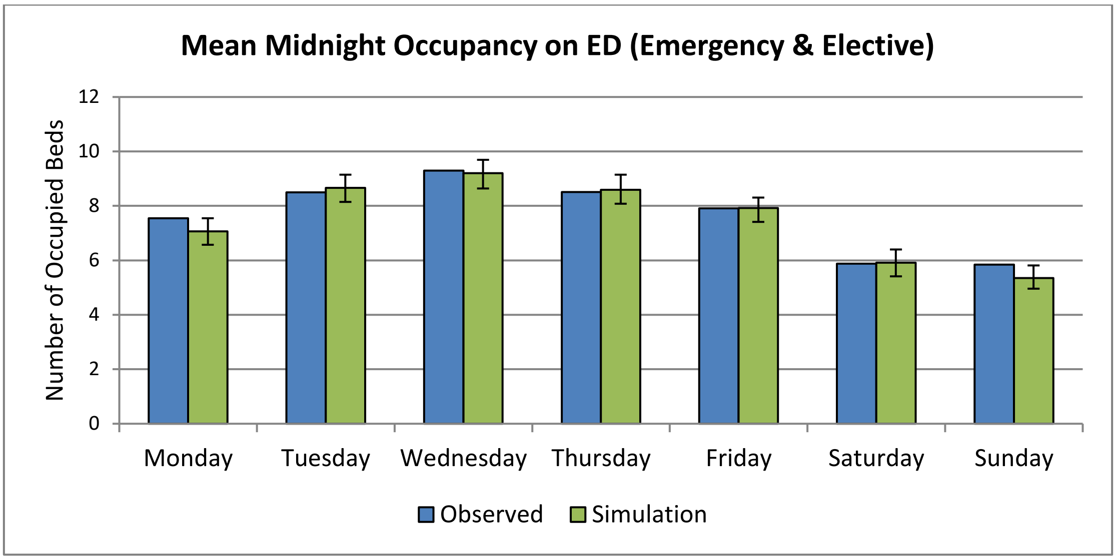


Figure A.1.2: Intensive Care

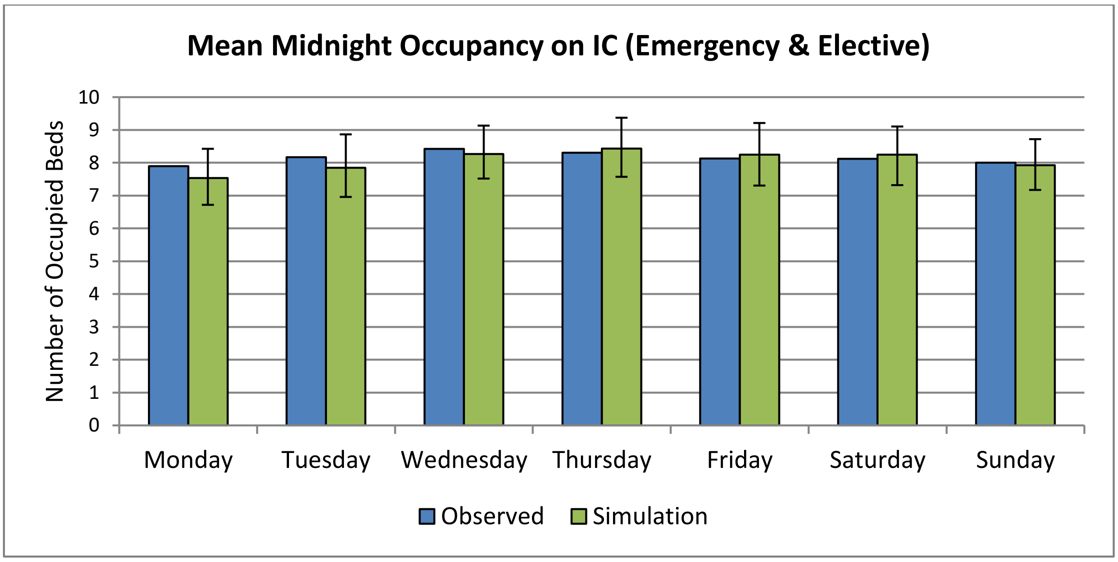


Figure A.1.3: Ward 4D

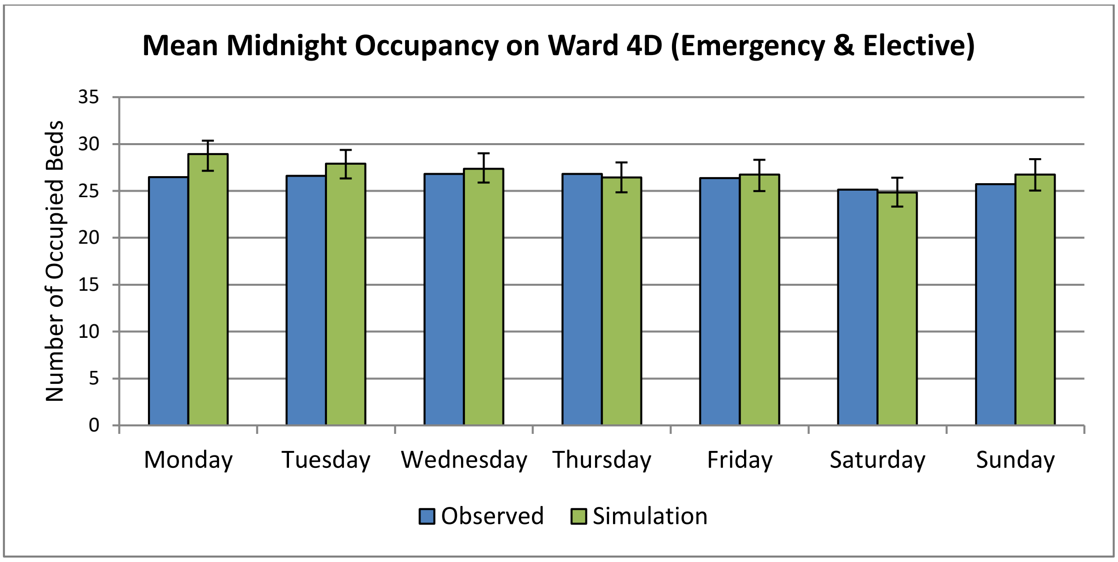


Figure A.1.4: Ward 4K

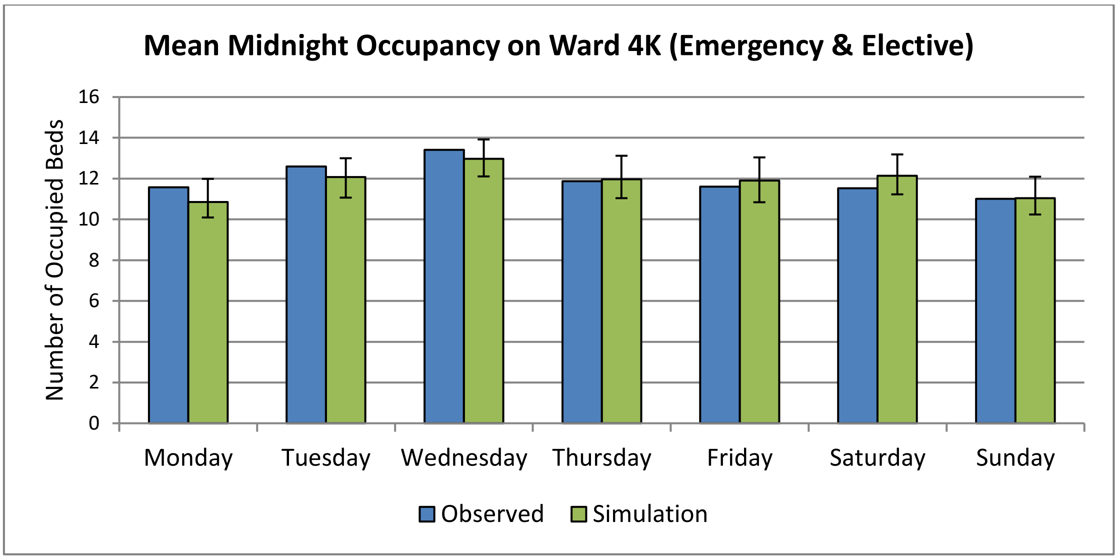


Figure A.1.5: Ward 5A

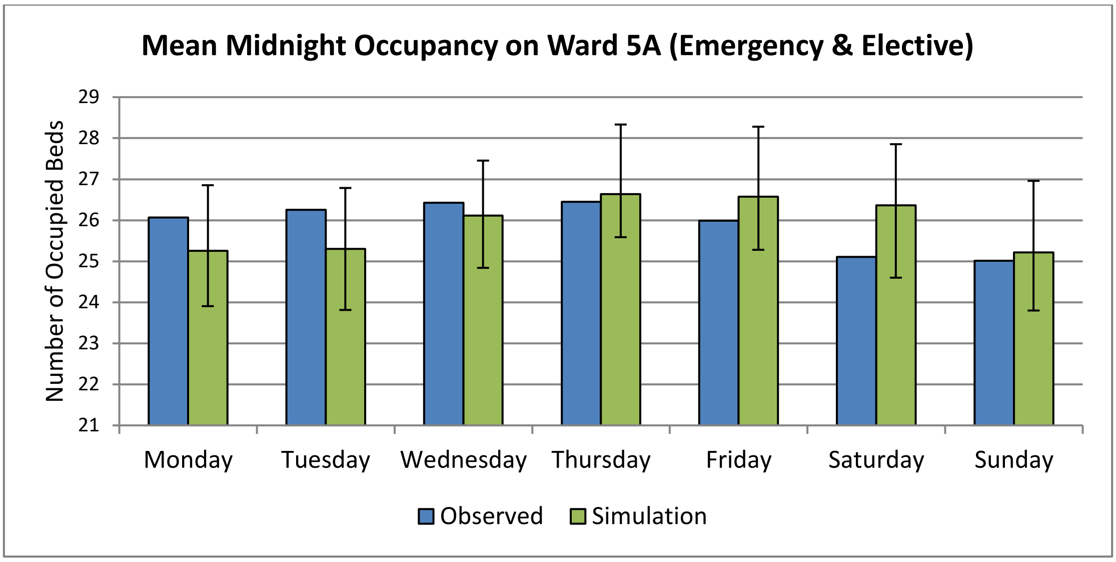


Figure A.1.6: Ward 5B

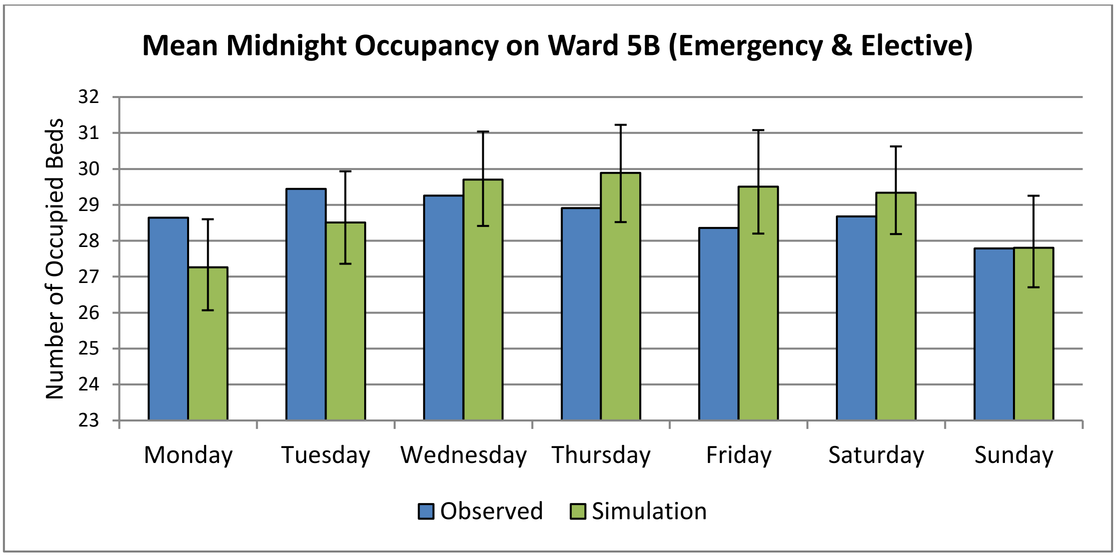


Figure A.1.7: Ward 5D

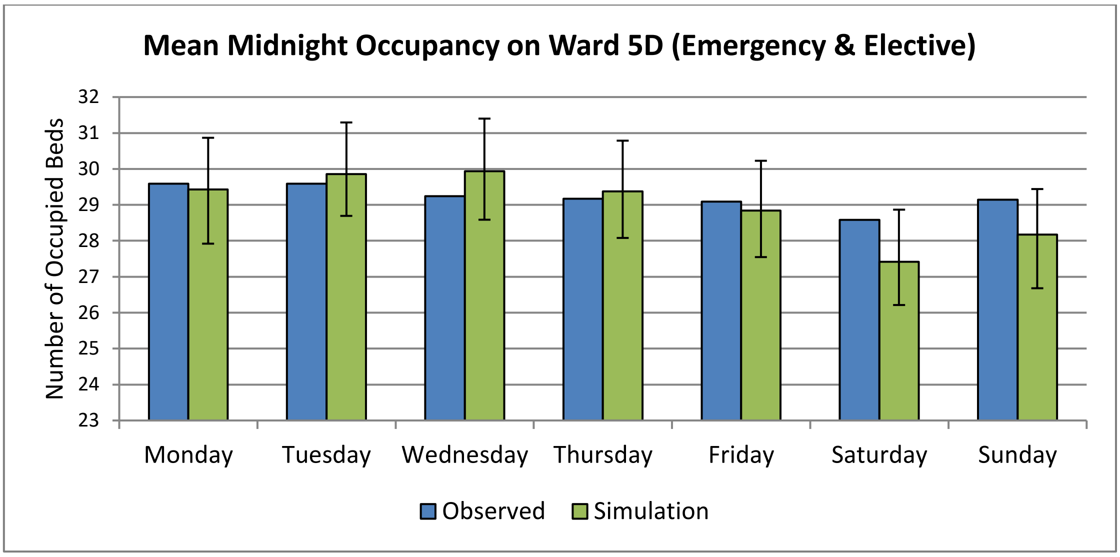


Figure A.1.8: Ward 6D

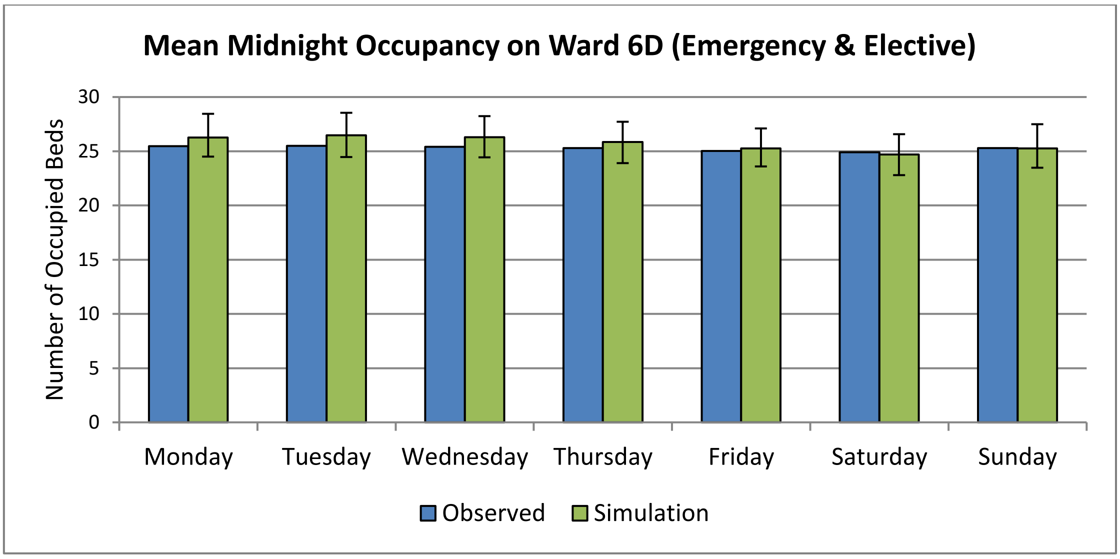


Figure A.1.9: Northside

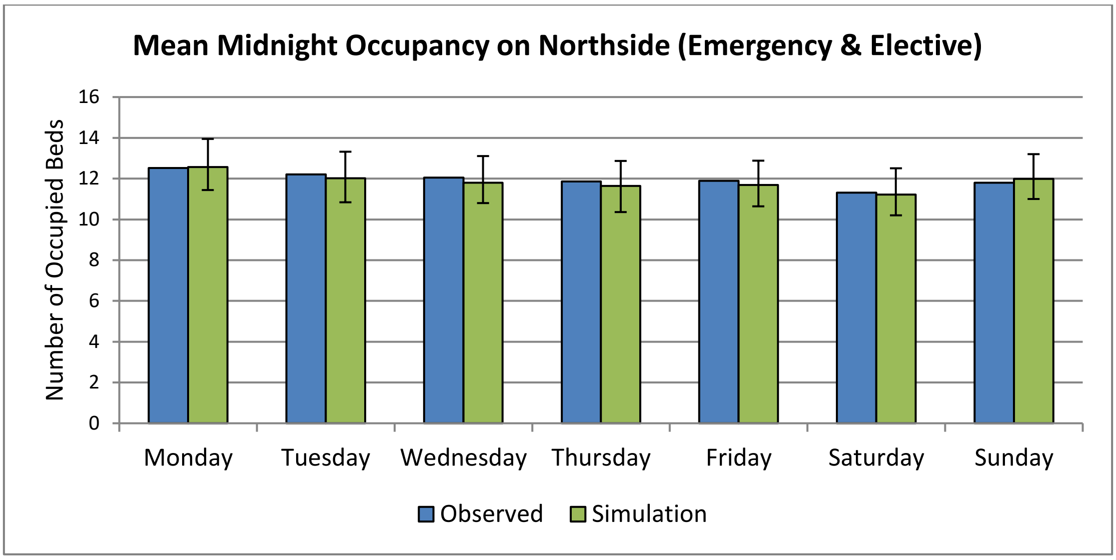


Figure A.1.10: Other

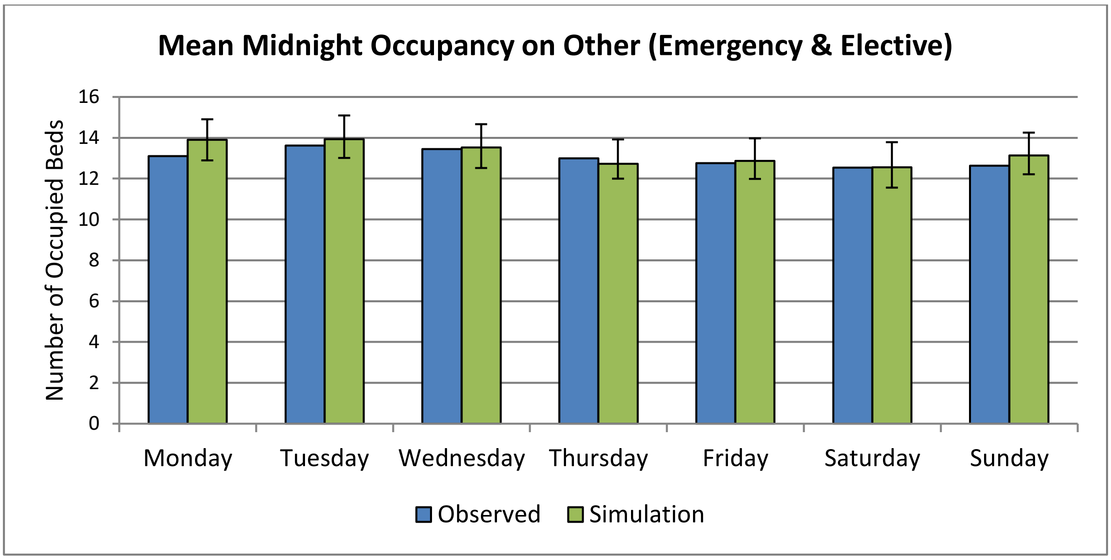
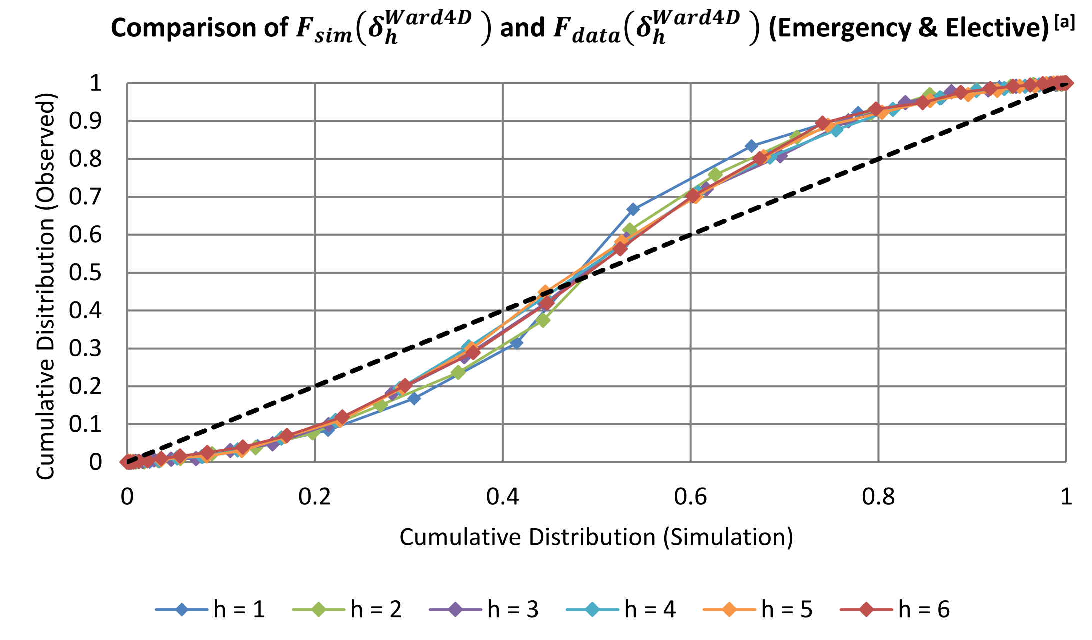


Figure A.2.1: Ward 4D



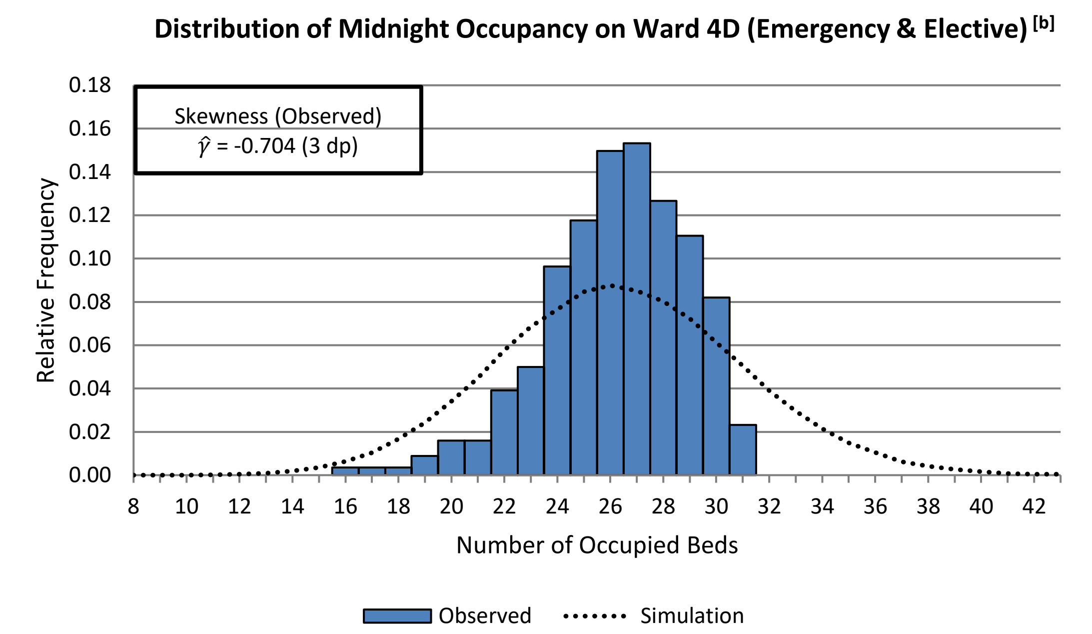


Figure A.2.2: Ward 5B

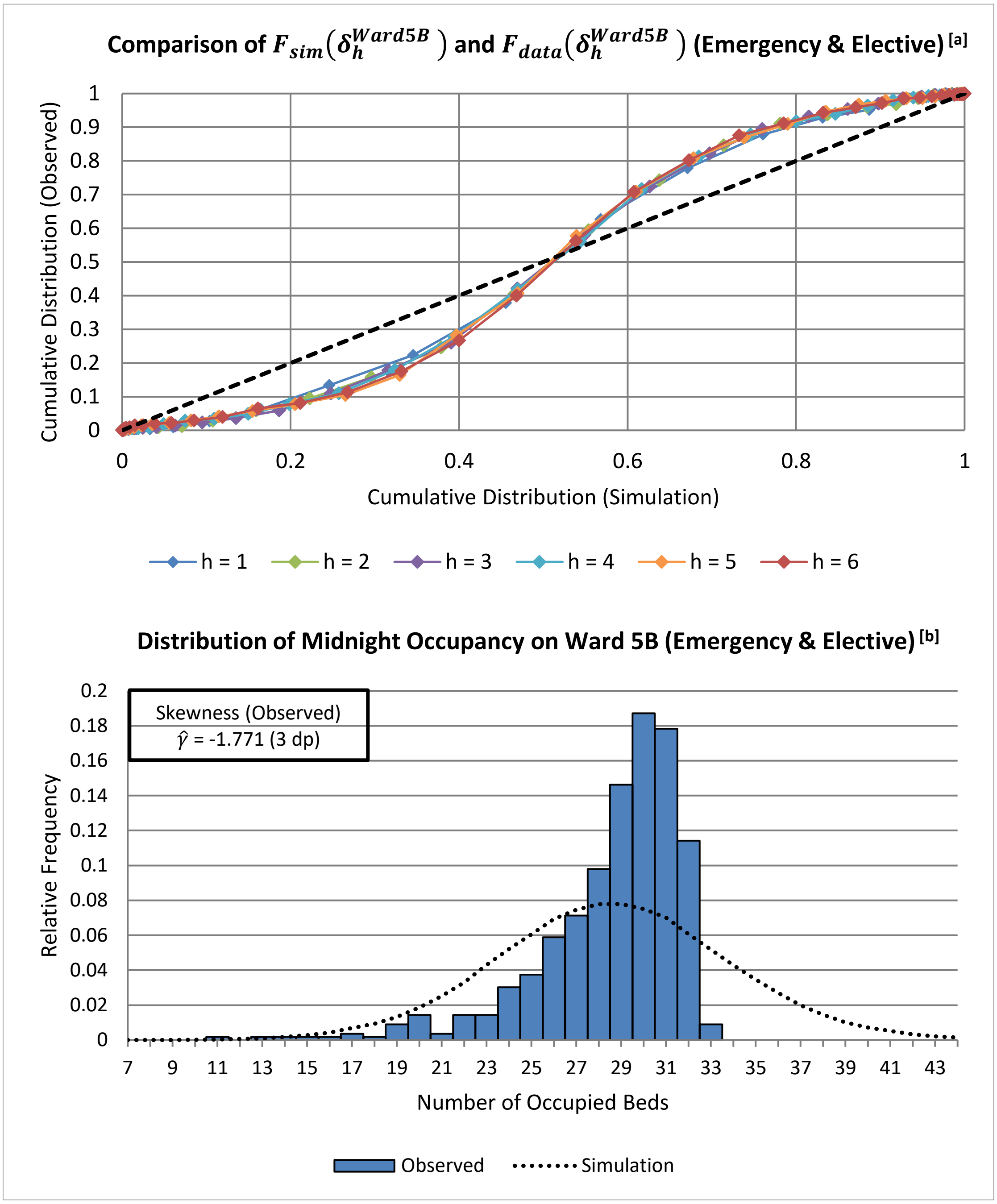


Figure A.2.3: Intensive Care

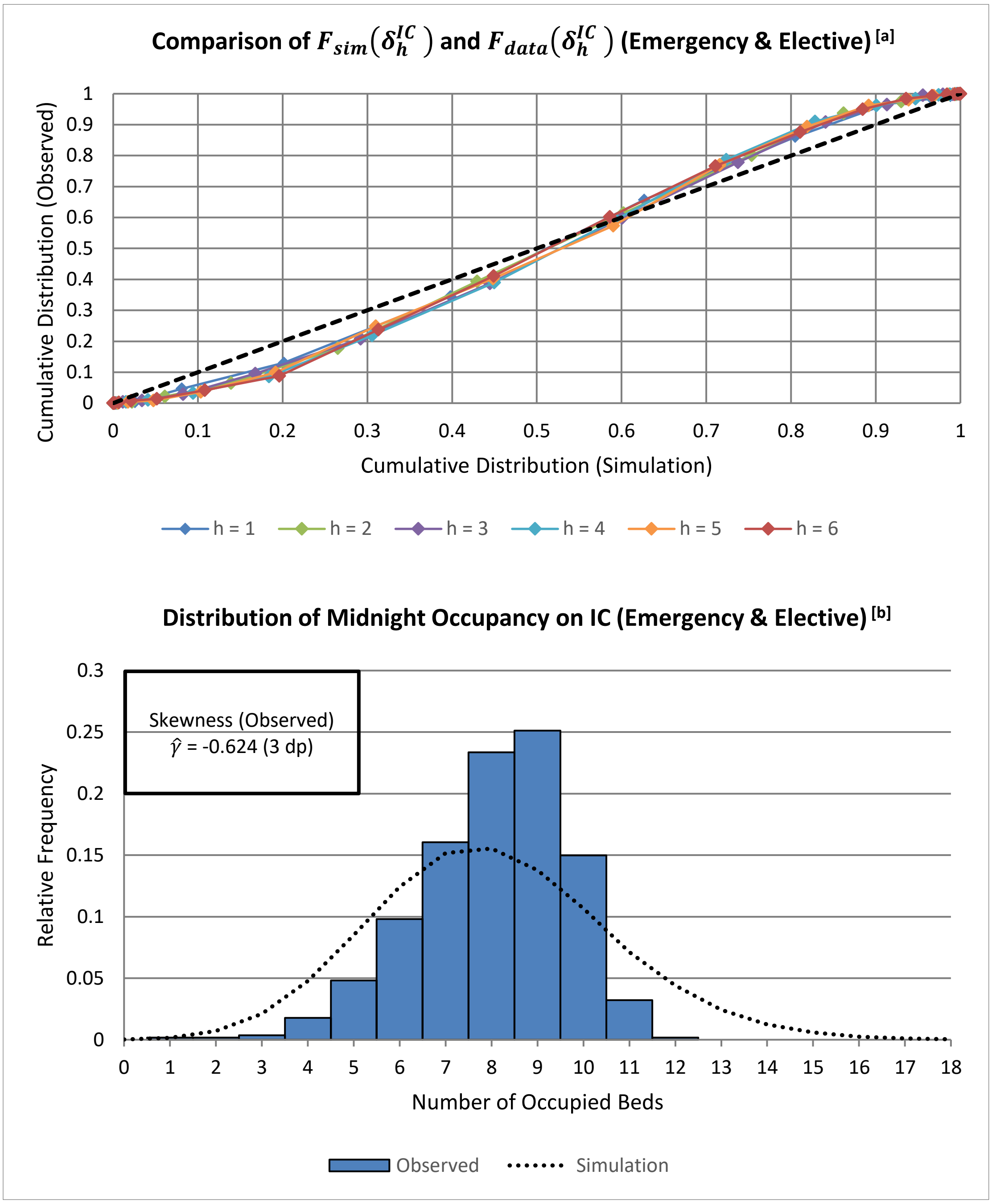


Figure A.2.4: Ward 5A

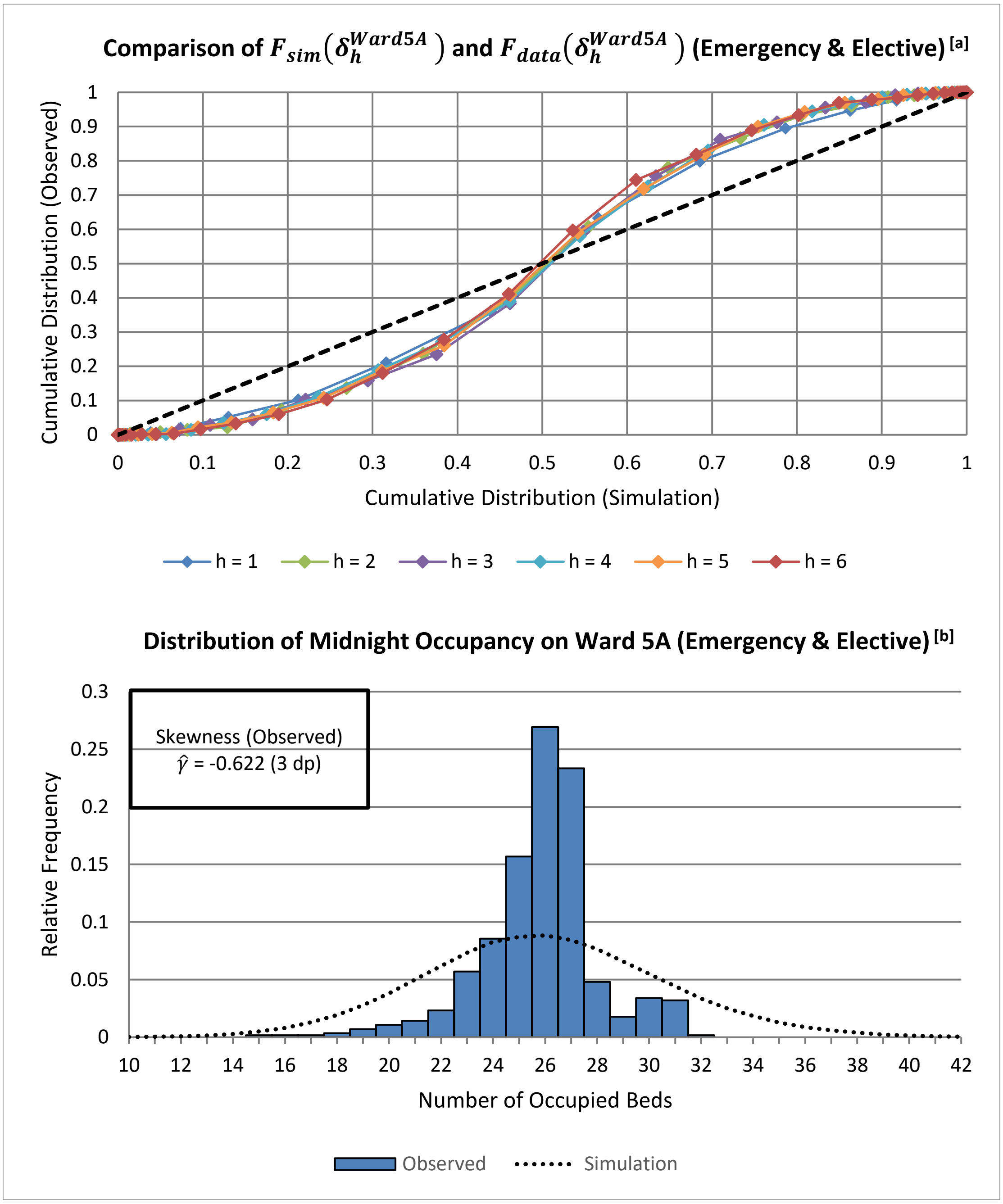


Figure A.2.5: Ward 6D

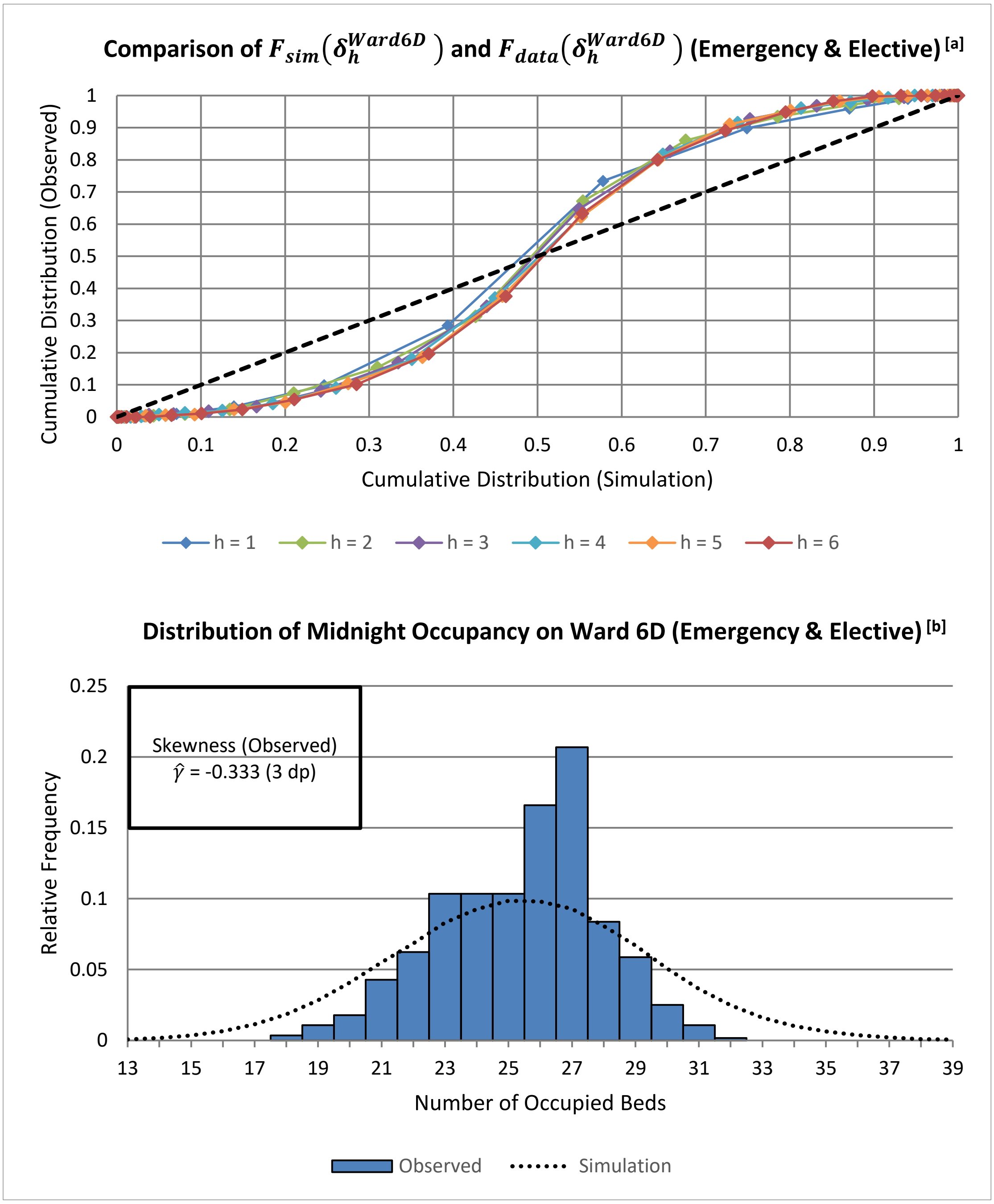


Figure A.2.6: Ward 4K

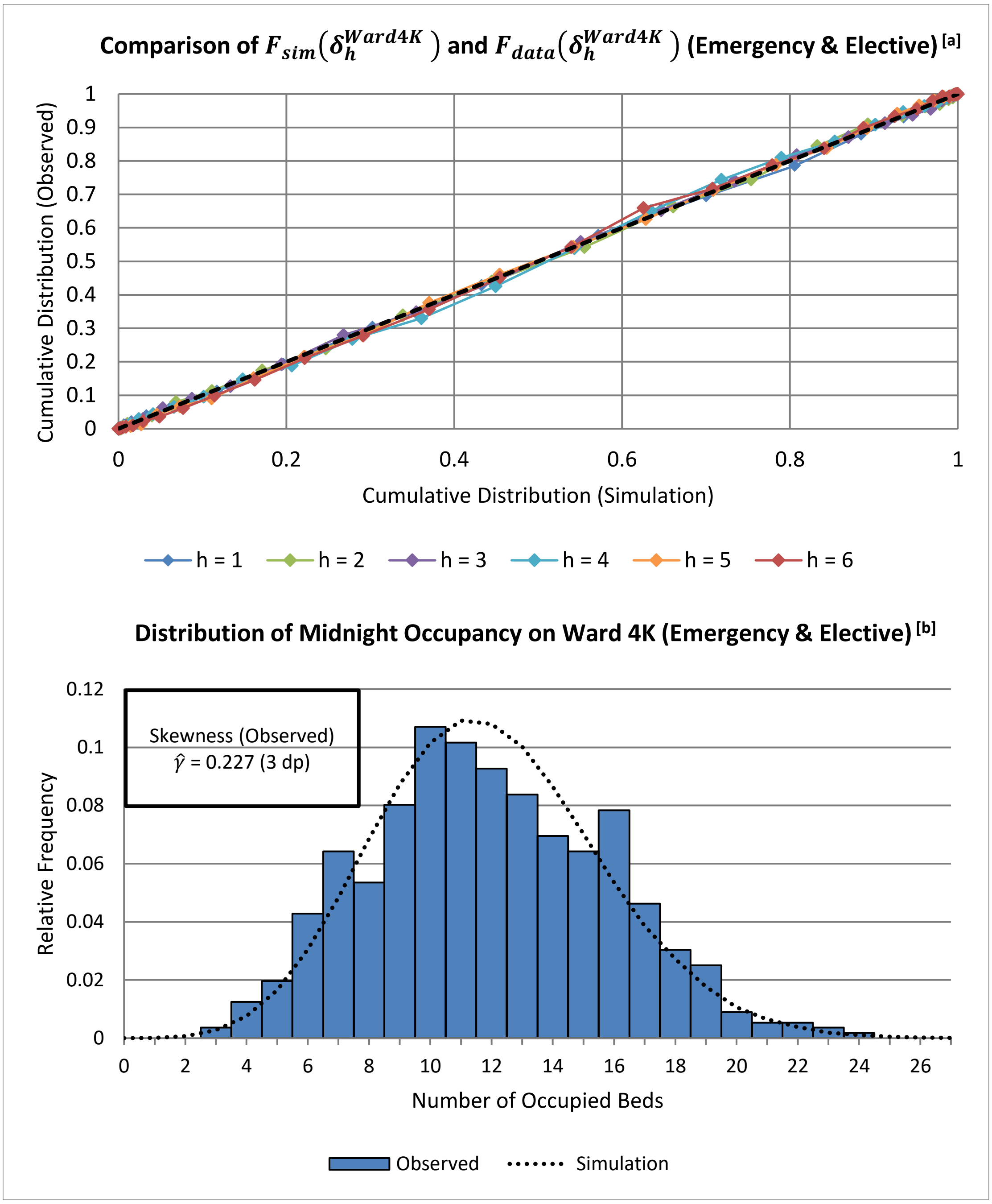


Figure A.2.7: Northside

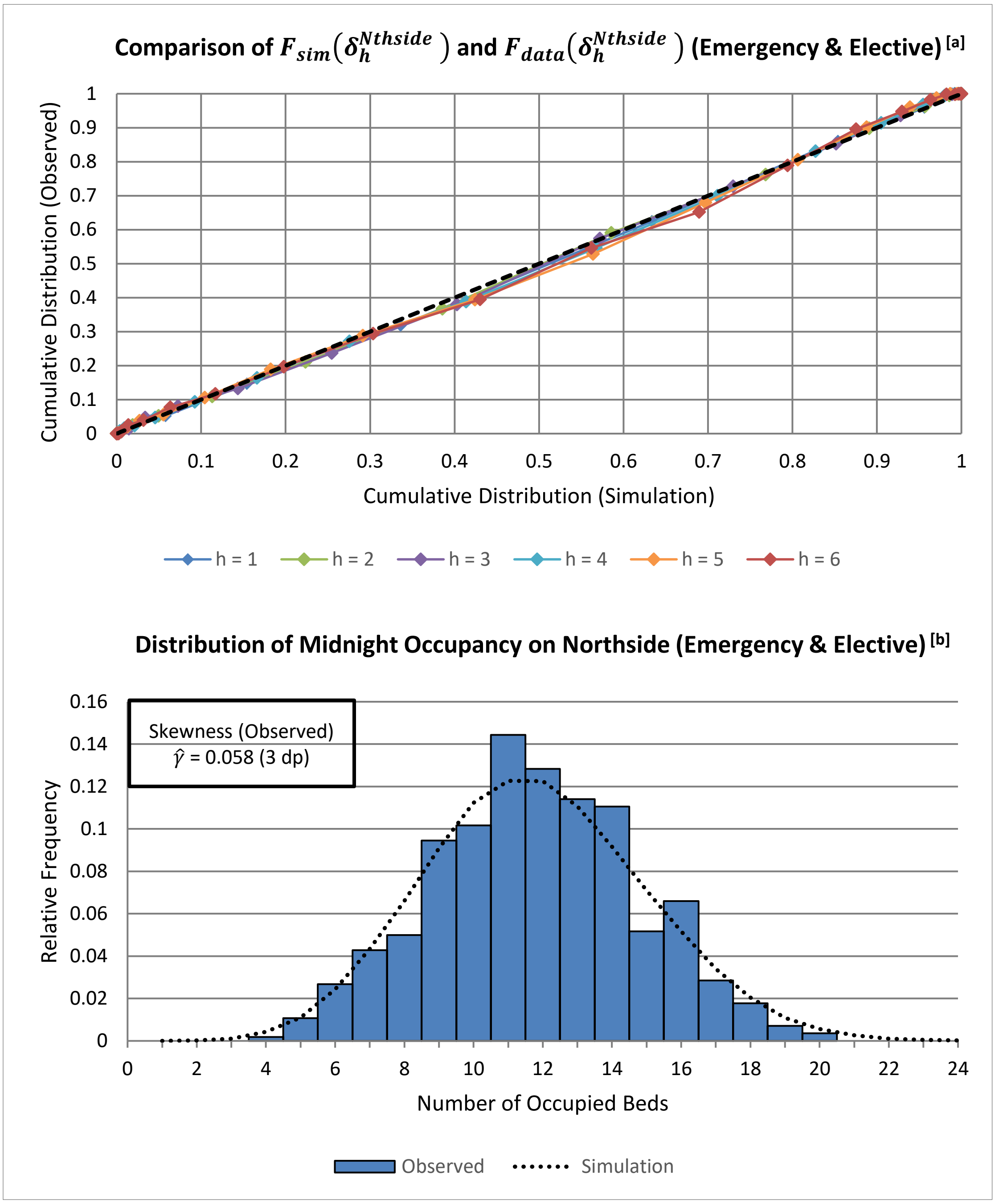
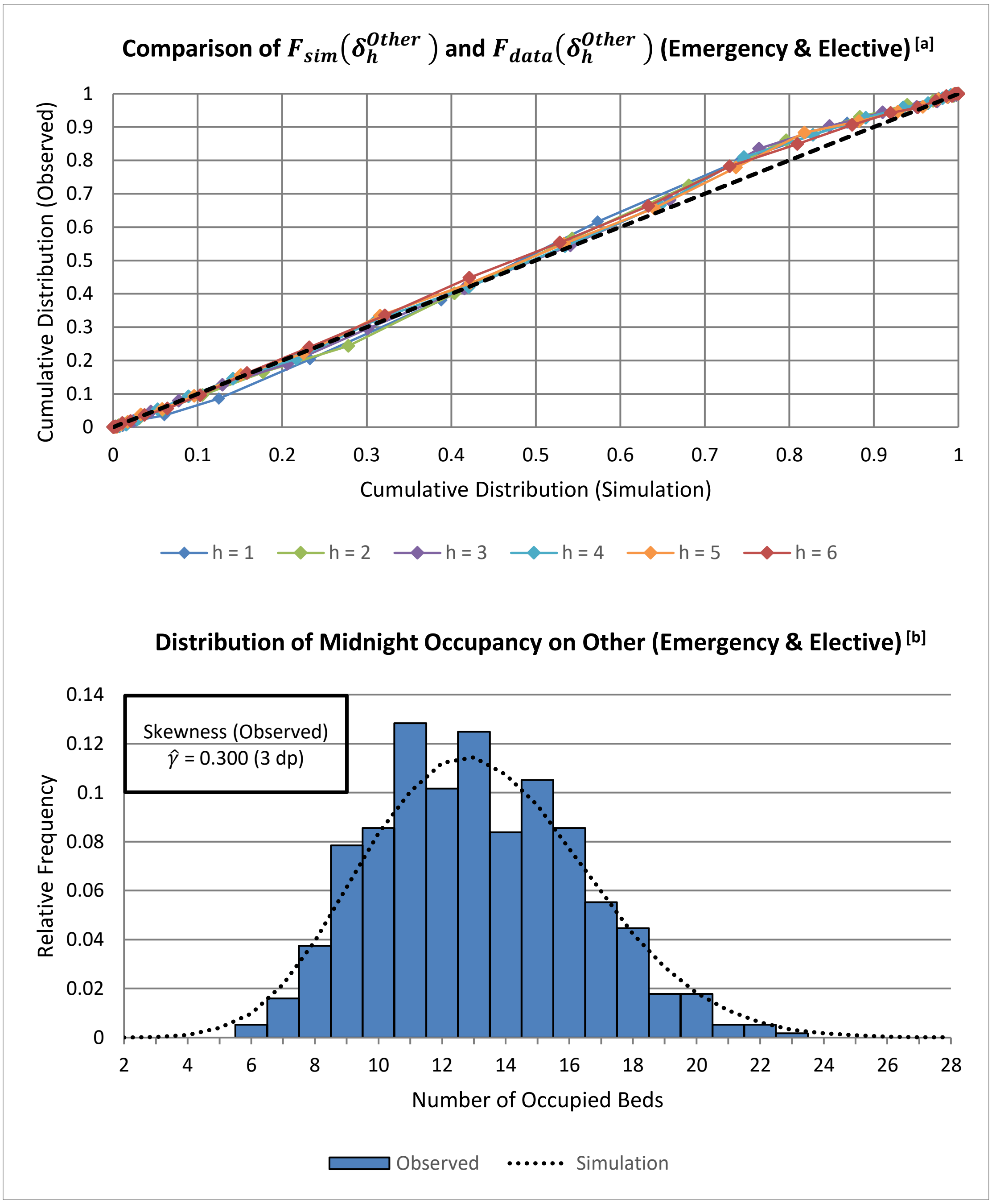


Figure A.2.8: Other



Appendix B
Appendix to Chapter 5

Figure B.1: Ward 4D

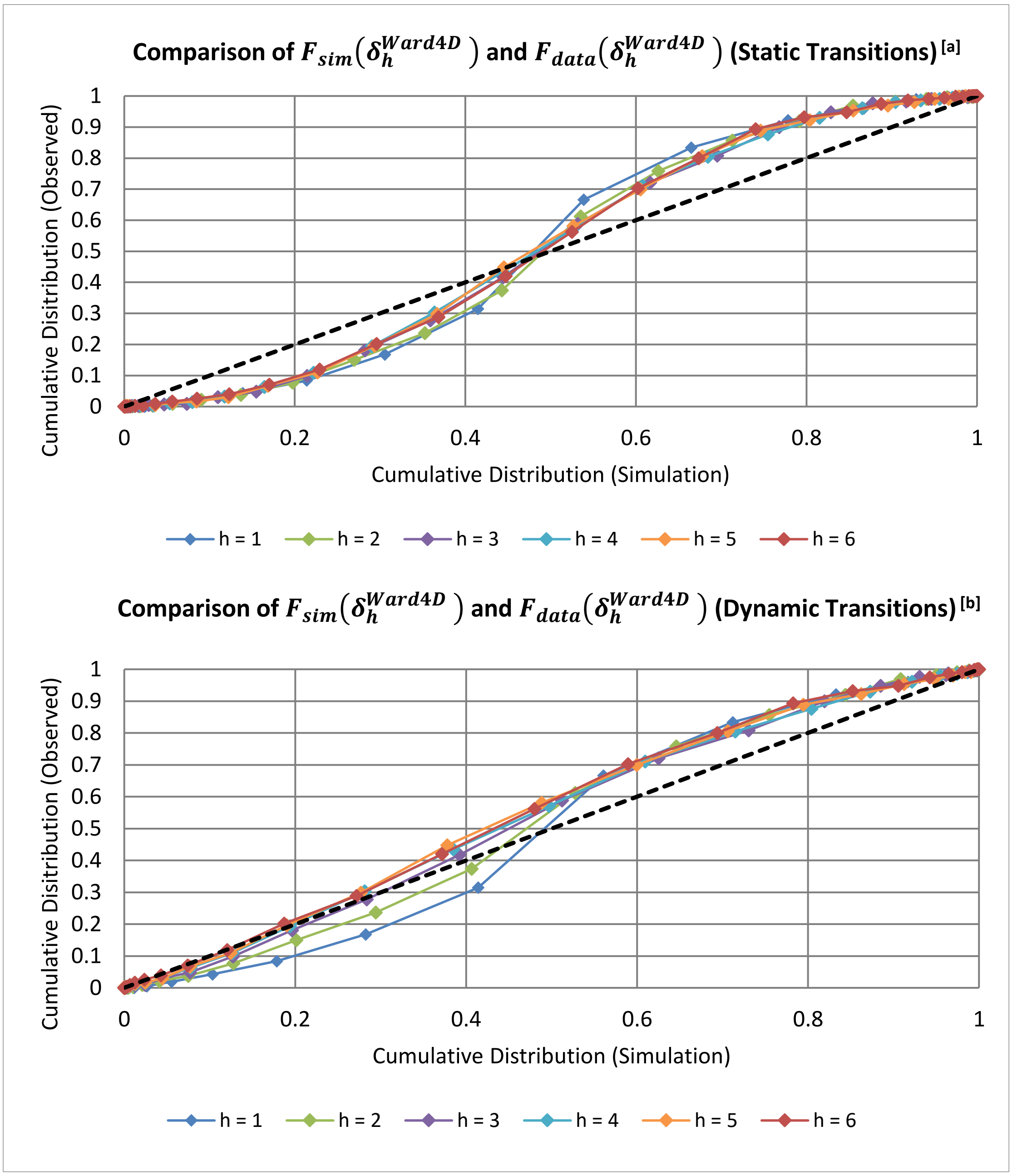


Figure B.2: Ward 4D

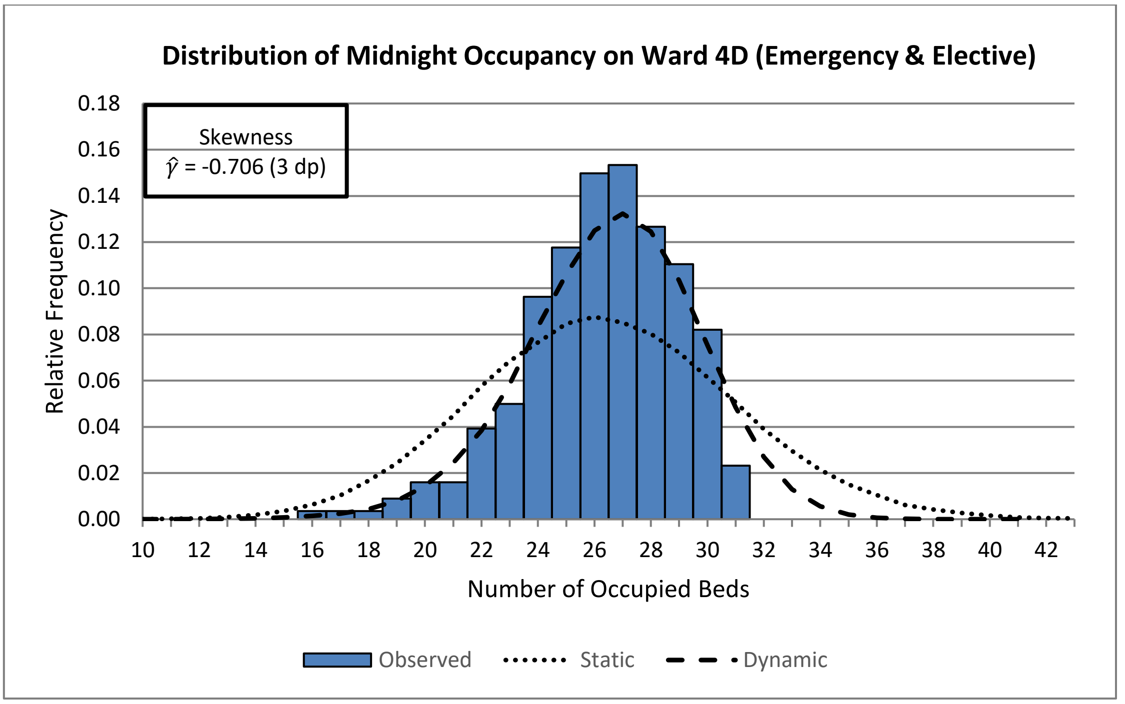


Figure B.3: Ward 5B

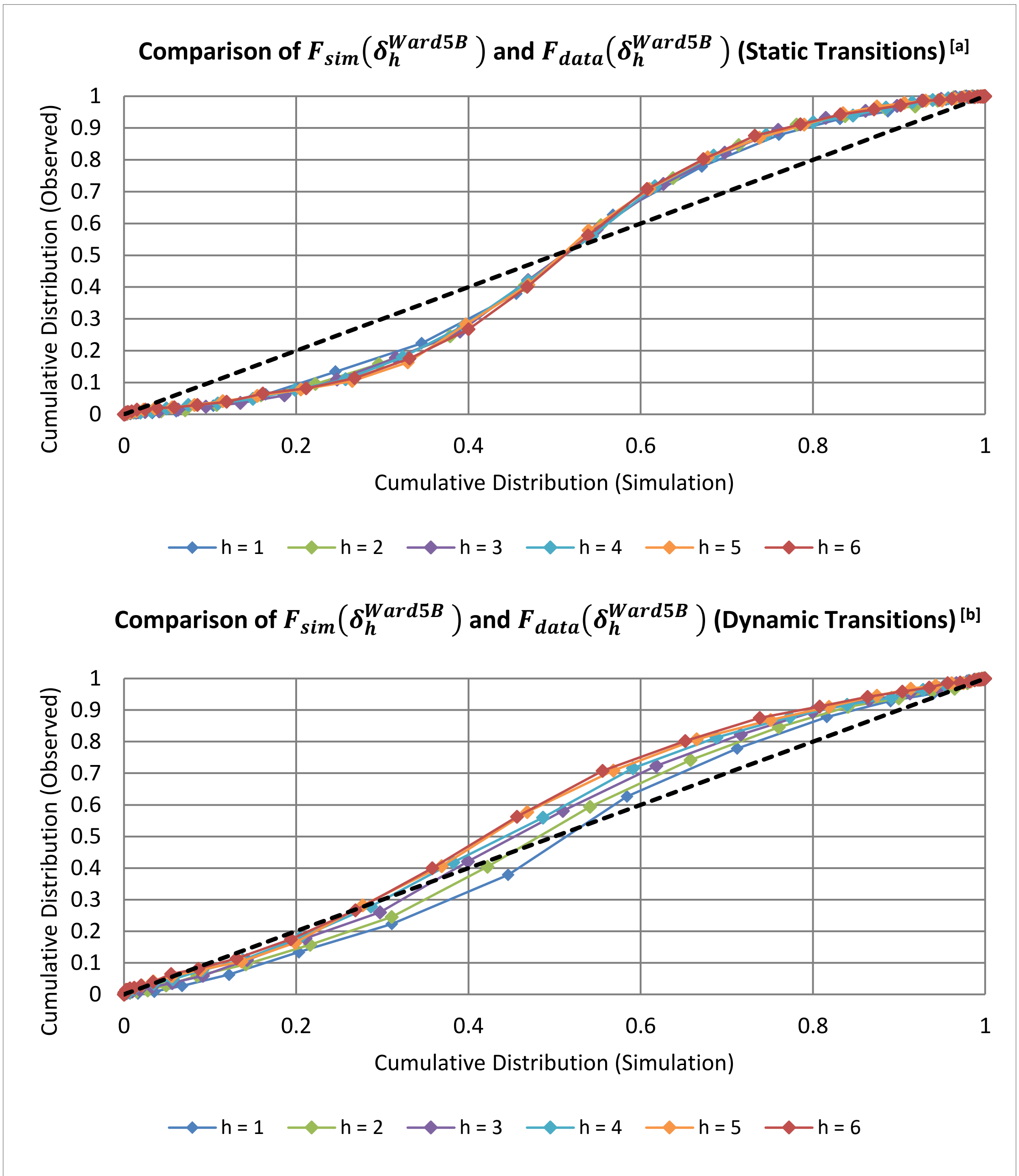


Figure B.4: Ward 5B

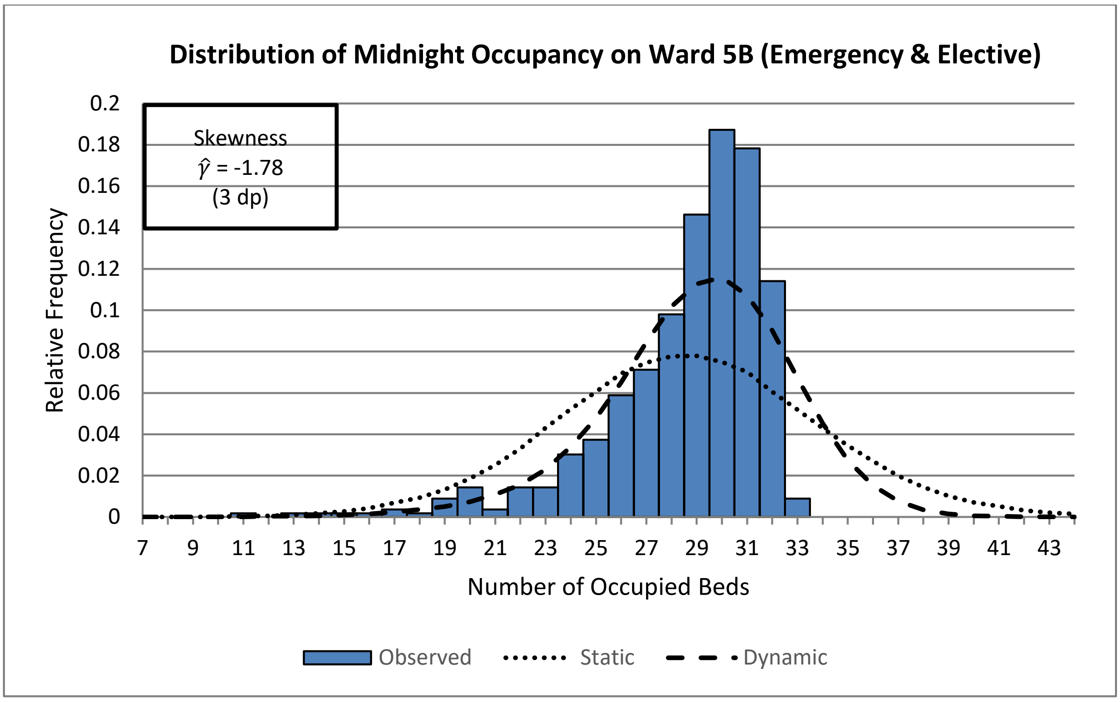
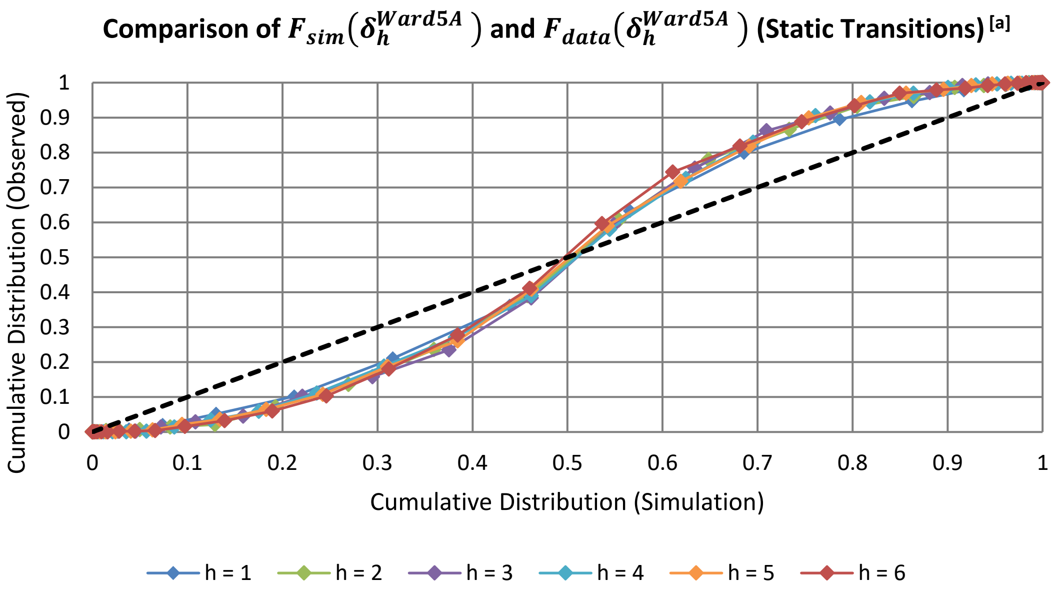


Figure B.5: Ward 5A



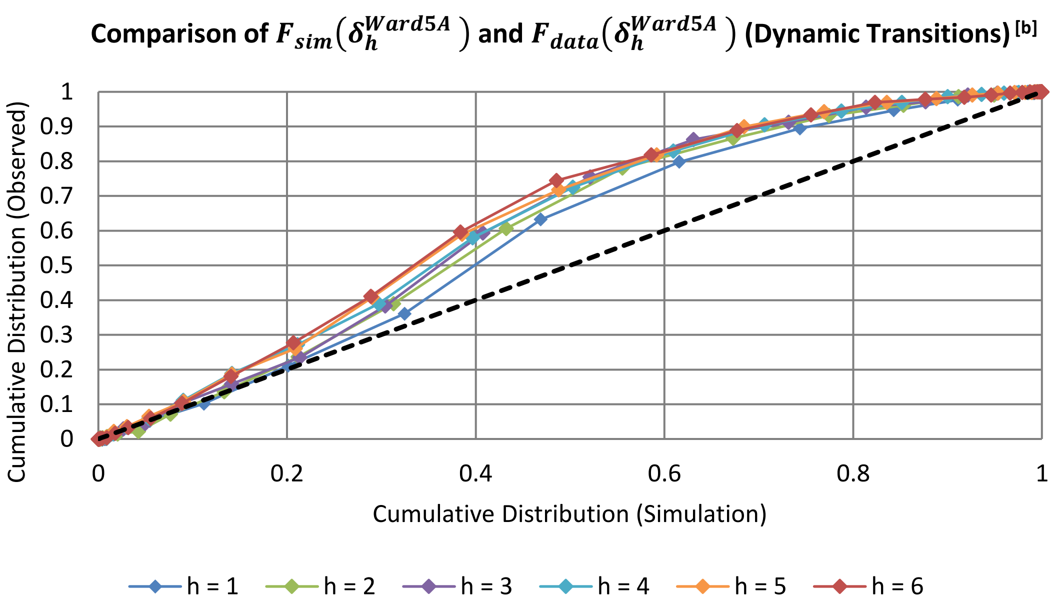


Figure B.6: Ward 5A

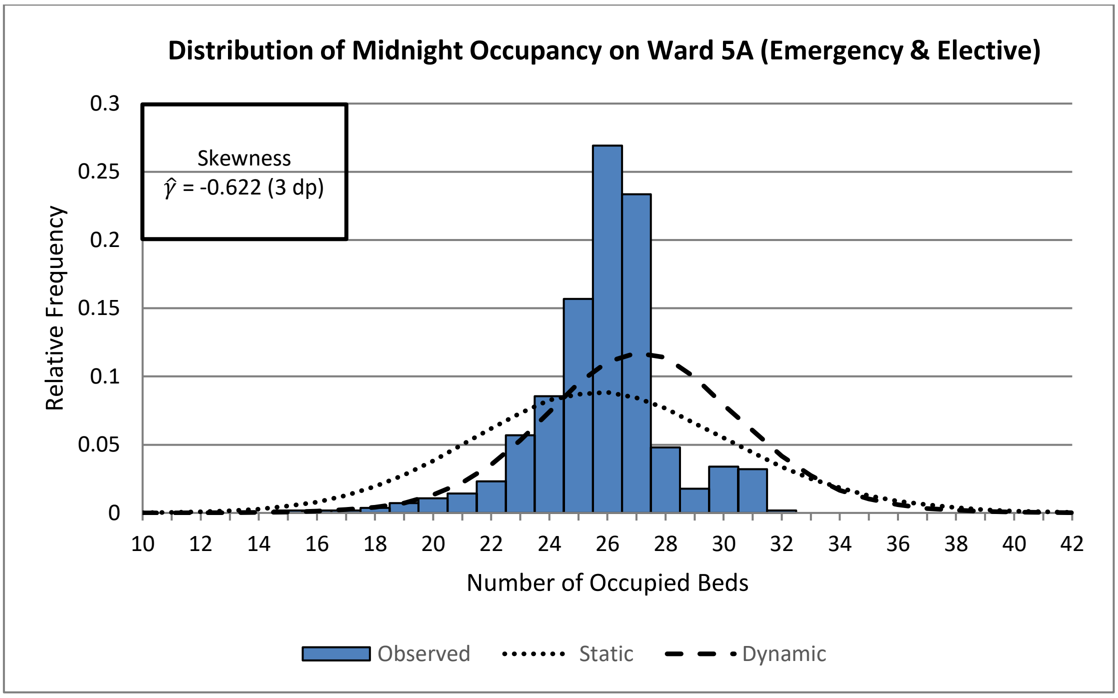
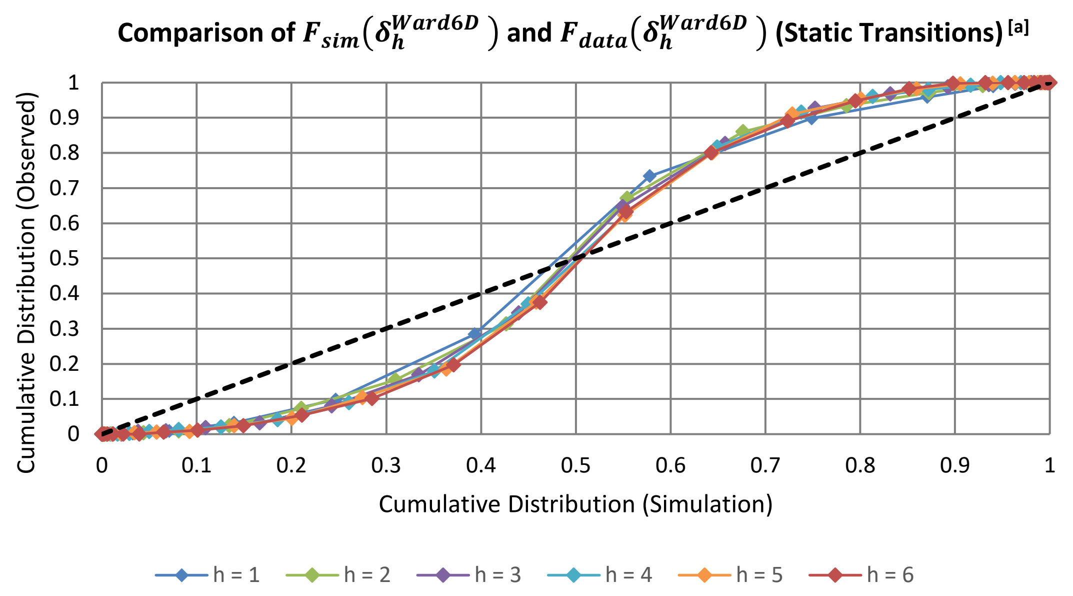


Figure B.7: Ward 6D



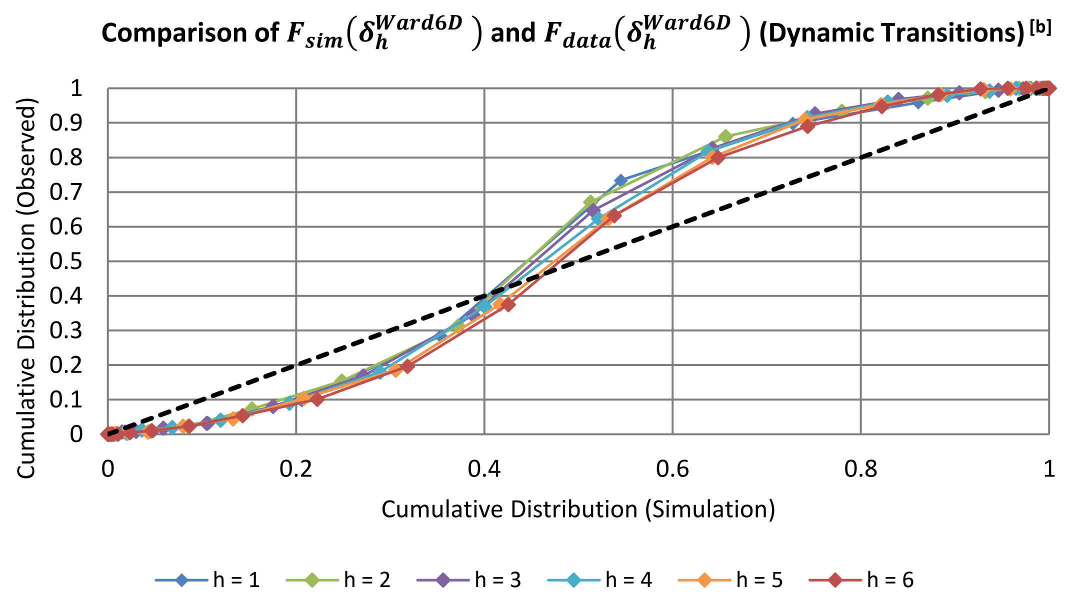


Figure B.8: Ward 6D

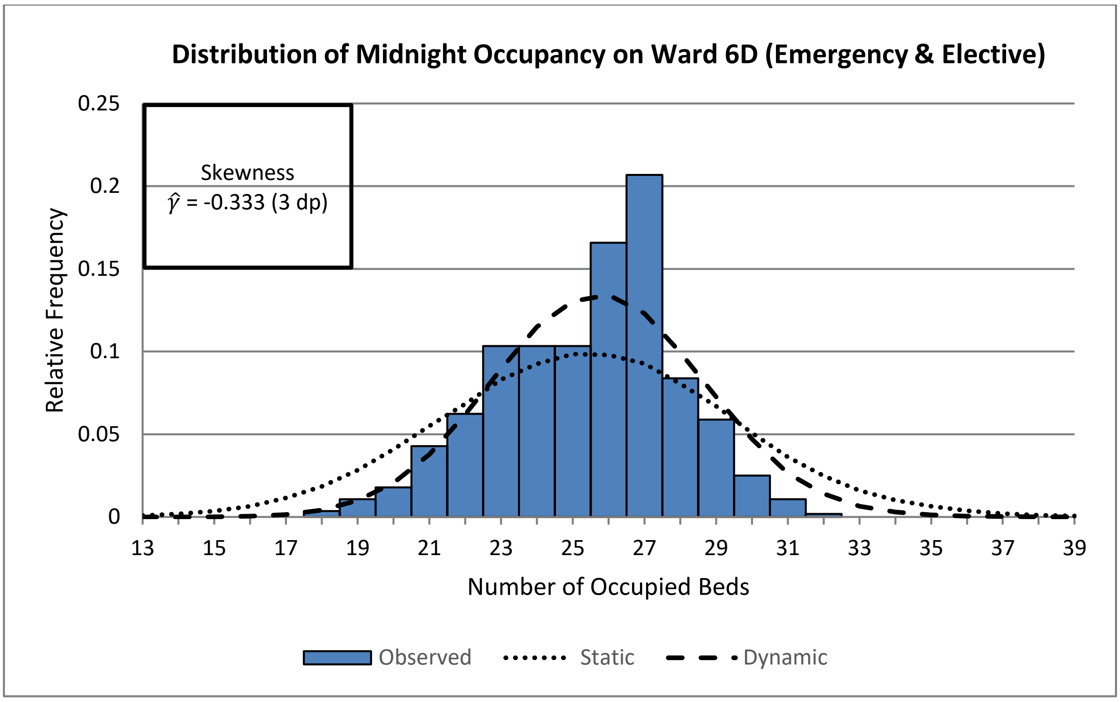


Figure B.9: Ward 4K

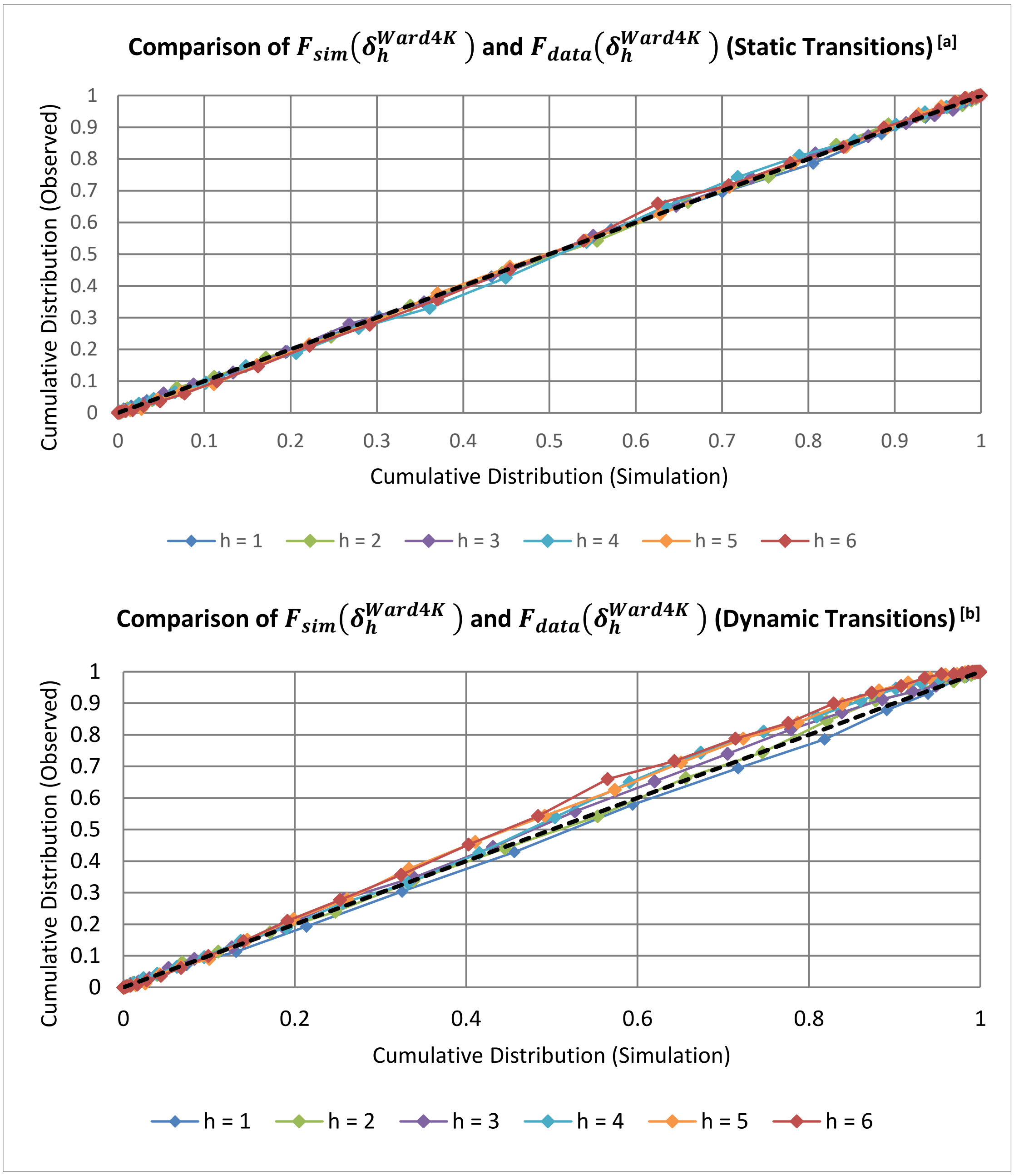


Figure B.10: Ward 4K

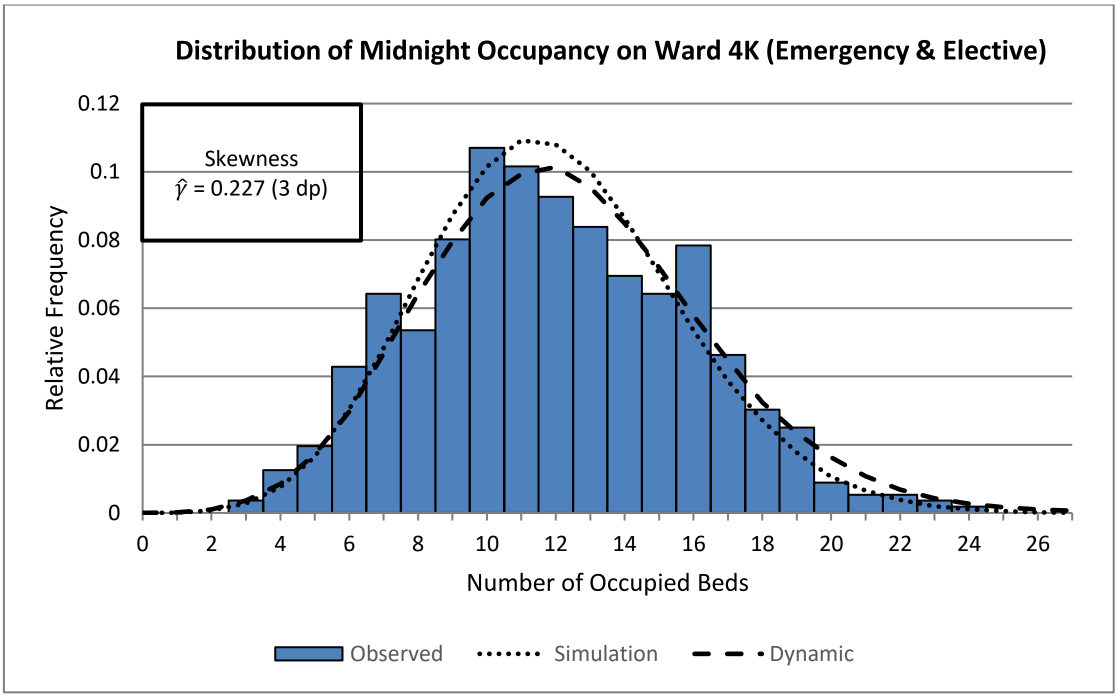


Figure B.11: Northside

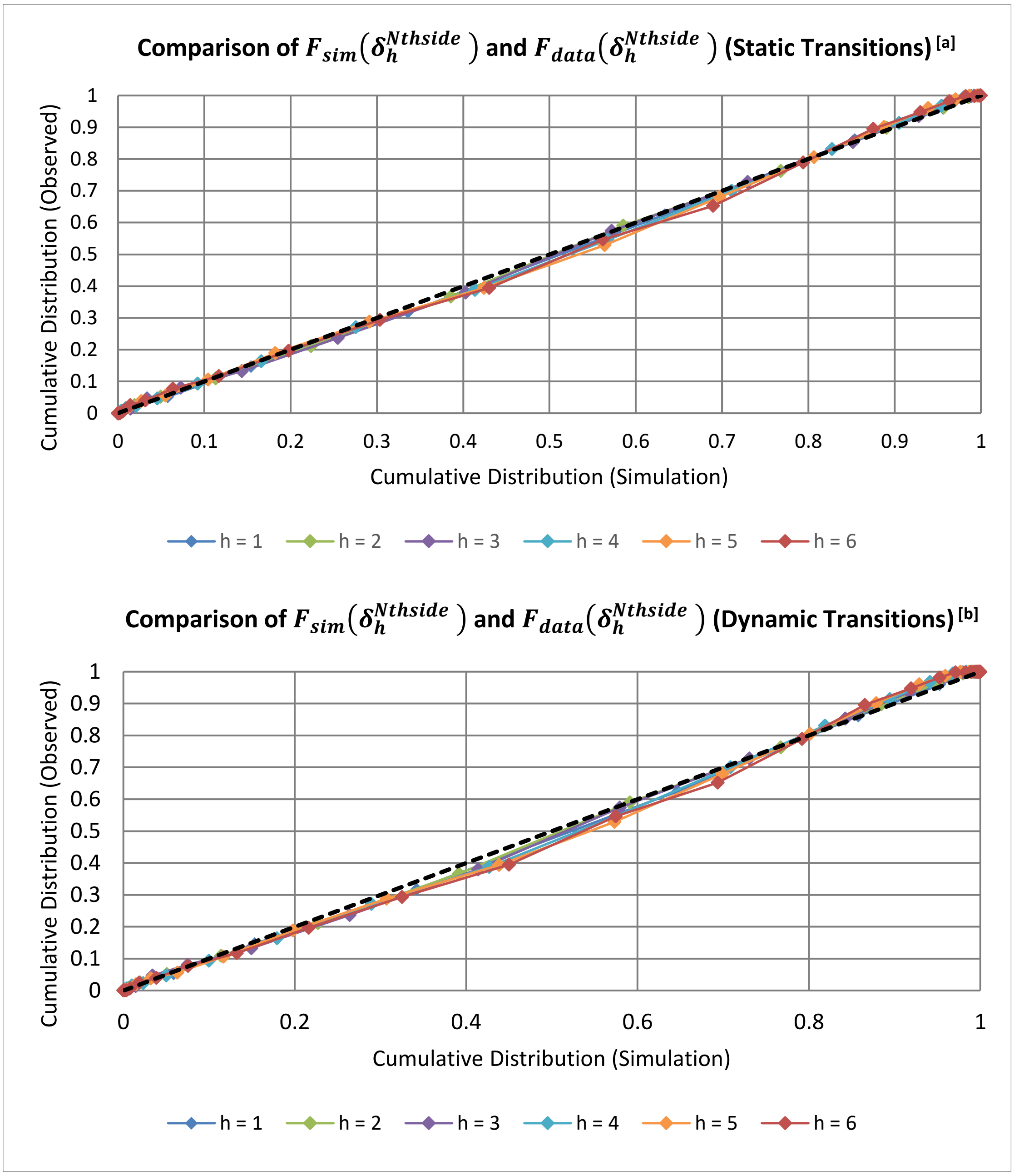


Figure B.12: Northside

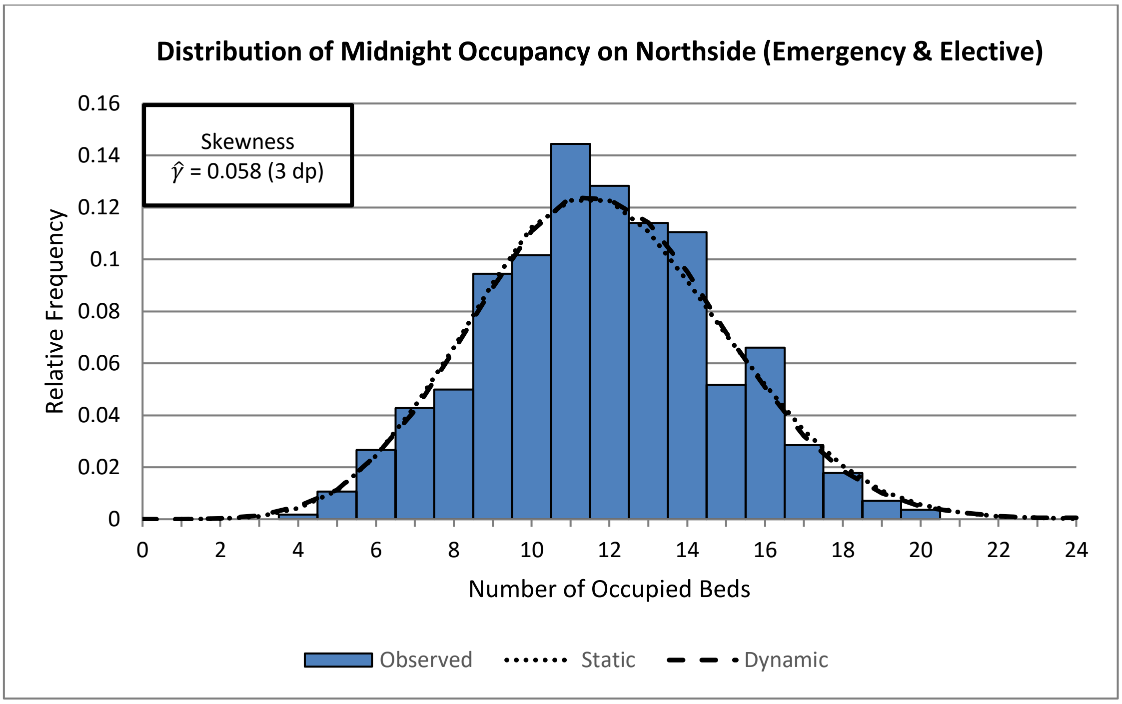


Figure B.13: Other

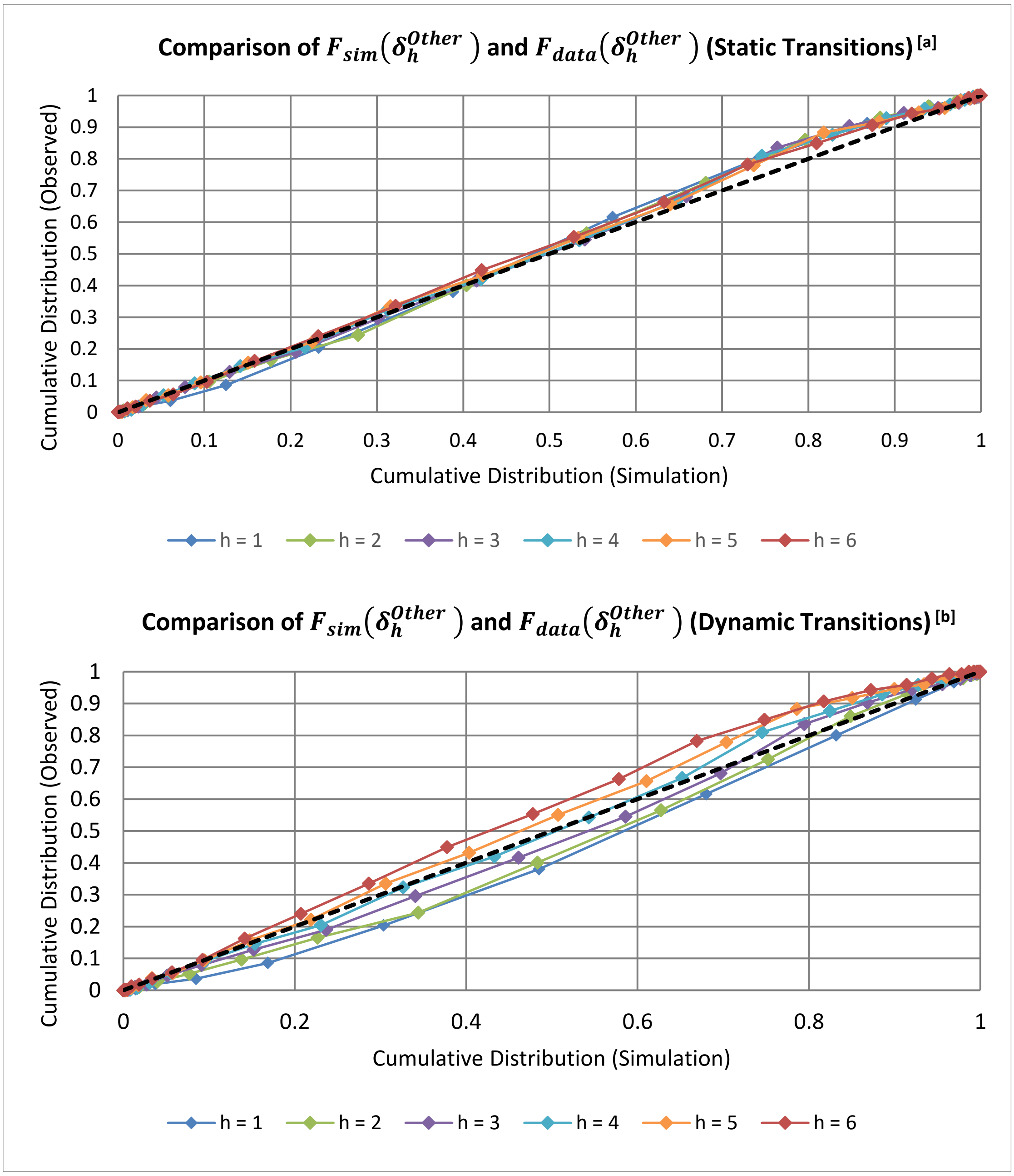
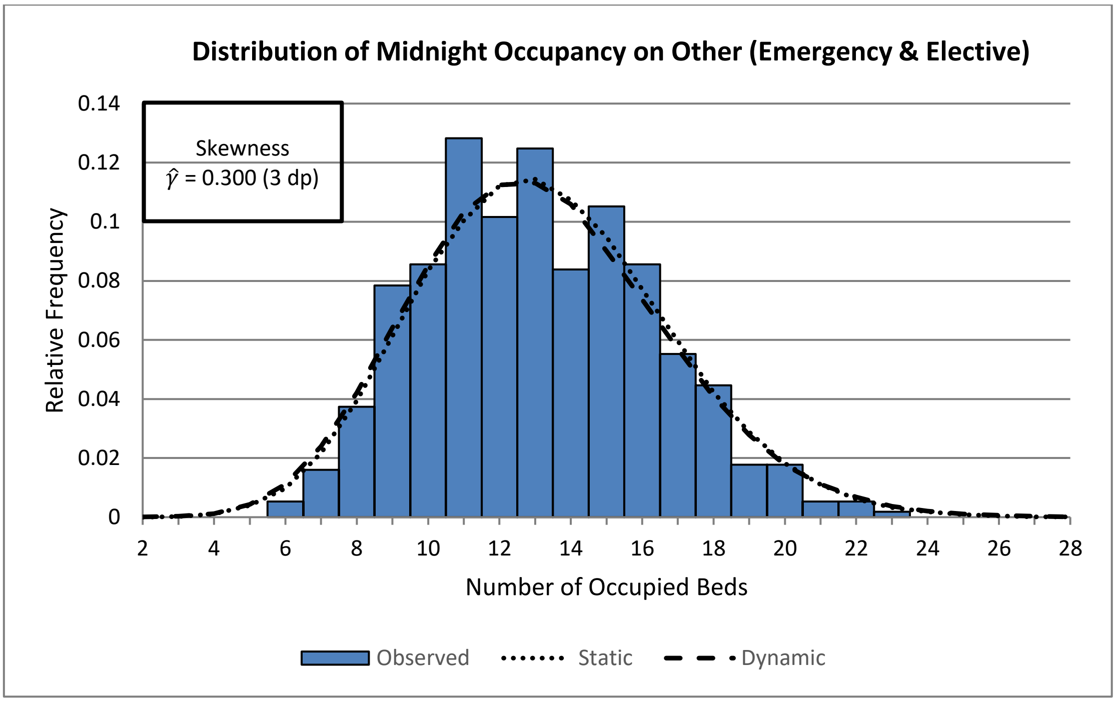


Figure B.14: Other



Appendix C
Reviewer's Guide (STRESS-DES)

The purpose of this appendix is to summarise the simulation study in a way which facilitates the review and reproduction of the model and experiments as necessary. In the interest of standardisation, the summary is framed using the STRESS-DES Guidelines recommended by Monks et al. (2018). In instances where the content of the thesis already satisfies the guidelines, links are provided to the relevant sections.

C.1 Objectives

C.1.1 Purpose of the model

The online discrete-event simulation (ODES) reported in this thesis is designed to investigate the impact of admissions scheduling decisions on short-term bed demand (via the midnight bed census), within a network of wards where bed resources may be pooled for emergency and elective care. For further details see Section 1.4 (Expected Contributions) and Section 2.5 (Research Questions).

C.1.2 Model Outputs

The primary output generated by the ODES are the realisations of the midnight bed census for each modelled ward and patient type (emergency/elective admission status). Key transformations of this data include:

1. The Δ-occupancy random variable for validating simulations whose outputs follow time-dependent distributions (see Section 4.7.4)
2. The estimated probability of bed demand exceeding available capacity (see Section 6.2.2)
3. Bed-Midnights Over Capacity (BMOC) which summarises (across all wards) the average number of midnights for which bed demand exceeds available capacity, multiplied by the extra beds which would be required to satisfy demand (see Section 6.2.2).

The computation of other statistics, such as prediction intervals which help to assess the validity of the offline model, are sufficiently described in the relevant sections of this thesis.

C.1.3 Experimentation Aims

Seven simulation experiments are conducted and reported in the body of this thesis:

- *Experiment 1 - Offline Model Validation:* This experiment assesses the validity of the offline model over longer simulation runs before the state-matching component of the online model is added. The Base Model (offline) with Static Transition Matrices (STMs) is used in this simulation experiment. For further details, see Section 4.5.
- *Experiment 2 - Online Model Validation:* This experiment assesses the validity of the online model over one-week intervals, initialised with each of the 560 hospital states observed in the patient administration database. This experiment also demonstrates the use of the Δ-

occupancy technique for validating models with time-dependent outputs. Performance is judged on how well the Δ-occupancy distributions fit the same distributions generated by the PA data. The Base Model (online) with STMs is used in this simulation experiment. For further details, see Section 4.7.

- *Experiment 3 - STMs vs DTMs:* In this simulation experiment, the effect of implementing Dynamic Transition Matrices (DTMs) which respond to bed occupancy levels, is investigated. The Base Model (online) with STMs is compared against the same model, instead using the DTM routing policy. Performance is judged on how well the Δ-occupancy distributions fit with the same distributions generated by the PA data. For both the STM and DTM runs, the structure of the experiment (initial conditions, run length and number of replications) is the same as Experiment 2. For further details, see Section 5.4.
- *Experiment 4 - High Risk Plans:* This scenario-based experiment demonstrates how the ODES could be used to reduce the risk of encountering capacity related issues over the course of a one-week planning horizon. The Base Scenario is the online model (with DTMs), using the elective admissions pattern which is observed in the PA data during Week 47. Two alternative scenarios are tested. The Postponement Schedule allows postponements to be made with respect to the observed schedule. The Postponement and Cancellation Schedule allows both postponements and cancellations to be made with respect to the observed schedule. Performance is judged on the

estimated probability of bed demand exceeding available capacity, and BMOC. For further details, see Section 6.2.

- *Experiment 5 - Additional Information and Observed Schedules:* This scenario-based experiment assesses how clinicians' estimates of length-of-stay could affect the estimated probability of bed demand exceeding capacity. Two Base Scenarios are considered using the online model with DTMs; the pattern of observed elective admissions for Week 47 and Week 53. For each Base Scenario, alternative scenarios are tested by increasing the accuracy of clinicians' estimates from d=0 to d=0.5, and again to d=0.75, to illustrate the impact of incorporating additional length-of-stay information from clinicians, on the probability of encountering capacity issues. Performance is judged on the estimated probability of bed demand exceeding available capacity, and BMOC. For further details, see Section 6.3.2.
- *Experiment 6 - Additional Information and Alternative Schedules:* This scenario-based experiment combines alternative scenarios from Experiment 4 and Experiment 5, to test the effect of using additional information (d=0.75) on the Postponement and Cancellation Schedule. The purpose is to assess whether further action could be taken to reduce the probability of encountering capacity issues, while also considering clinicians' estimates of length-of-stay. For further details, see Section 6.3.3.
- *Experiment 7 - Variance Reduction:* This scenario-based experiment assesses how the accuracy of clinicians' estimates of length-of-stay, translate to reductions in the variance of the midnight bed census. The

Base Scenario is the online model with DTMs and d=0, initialised every Monday, and run for the entire 560-day observation period. Alternative scenarios are generated by increasing d from 0 to 1 by 0.25 increments. Performance is judged on the variance of midnight occupancy, averaged over each day of the week, and the reduction in average variance, relative to the d=0 model. For further details, see Section 6.3.4.

C.2 Logic

C.2.1 Base model overview diagram

A diagram of the generic model structure can be found in Section 4.3.6 (Conceptual Model Diagram). A diagram of the wards which have been selected for individual modelling (from the PA data) can be found in Section 4.4.1 (Modelled Wards).

C.2.2 Base model logic

The base model logic is described in this thesis two parts. The first part consists of a description of the offline model, including arrival patterns (Section 4.4.2), length-of-stay (Section 4.4.3) and transitions between wards (Section 4.4.4). The second part consists of a description of the components required to bring the offline model online, including loading the real hospital state at initialisation (Section 4.6.1) and conditional length-of-stay distributions (Section 4.6.2).

C.2.3 Scenario logic

Two alternative scenarios are tested which change the base model's logic. The first alternative is the use of Dynamic Transition Matrices to govern patient transfers between wards (see Chapter 5). The second alternative evaluates the impact of using additional information to make better LOS predictions. This is applied to the patients occupying a bed at the time the ODES is initialised, and the elective patients which arrive during the planning period (see Chapter 6: Case Study 2). Other alternative scenarios are tested which include changes to the elective admissions schedules, however these do not change the logic of the base model.

C.2.4 Algorithms

Two important algorithms are used to draw realisations from empirical length of stay distributions. These algorithms are not included in Micro Saint Sharp by default.

1. *SampleECDF:* Uses the so-called Inversion Method (Devroye, 1986) to draw realisations from an empirical distribution function. In the C# code below, "ECDF" is a list containing values of the cumulative distribution. For example, the entry in the list for which ECDF{i}=0.5 is the median.

```
double U=Distributions.Rectangular(0.5,0);
int index=0;
double Prob=ECDF[index];
while(U>Prob){
	index++;
	Prob=ECDF[index];
}
return index;
```

2. *ScaleECDF:* Re-scales the ECDF to account for time already spent on the ward for the simulated patients loaded at initialisation. The algorithm requires an ECDF to be passed to it in the form of a list (as does SampleECDF), and an integer number of midnights already spent on the ward at the time of initialisation (ScaleTo). The algorithm returns a conditional distribution in the form of a new list which can be passed to SampleECDF for sampling remaining LOS. This is an implementation of Equation 4.3 (see Section 4.6.2).

```
//Handles the instance where no scaling is required.
if(ScaleTo==0){return ECDF;}

//Handles the instance where LOS_to_Date is already greater than the support of the LOS distribution.
//In this case patients' remaining LOS should be zero.
else if((int)ScaleTo>=ECDF.Count){
	var SubECDF = new List<double>();
	SubECDF.Add(1);
	return SubECDF;
}
//Handles the case where LOS_to_Date is non-zero, but within the range of the empirical distribution.
else{
	var SubECDF = new List<double>();
	SubECDF=ECDF.GetRange((int)ScaleTo,(int)(ECDF.Count-ScaleTo));
	for(int i=0; i<=SubECDF.Count-1; i++){
		SubECDF[i]=SubECDF[i]-ECDF[(int)ScaleTo-1];
	}
	for(int i=0; i<=SubECDF.Count-1; i++){
		SubECDF[i]=SubECDF[i]/(1-ECDF[(int)ScaleTo-1]);
	}
	return SubECDF;
}
```

C.2.5 Components

C.2.5.1 Entities

Patients are represented by simulation entities. All simulation entities in the model represent patients moving through the ward network. Every simulation entity is given the following attributes listed in Table C.1 upon creation.

Attribute	Description	Values
Entity.PatientType	Identifies whether an entity has emergency or elective admission status.	"Emer", "Elec".
Entity.ArrivalDay	Weekday of arrival to the current ward. This informs the distribution from which realisations of LOS are drawn.	"Monday", "Tuesday", "Wednesday", "Thursday", "Friday", "Saturday", "Sunday".
Entity.FirstWard	For elective patients, this is the ward identifier for the scheduled ward of arrival. The wards are identified using numbered strings. Defaults to blank for emergency patients.	"1" (ED), "2" (IC), "3" (4D), "4" (4K), "5" (5A), "6" (5B), "7" (5D), "8" (6D), "9" (Northside), "10" (Other).
Entity.PrevWard	Identifier for the previously visited ward. This is especially important for the patients loaded at initialisation, and the first ward-stay for the elective patients, since it determines whether additional LOS information will be used.	"Loaded","Scheduled", "1" (ED), "2" (IC), "3" (4D), "4" (4K), "5" (5A), "6" (5B), "7" (5D), "8" (6D), "9" (Northside), "10" (Other).
Entity.LOS_to_Date	Assigns the time already spent on the current ward for the entities loaded at initialisation, in whole midnights. Defaults to zero for patients which are created after initialisation.	Integer ≥ 0
Entity.LOS_Rem	Assigns the estimate of remaining LOS on the current ward (initialised	Integer ≥ 0

Attribute	Description	Values
	entities) or the estimate of LOS on the first ward (elective entities), in whole midnights. In practice, this would be loaded with clinicians' estimates. Presently, it is loaded with real midnights-remaining from the PA data.	

Table C.1: Attributes of each simulation entity or simulated patient.

C.2.5.2 Activities

Simulation entities (patients) engage in a series of ward stays between admission and discharge (Entry/Exit). The sequence of ward stays is governed by routing probabilities, except for the current/first ward for the loaded/elective entities. The details of the routing probabilities when Static Transition Matrices are used are detailed in Section 4.4.4. Dynamic Transition Matrices are also used to govern patient transitions in a way which responds to ward occupancy. An overview of the multinomial logistic regression models which compute the transition probabilities as a function of ward occupancy is provided in Section 5.3.2. Details of the equations, including estimated regression coefficients, are listed in Appendix D in Section D.1 for the emergency patients and Section D.2 for the elective patients.

C.2.5.3 Resources

The resources used by the simulated patients are the ten modelled wards detailed in Section 4.4.1.

C.2.5.4 Queues

There are no queues modelled in the simulation since all wards are treated as uncapacitated servers.

C.2.5.5 Entry/Exit Points

For the emergency patients, a dummy entry node exists (with a service time of zero) which routes patients to their first ward based on the transition probabilities estimated from the data (STMs and DTMs). For the elective patients, a dummy node also exists, although the routing rules are deterministic, based on the admission ward from the elective admissions in the PA data. Details of the arrival mechanisms are discussed in Section 4.4.2.

C.3 Data

C.3.1 Data Sources

The data which parameterises the ODES is an extract of the patient administration database from an Australian General Hospital, which is routinely collected. The raw extract contains information for all patients who occupy a bed between the 1st of January 2010, and the 30th of June 2012.

C.3.2 Pre-processing

The pre-processing steps which ready the database for use in parameterising the ODES model are discussed in Section 3.3 (Scope and Filtering) and Section 3.4 (Preliminary Analysis). After the data pre-processing is complete, the database which is used to parameterise the model covers the period from the 22nd of March 2010 to the 3rd of October 2011, and contains data relating to 16,276 distinct inpatient episodes.

C.3.3 Input parameters

The main input parameters for the ODES model are the arrival processes, length-of-stay distributions and ward transition probabilities. Since the details of the ward transition probabilities have already been discussed earlier within this appendix (Section C.2.5.2), this section focuses on the details of patient arrivals to simulated hospital, and the length-of-stay distributions used on each ward.

Emergency Admissions

As discussed in Section 4.3.5 (Uncontrolled Variables) and Section 4.4.2 (Modelling Arrivals), the number of emergency admissions per day is determined by empirical distributions generated from the PA data. Within-day admission patterns are not considered because bed occupancy is only captured at once per day, at midnight. Figure C.1 charts the empirical CDFs which govern the number of emergency arrivals for each day of the week. Table C.2 lists the values of the ECDFs explicitly.

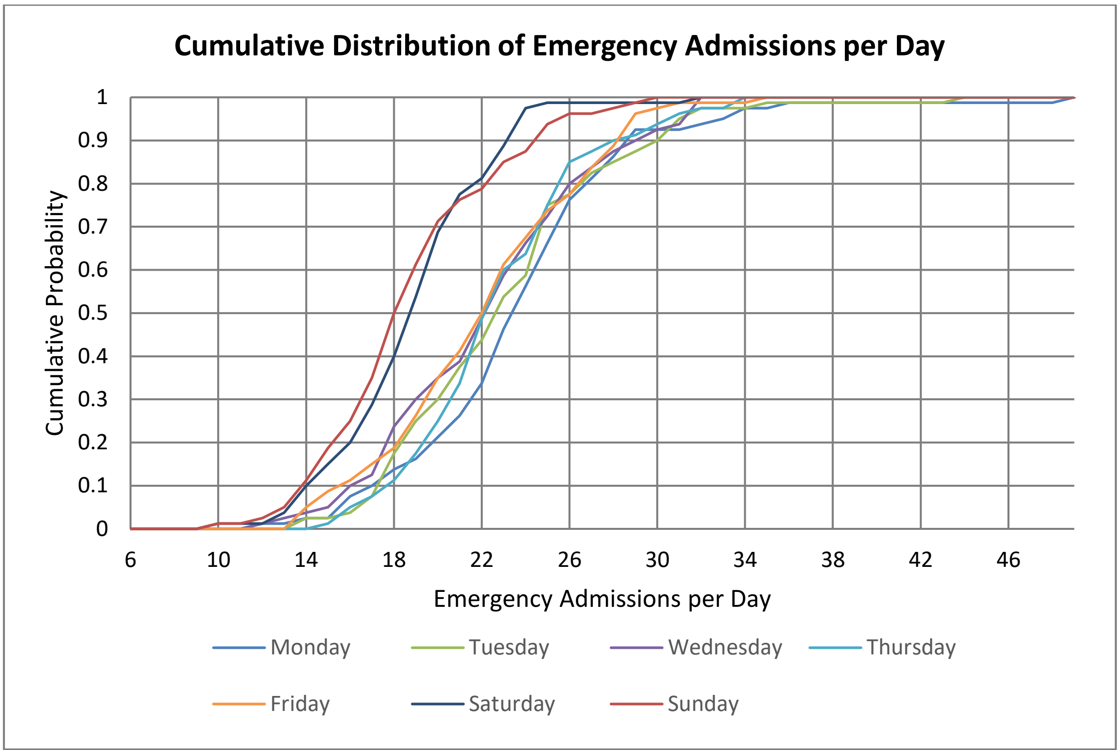


Figure C.1: Empirical CDFs of the number of emergency admissions per day, by day of the week. The ODES draws from one of these CDFs each day, depending on the week day being simulated.

	Cumulative Distribution of the Number of Emergency Admissions						
Admissions	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
10	0.000	0.000	0.000	0.000	0.000	0.013	0.013
11	0.000	0.000	0.000	0.000	0.000	0.013	0.013
12	0.013	0.000	0.013	0.000	0.000	0.013	0.025
13	0.013	0.000	0.025	0.000	0.000	0.038	0.050
14	0.025	0.025	0.038	0.000	0.050	0.100	0.113
15	0.025	0.025	0.050	0.013	0.088	0.150	0.188
16	0.075	0.038	0.100	0.050	0.113	0.200	0.250
17	0.100	0.075	0.125	0.075	0.150	0.288	0.350
18	0.138	0.175	0.238	0.113	0.188	0.400	0.500
19	0.163	0.250	0.300	0.175	0.263	0.538	0.613
20	0.213	0.300	0.350	0.250	0.350	0.688	0.713
21	0.263	0.375	0.388	0.338	0.413	0.775	0.763
22	0.338	0.438	0.488	0.488	0.500	0.813	0.788
23	0.463	0.538	0.588	0.600	0.613	0.888	0.850
24	0.563	0.588	0.663	0.638	0.675	0.975	0.875
25	0.663	0.750	0.725	0.750	0.738	0.988	0.938
26	0.763	0.775	0.800	0.850	0.775	0.988	0.963
27	0.813	0.825	0.838	0.875	0.838	0.988	0.963
28	0.863	0.850	0.875	0.900	0.888	0.988	0.975
29	0.925	0.875	0.900	0.913	0.963	0.988	0.988
30	0.925	0.900	0.925	0.938	0.975	0.988	1.000
31	0.925	0.950	0.938	0.963	0.988	0.988	
32	0.938	0.975	1.000	0.975	0.988	1.000	
33	0.950	0.975		0.975	0.988		

	Cumulative Distribution of the Number of Emergency Admissions						
Admissions	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
34	0.975	0.975		1.000	0.988		
35	0.975	0.988			1.000		
36	0.988	0.988					
37	0.988	0.988					
38	0.988	0.988					
39	0.988	0.988					
40	0.988	0.988					
41	0.988	0.988					
42	0.988	0.988					
43	0.988	0.988					
44	0.988	1.000					
45	0.988						
46	0.988						
47	0.988						
48	0.988						
49	1.000						

Table C.2: Listing of the values of the ECDFs from which the number of emergency admissions each day is drawn.

Elective Admissions

The elective admissions occur deterministically from schedule stored as an array within the ODES model. The "base case" data is the admissions pattern extracted from the PA data. Alternatives can be tested by entering user-defined schedules, as is the case for Experiment 4 and Experiment 6 (see Section C.1.3 of this appendix). Any admissions schedule consists of one row for each admitted patient and requires the admission date as an integer and the admission ward using its numeric identifier. Additionally, the schedule can contain an estimate of the LOS (in midnights) for the admitting ward, as is the case for Experiments 5 - 7.

Storing the admission date as an integer is a common way of handling date calculations in data software such as Microsoft Excel or SAS. In Excel, dates are stored as the number of days since January 1st, 1900 which is also how dates are stored in the elective admissions schedule. Table C.3 shows an example schedule in which two patients are admitted on two different days.

Admission Day	Ward of First Admission (Entity.FirstWard)	Estimated LOS on First Ward (Entity.LOS_Rem)
40588 (14/02/2011)	"4" (Ward 4K)	3
40589 (15/02/2011)	"6" (Ward 5B)	1

Table C.3: Example elective admissions schedule in which two patients are admitted over two days. The column headings indicate the entity attributes set by the schedule.

Length of Stay

The length of stay distributions for each modelled ward are graphed and listed in Appendix E; in Section E.1 for the emergency patients, and Section E.2 for the elective patients.

Accuracy of Length of Stay (d)

In Experiments 5 – 7 (Section 6.3.2 – Section 6.3.4) additional LOS information is added to the simulation via the elective admissions schedule and the state loaded at initialisation. Since this information is taken from the observations of LOS in the PA data, the parameter d is used to model the uncertainty associated with clinicians' LOS estimates. This parameter could also be used in an operational setting, if calibrated by real estimates of clinicians' LOS prediction accuracy. In the ODES model, d is a single floating-point number, set globally by the user for the whole hospital. For further details, refer to the simulation experiments reported in Chapter 6 (Case Study 2) which make use of this input parameter.

C.3.4 Assumptions

Few assumptions are made in the development on the ODES, since the level of detail of the conceptual model is well supported by the PA data. However, one abstraction is the use of "uncapacitated" or "infinite server" nodes in the simulation to represent wards in the real hospital. Modelling wards in this way allows for the straightforward calculation of the probability of bed demand exceeding available capacity, and other capacity-related metrics such as the average bed-midnights over capacity (BMOC). See Section 4.3.3 (Uncapacitated Wards) for a discussion of the use of uncapacitated wards, and the case studies reported in Chapter 6 for examples of how they can be used to inform operational decision making.

C.4 Experimentation

C.4.1 Initialisation

The ability to load the current system state at initialisation is one of the core properties of an online simulation. In this application, the initial conditions are loaded from a patient-level array which contains one record (or row) for each patient occupying a bed at run-time. Within each row, the columns set the attributes of each initialised patient; including initialisation time, the ward identifier for where the patient is staying, the day of the week when the ward-stay began, and the time already spent on the ward. Separate arrays are created for the emergency and elective patients; therefore, the patient type attribute is set based on which array is queried at run-time.

In its current form, the ODES contains sets of initial conditions (deterministic), taken from the PA data, which can be queried at run-time for the purpose of validation and experimentation. However, in practice, a connection with the real patient database should be established to initialise the model, including queries to transform the raw data as necessary. Table C.4 provides an example of an array which would load three patients at initialisation on the 23rd of March 2010.

Initialisation Day	Current Ward (Entity.FirstWard)	Midnights spent on Current Ward (Entity.LOS_to_Date)	Weekday of Arrival on Current Ward (Entity.ArrivalDay)	Estimated remaining LOS on Current Ward (Entity.LOS_Rem)
40259 (22/03/2010)	"2" (ICU)	1	"Sunday"	0
40259 (22/03/2010)	"5" (Ward 5A)	37	"Saturday"	7
40259 (22/03/2010)	"6" (Ward 5B)	16	"Saturday"	18

Table C.4: Example array which would load three patients to three distinct wards at initialisation. The column headings indicate the entity attributes set by the array of initial conditions.

For the experiments using the ODES, the simulation terminates after seven days. However, this run length is not determined by any physical property of the system being modelled (such as closing time, or scheduled jobs completed) but rather the period for which the initial conditions continue to influence the simulation outputs. In Section 4.7.2, it is shown that the variance of the midnight occupancy predictions is similar to that of an equivalent offline model after running for more than seven days. Therefore, the value of using this ODES model lies within one-week (or shorter) planning horizons.

Although the ODES is used for most of the simulation experiments reported in this thesis, Experiment 1 is concerned with validating the *offline* model, and therefore warm-up period data is generated during this experiment. The length of the warm-up period is determined visually, by charting the midnight census time-series. See Section 4.5.2 for the details of the warm-up period associated with the offline model.

C.4.2 Run length

Run length and number of replications are noted within the run configuration section which precedes each simulation experiment. Experiments either simulate the full 560-day observation period or selected one-week planning horizons. See Section C.1.3 of this appendix for references to each of the experimentation sections, which state the run length and number of replications.

C.4.3 Estimation Approach

All simulation experiments use multiple replications to account for stochastic variation in the output data (midnight occupancy). When the full 560-day observation period is simulated (offline and online models), 100 replications are executed. When one-week planning horizons are simulated to investigate the impact of alternative admissions schedules, or the effect of additional patient information, 400 replications are executed. The shorter run length means that computational effort can be allocated to increasing the number of replications, resulting in smoother approximations of the distributions of midnight occupancy.

C.5 Implementation

C.5.1 Software or programming language

The simulation models reported in this thesis are developed and run in Micro Saint Sharp version 3.6 (build 3.6.4.528). Micro Saint Sharp is a commercial DES software package and is not open source.

C.5.2 Random sampling

To generate pseudo-random numbers Micro Saint Sharp uses the System.Random call within the Microsoft .NET framework. The algorithm implemented within this class is based on the subtractive random number generator reported by Knuth (1998).

C.5.3 Model execution

The parallel execution facility within Micro Saint Sharp has been used for all simulation experiments to decrease real experiment run time.

C.5.4 System specification

All simulation experiments in were conducted on a Windows 7 PC using an Intel Core i3 CPU (2.4 GHz, 2 Cores) with 4 GB of installed physical memory. Approximate run-times for each simulation experiment are listed in Table C.5.

	Run Length	Number of Replications	Run-time
Experiment 1	560 simulated days	100	~12 minutes
Experiment 2	560 simulated days	100	~12.5 minutes
Experiment 3	560 simulated days	100	~15 minutes
Experiment 4	7 simulated days	400	~40 seconds
Experiment 5	7 simulated days	400	~40 seconds
Experiment 6	7 simulated days	400	~40 seconds
Experiment 7	560 simulated days	100	~15 minutes

Appendix D
Fitted MLR Models

In this appendix, the systems of equations which define the Dynamic Transition Matrices are reported in detail, including estimates of the regression coefficients, and an example of the SAS code used to fit them. In Sections D.2 and D.3, the left-hand side of each equation is defined in terms of the numeric identifier for the ward. The concordance between numeric identifier and ward name is as follows:

1. Emergency Department
2. Intensive Care Unit
3. Ward 4D
4. Ward 4K
5. Ward 5A
6. Ward 5B
7. Ward 5D
8. Ward 6D
9. Northside ward
10. Other ward
11. Exit/Discharge

In instances where an equation is set identically equal to zero, this is to either prevent reflexive transfers, or because no transfer between the wards was observed in the PA data.

D.1 MLR Fitting using AIC

The following SAS code fits the MLR which governs transitions away from the Emergency Department, for the emergency patients, which can be thought of as a single row of probabilities in the DTM. The same procedure is used for all other wards.

```
*1. Read in data of the form shown in Figure 5.1, keeping records for
the emergency patients who transition from the Emergency Department;

PROC LOGISTIC DATA = TransfersOccupancy

             (WHERE = (LocationID="ED" and Admission_Type='Emer'));


*2. Specify Next_LocationID as a categorical variable, with "Exit" as
the reference outcome;

CLASS Next_LocationID (REF = "Exit") / PARAM = REF;


*3. Set the dependent (LHS) and explanatory (RHS) variables and the
link function between them (generalised logit). All wards are to be
considered in the sequential search, including their two-factor
interaction terms (using the "@ 2" syntax);

MODEL Next_LocationID = ED | IC | WARD4D | WARD4K | WARD5A | WARD5B |

                        WARD5D | WARD6D| NORTHSIDE | OTHER @ 2 /

                        LINK = GLOGIT;


*4. Set the variable selection method and the p-values for entering
and remaining in the model. As mentioned previously, p-values are not
the primary selection criteria, however setting relatively high p-
values eliminates the least significant variables from the search and
speeds up the procedure.

SELECTION = Stepwise SLENTRY=0.5 SLSTAY=0.5;


*5. Create two datasets which contain the results of the procedure.
SUM reports the variables which enter or are removed during the stepwise
search. FIT contains the AIC which is evaluated each time a variable
enters or is removed.

ODS OUTPUT ModelBuildingSummary=SUM; ODS OUTPUT FitStatistics=FIT;

RUN;
```

```
*6. Select the model summary which minimises the AIC, by combining the
SUM and FIT datasets.

PROC SQL;

      CREATE TABLE OPT_AIC1 AS SELECT

            A.CRITERION,

            A.INTERCEPTONLY,

            A.INTERCEPTANDCOVARIATES,

            A.STEP AS MODEL_NUM,

            B.STEP AS VAR_NUM,

            B.EFFECTENTERED,

            B.EFFECTREMOVED,

            B.NUMBERINMODEL,

            MIN(A.INTERCEPTANDCOVARIATES) AS OPT_AIC

      FROM FIT AS A LEFT JOIN SUM AS B

      ON A.STEP>=B.STEP

      WHERE A.CRITERION='AIC' AND A.INTERCEPTANDCOVARIATES LT
      A.INTERCEPTONLY

      ORDER BY A.STEP,B.STEP;


      CREATE TABLE OPT_AIC2 AS SELECT * FROM OPT_AIC1

      WHERE INTERCEPTANDCOVARIATES=OPT_AIC;

QUIT;
```

D.2 MLR models for the Emergency Patients

	Fitted Multinomial Logistic Regression models by ward of departure (Emergency Patients)
Entry	Num1=Math.Exp(39.196+0.0731*ED+-0.7421*IC+0.1368*Ward4D+0.0598*Ward4K+-0.0418*Ward5A+0.00465*Ward4K*Ward5A+-0.082*Ward5B+-0.00431*ED*Ward5B+0.02*IC*Ward5B+-0.00671*Ward4K*Ward5B+-1.1998*Ward5D+0.00101*Ward5B*Ward5D+-1.4196*Ward6D+0.049*Ward5D*Ward6D+0.5473*Northside+-0.0111*Ward4D*Northside+-0.00734*Ward5D*Northside+-0.1143*Other+0.0127*IC*Other); Num2=Math.Exp(5.9939+-0.3115*ED+-0.1582*IC+0.0882*Ward4D+0.6134*Ward4K+0.2665*Ward5A+-0.0241*Ward4K*Ward5A+0.1213*Ward5B+0.0074*ED*Ward5B+-0.00175*IC*Ward5B+0.00335*Ward4K*Ward5B+-0.4628*Ward5D+-0.0096*Ward5B*Ward5D+-1.522*Ward6D+0.0538*Ward5D*Ward6D+1.9607*Northside+0.000207*Ward4D*Northside+-0.064*Ward5D*Northside+-0.0159*Other+0.00237*IC*Other); Num3=Math.Exp(8.1549+-0.096*ED+0.368*IC+0.3982*Ward4D+0.5289*Ward4K+-0.0372*Ward5A+-0.00353*Ward4K*Ward5A+-0.0151*Ward5B+0.00341*ED*Ward5B+-0.00037*IC*Ward5B+-0.017*Ward4K*Ward5B+-0.8133*Ward5D+0.00581*Ward5B*Ward5D+-0.8593*Ward6D+0.0263*Ward5D*Ward6D+1.6016*Northside+-0.0466*Ward4D*Northside+-0.0135*Ward5D*Northside+0.2842*Other+-0.03*IC*Other); Num4=Math.Exp(-4.0615+-0.3726*ED+1.1627*IC+0.0778*Ward4D+0.4858*Ward4K+-0.0469*Ward5A+0.00481*Ward4K*Ward5A+0.6352*Ward5B+0.0127*ED*Ward5B+-

	Fitted Multinomial Logistic Regression models by ward of departure (Emergency Patients)
	0.0363*IC*Ward5B+-0.0236*Ward4K*Ward5B+-0.4283*Ward5D+-0.00648*Ward5B*Ward5D+-0.5797*Ward6D+0.0205*Ward5D*Ward6D+0.2423*Northside+-0.00728*Ward4D*Northside+0.000435*Ward5D*Northside+0.0878*Other+-0.00134*IC*Other); Num5=Math.Exp(7.2811 +0.0772*ED +0.114*IC +0.1268*Ward4D +1.0425*Ward4K +0.0482*Ward5A +-0.0256*Ward4K*Ward5A +0.0815*Ward5B +-0.00452*ED*Ward5B +0.0184*IC*Ward5B +-0.0166*Ward4K*Ward5B +-0.6802*Ward5D +-0.00078*Ward5B*Ward5D +-0.8563*Ward6D +0.0308*Ward5D*Ward6D +0.6534*Northside +-0.0115*Ward4D*Northside +-0.0111*Ward5D*Northside +0.3302*Other +-0.0415*IC*Other); Num6=Math.Exp(17.8118 +0.0705*ED +0.3421*IC +0.0738*Ward4D +0.4051*Ward4K +-0.2515*Ward5A +0.0177*Ward4K*Ward5A +0.8477*Ward5B +-0.00483*ED*Ward5B +-0.0162*IC*Ward5B +-0.0309*Ward4K*Ward5B +-0.8286*Ward5D +-0.0157*Ward5B*Ward5D +-1.5996*Ward6D +0.0552*Ward5D*Ward6D +0.9986*Northside +-0.0151*Ward4D*Northside +-0.0218*Ward5D*Northside +0.0221*Other +0.000692*IC*Other); Num7=Math.Exp(8.409 +-0.1707*ED +0.115*IC +0.3511*Ward4D +0.8396*Ward4K +-0.0257*Ward5A +-0.00036*Ward4K*Ward5A +0.3826*Ward5B +0.00687*ED*Ward5B +-0.00393*IC*Ward5B +-0.0312*Ward4K*Ward5B +-0.805*Ward5D +-0.00471*Ward5B*Ward5D +-1.317*Ward6D +0.0435*Ward5D*Ward6D +1.9163*Northside +-0.035*Ward4D*Northside +-0.0348*Ward5D*Northside +0.1488*Other +-0.00906*IC*Other);

	Fitted Multinomial Logistic Regression models by ward of departure (Emergency Patients)

```
Num8=Math.Exp(-27.2634 +0.0893*ED +1.1856*IC +0.3954*Ward4D +0.6745*Ward4K +-
0.00679*Ward5A +-0.0122*Ward4K*Ward5A +0.3578*Ward5B +-0.00241*ED*Ward5B +-
0.0304*IC*Ward5B +-0.0189*Ward4K*Ward5B +0.1691*Ward5D +0.0025*Ward5B*Ward5D
+0.5028*Ward6D +-0.023*Ward5D*Ward6D +0.541*Northside +-0.0362*Ward4D*Northside
+0.0179*Ward5D*Northside +0.1594*Other +-0.0177*IC*Other);

Num9=Math.Exp(19.9881 +0.0456*ED +-0.662*IC +-0.0322*Ward4D +0.3854*Ward4K
+0.0885*Ward5A +-0.00516*Ward4K*Ward5A +0.8242*Ward5B +-0.00256*ED*Ward5B
+0.0181*IC*Ward5B +-0.00991*Ward4K*Ward5B +-0.6182*Ward5D +-0.0303*Ward5B*Ward5D +-
1.793*Ward6D +0.0623*Ward5D*Ward6D +0.3051*Northside +0.00232*Ward4D*Northside +-
0.0113*Ward5D*Northside +-0.0775*Other +0.0092*IC*Other);

Num10=1/(1 +Num1 +Num2 +Num3 +Num4 +Num5 +Num6 +Num7 +Num8 +Num9);

if(Location==1){return Num1*Num10;}
else if(Location==2){return Num2*Num10;}
else if(Location==3){return Num3*Num10;}
else if(Location==4){return Num4*Num10;}
else if(Location==5){return Num5*Num10;}
else if(Location==6){return Num6*Num10;}
else if(Location==7){return Num7*Num10;}
else if(Location==8){return Num8*Num10;}
else if(Location==9){return Num9*Num10;}
```

	Fitted Multinomial Logistic Regression models by ward of departure (Emergency Patients)
	else if(Location==10){return Num10;} else{return 0;}
Emergency Department (ED)	Num2=Math.Exp(9.3342 +-0.2542*Ward5D +-0.0705*Other +-0.2915*Ward4D +0.6459*Ward5B +- 0.2878*Ward5A +0.0122*ED +0.3688*Ward6D +-1.8121*IC +0.0607*Ward4D*IC +-0.0666*Ward4K +-0.0111*Ward4D*Ward5B +0.0011*Ward5D*Ward4D +-0.00105*Ward5B*Ward5A +0.00834*Ward5D*Ward5A +-0.00101*Ward5D*Other +0.00252*Ward5D*Ward4K +- 0.00343*Other*ED +-0.0128*Ward5B*Ward6D); Num3=Math.Exp(20.9576 +-0.379*Ward5D +-0.2206*Other +-0.2097*Ward4D +0.2967*Ward5B +- 0.6219*Ward5A +-0.0346*ED +0.1574*Ward6D +-0.9896*IC +0.0324*Ward4D*IC +0.3454*Ward4K +-0.0099*Ward4D*Ward5B +0.00164*Ward5D*Ward4D +0.00509*Ward5B*Ward5A +0.0137*Ward5D*Ward5A +0.00591*Ward5D*Other +-0.0132*Ward5D*Ward4K +-0.00209*Other*ED +-0.00589*Ward5B*Ward6D); Num4=Math.Exp(23.5242 +-0.2879*Ward5D +0.0702*Other +-0.5232*Ward4D +0.1753*Ward5B +- 0.5369*Ward5A +-0.0319*ED +0.2024*Ward6D +-1.3928*IC +0.049*Ward4D*IC +0.0584*Ward4K +0.00549*Ward4D*Ward5B +-0.00342*Ward5D*Ward4D +-0.00373*Ward5B*Ward5A +0.0182*Ward5D*Ward5A +-0.00652*Ward5D*Other +-0.00249*Ward5D*Ward4K +- 0.00211*Other*ED +-0.00728*Ward5B*Ward6D); Num5=Math.Exp(36.5854 +-0.6878*Ward5D +0.0975*Other +-0.7564*Ward4D +0.1786*Ward5B +- 0.9159*Ward5A +-0.077*ED +0.1982*Ward6D +-1.4821*IC +0.0528*Ward4D*IC +0.1057*Ward4K +0.00608*Ward4D*Ward5B +0.00554*Ward5D*Ward4D +-0.0051*Ward5B*Ward5A

	Fitted Multinomial Logistic Regression models by ward of departure (Emergency Patients)
	+0.0287*Ward5D*Ward5A +-0.00747*Ward5D*Other +-0.00383*Ward5D*Ward4K +-0.001*Other*ED +-0.00683*Ward5B*Ward6D); Num6=Math.Exp(15.6085 +-0.3905*Ward5D +-0.1693*Other +-0.6054*Ward4D +0.4031*Ward5B +-0.353*Ward5A +-0.0779*ED +0.561*Ward6D +-1.0994*IC +0.0384*Ward4D*IC +0.2871*Ward4K +0.00575*Ward4D*Ward5B +0.00396*Ward5D*Ward4D +-0.00675*Ward5B*Ward5A +0.0149*Ward5D*Ward5A +0.00196*Ward5D*Other +-0.00975*Ward5D*Ward4K +-0.00019*Other*ED +-0.0192*Ward5B*Ward6D); Num7=Math.Exp(21.4872 +-0.2098*Ward5D +-0.4093*Other +-0.00375*Ward4D +0.0471*Ward5B +-0.7331*Ward5A +-0.1485*ED +0.3275*Ward6D +-0.9867*IC +0.0344*Ward4D*IC +0.3966*Ward4K +0.0059*Ward4D*Ward5B +-0.0182*Ward5D*Ward4D +0.00471*Ward5B*Ward5A +0.0177*Ward5D*Ward5A +0.00885*Ward5D*Other +-0.0141*Ward5D*Ward4K +0.00426*Other*ED +-0.0117*Ward5B*Ward6D); Num8=Math.Exp(13.6554 +0.0778*Ward5D +-0.3028*Other +-0.4661*Ward4D +0.2107*Ward5B +-0.1989*Ward5A +-0.129*ED +0.289*Ward6D +-1.1415*IC +0.0368*Ward4D*IC +0.532*Ward4K +0.00113*Ward4D*Ward5B +0.000421*Ward5D*Ward4D +0.00403*Ward5B*Ward5A +-0.0018*Ward5D*Ward5A +0.00609*Ward5D*Other +-0.0186*Ward5D*Ward4K +0.00471*Other*ED +-0.0139*Ward5B*Ward6D); Num9=Math.Exp(-112.1 +2.0455*Ward5D +0.6539*Other +4.5601*Ward4D +1.9038*Ward5B +-0.5062*Ward5A +-0.2237*ED +0.4332*Ward6D +0.771*IC +-0.0226*Ward4D*IC +-0.0567*Ward4K

	Fitted Multinomial Logistic Regression models by ward of departure (Emergency Patients)

```
+-0.0827*Ward4D*Ward5B +-0.0691*Ward5D*Ward4D +0.0305*Ward5B*Ward5A +-
0.0233*Ward5D*Ward5A +-0.0303*Ward5D*Other +0.0108*Ward5D*Ward4K +0.00765*Other*ED +-
0.0206*Ward5B*Ward6D);

Num10=Math.Exp(16.534 +-0.2193*Ward5D +-0.00022*Other +-0.412*Ward4D +0.1271*Ward5B
+-0.4875*Ward5A +-0.0757*ED +0.2532*Ward6D +-0.8757*IC +0.0294*Ward4D*IC
+0.2889*Ward4K +-0.00091*Ward4D*Ward5B +0.00619*Ward5D*Ward4D +0.00538*Ward5B*Ward5A
+0.00821*Ward5D*Ward5A +-0.00317*Ward5D*Other +-0.0102*Ward5D*Ward4K
+0.00284*Other*ED +-0.0099*Ward5B*Ward6D);

Num11=1/(1 +Num2 +Num3 +Num4 +Num5 +Num6 +Num7 +Num8 +Num9 +Num10);

if(Location==1){return 0;}
else if(Location==2){return Num2*Num11;}
else if(Location==3){return Num3*Num11;}
else if(Location==4){return Num4*Num11;}
else if(Location==5){return Num5*Num11;}
else if(Location==6){return Num6*Num11;}
else if(Location==7){return Num7*Num11;}
else if(Location==8){return Num8*Num11;}
else if(Location==9){return Num9*Num11;}
else if(Location==10){return Num10*Num11;}
else{return Num11;}
```

	Fitted Multinomial Logistic Regression models by ward of departure (Emergency Patients)
Intensive Care (IC)	Num1=Math.Exp(-18.8178 +0.2415*Ward5D +0.1667*Ward4D +0.0682*Ward5A +-0.0987*Ward5B +0.5496*IC); Num3=Math.Exp(-5.9834 +0.0749*Ward5D +-0.129*Ward4D +0.0417*Ward5A +0.0919*Ward5B +0.2822*IC); Num4=Math.Exp(-6.721 +0.1773*Ward5D +0.000626*Ward4D +0.1056*Ward5A +-0.1154*Ward5B +-0.0132*IC); Num5=Math.Exp(-6.5708 +0.1952*Ward5D +0.0894*Ward4D +-0.1718*Ward5A +0.05*Ward5B +0.1459*IC); Num6=Math.Exp(-7.7565 +0.0795*Ward5D +0.1714*Ward4D +0.0639*Ward5A +-0.1102*Ward5B +0.1435*IC); Num7=Math.Exp(1.2798 +-0.1198*Ward5D +-0.0217*Ward4D +0.0148*Ward5A +-0.0329*Ward5B +0.1981*IC); Num8=Math.Exp(-0.2804 +-0.0352*Ward5D +0.0673*Ward4D +-0.0374*Ward5A +-0.0395*Ward5B +0.0525*IC); Num10=Math.Exp(-6.7065 +0.0787*Ward5D +0.0048*Ward4D +-0.0663*Ward5A +0.0818*Ward5B +0.2321*IC);

	Fitted Multinomial Logistic Regression models by ward of departure (Emergency Patients)
	Num11=1/(1 +Num1 +Num3 +Num4 +Num5 +Num6 +Num7 +Num8 +Num10); if(Location==1){return Num1*Num11;} else if(Location==2){return 0;} else if(Location==3){return Num3*Num11;} else if(Location==4){return Num4*Num11;} else if(Location==5){return Num5*Num11;} else if(Location==6){return Num6*Num11;} else if(Location==7){return Num7*Num11;} else if(Location==8){return Num8*Num11;} else if(Location==9){return 0;} else if(Location==10){return Num10*Num11;} else{return Num11;}
Ward 4D	Num1=Math.Exp(-26.2331 +0.2164*IC +0.1205*Ward6D +-0.1927*Other +0.5519*Ward5D +0.1883*Ward4D +-0.00358*Ward5B +0.000771*Other*Ward5B +0.4561*Ward5A +-0.0168*Ward5D*Ward5A); Num2=Math.Exp(3.9475 +-0.0263*IC +0.00248*Ward6D +-0.8402*Other +0.0782*Ward5D +0.0546*Ward4D +-0.2801*Ward5B +0.0251*Other*Ward5B +-0.0936*Ward5A +0.00161*Ward5D*Ward5A); Num4=Math.Exp(-50.6975 +0.33*IC +0.7564*Ward6D +-0.1778*Other +1.0506*Ward5D +0.1654*Ward4D +-0.5286*Ward5B +-0.00134*Other*Ward5B +1.7709*Ward5A +-0.0589*Ward5D*Ward5A);

	Fitted Multinomial Logistic Regression models by ward of departure (Emergency Patients)
	Num5=Math.Exp(-55.6678 +-0.125*IC +0.1082*Ward6D +0.1755*Other +1.7038*Ward5D +0.0948*Ward4D +0.1611*Ward5B +-0.00764*Other*Ward5B +1.7319*Ward5A +-0.0668*Ward5D*Ward5A); Num6=Math.Exp(-29.6492 +-0.0142*IC +0.0681*Ward6D +0.1493*Other +0.7441*Ward5D +0.0591*Ward4D +0.0213*Ward5B +-0.0063*Other*Ward5B +1.061*Ward5A +-0.0351*Ward5D*Ward5A); Num7=Math.Exp(-15.7105 +0.0928*IC +0.0221*Ward6D +0.2008*Other +0.3404*Ward5D +0.0113*Ward4D +0.1682*Ward5B +-0.00821*Other*Ward5B +0.45*Ward5A +-0.0191*Ward5D*Ward5A); Num8=Math.Exp(-36.1964 +0.0446*IC +-0.1798*Ward6D +0.3728*Other +1.196*Ward5D +-0.1229*Ward4D +0.2919*Ward5B +-0.0162*Other*Ward5B +1.3436*Ward5A +-0.0471*Ward5D*Ward5A); Num10=Math.Exp(-21.0388 +0.1595*IC +0.0297*Ward6D +-0.4333*Other +0.7433*Ward5D +0.0206*Ward4D +-0.1755*Ward5B +0.0148*Other*Ward5B +0.8303*Ward5A +-0.027*Ward5D*Ward5A); Num11=1/(1 +Num1 +Num2 +Num4 +Num5 +Num6 +Num7 +Num8 +Num10);

	Fitted Multinomial Logistic Regression models by ward of departure (Emergency Patients)
	if(Location==1){return Num1*Num11;} else if(Location==2){return Num2*Num11;} else if(Location==3){return 0;} else if(Location==4){return Num4*Num11;} else if(Location==5){return Num5*Num11;} else if(Location==6){return Num6*Num11;} else if(Location==7){return Num7*Num11;} else if(Location==8){return Num8*Num11;} else if(Location==9){return 0;} else if(Location==10){return Num10*Num11;} else{return Num11;}
Ward 4K	Num1=Math.Exp(-9.1618 +0.1487*Other); Num2=Math.Exp(-0.9123 +-0.2639*Other); Num3=Math.Exp(-8.323 +0.084*Other); Num5=Math.Exp(-3.8061 +-0.1265*Other); Num6=Math.Exp(-5.5714 +0.0254*Other); Num7=Math.Exp(-2.8102 +-0.3501*Other); Num10=Math.Exp(-4.8001 +0.0122*Other); Num11=1/(1 +Num1 +Num2 +Num3 +Num5 +Num6 +Num7 +Num10); if(Location==1){return Num1*Num11;} else if(Location==2){return Num2*Num11;} else if(Location==3){return Num3*Num11;}

	Fitted Multinomial Logistic Regression models by ward of departure (Emergency Patients)
	else if(Location==4){return 0;} else if(Location==5){return Num5*Num11;} else if(Location==6){return Num6*Num11;} else if(Location==7){return Num7*Num11;} else if(Location==8){return 0;} else if(Location==9){return 0;} else if(Location==10){return Num10*Num11;} else{return Num11;}
Ward 5A	Num1=Math.Exp(-42.2601 +1.4058*ED +-0.8243*IC +-1.5569*Ward4D +1.2621*Ward4K +2.037*Ward5A +-0.037*Ward4K*Ward5A +-1.8189*Ward5B +0.0503*Ward4D*Ward5B +0.7643*Ward5D +0.8792*Ward6D +-0.0636*ED*Ward6D +-0.0512*Ward5A*Ward6D +0.0347*Ward5B*Ward6D +2.2841*Northside +0.0115*ED*Northside +0.0447*IC*Northside +-0.0126*Ward5B*Northside +-0.0498*Ward5D*Northside +0.4962*Other +-0.00347*Ward5D*Other +-0.0534*Northside*Other); Num2=Math.Exp(-4.3541 +-0.8145*ED +0.1688*IC +-0.3686*Ward4D +-0.2117*Ward4K +0.2291*Ward5A +0.00928*Ward4K*Ward5A +-0.7851*Ward5B +0.0142*Ward4D*Ward5B +0.9233*Ward5D +-0.4365*Ward6D +0.028*ED*Ward6D +-0.0137*Ward5A*Ward6D +0.0183*Ward5B*Ward6D +1.2752*Northside +0.00952*ED*Northside +-0.028*IC*Northside +0.0141*Ward5B*Northside +-0.0576*Ward5D*Northside +0.4283*Other +-0.0227*Ward5D*Other +0.0081*Northside*Other);

	Fitted Multinomial Logistic Regression models by ward of departure (Emergency Patients)
	Num3=Math.Exp(-71.5623 +0.5846*ED +-0.6811*IC +0.3282*Ward4D +1.1955*Ward4K +2.243*Ward5A +-0.0439*Ward4K*Ward5A +0.3542*Ward5B +-0.0182*Ward4D*Ward5B +0.308*Ward5D +0.952*Ward6D +-0.0149*ED*Ward6D +-0.0593*Ward5A*Ward6D +0.0259*Ward5B*Ward6D +0.2923*Northside +-0.0182*ED*Northside +0.0595*IC*Northside +-0.0371*Ward5B*Northside +0.00488*Ward5D*Northside +0.7163*Other +-0.0359*Ward5D*Other +0.02*Northside*Other); Num6=Math.Exp(-17.6825 +-0.0561*ED +0.5753*IC +-0.2864*Ward4D +-0.0512*Ward4K +-0.1988*Ward5A +0.0038*Ward4K*Ward5A +0.6709*Ward5B +0.0124*Ward4D*Ward5B +-0.0871*Ward5D +0.8651*Ward6D +0.00824*ED*Ward6D +0.013*Ward5A*Ward6D +-0.0497*Ward5B*Ward6D +0.5034*Northside +-0.0168*ED*Northside +-0.0443*IC*Northside +0.00733*Ward5B*Northside +0.00401*Ward5D*Northside +-0.00372*Other +0.00155*Ward5D*Other +-0.0162*Northside*Other); Num7=Math.Exp(11.6783 +0.1118*ED +-0.165*IC +0.3967*Ward4D +0.0792*Ward4K +-0.7557*Ward5A +-0.0029*Ward4K*Ward5A +-0.0734*Ward5B +-0.0152*Ward4D*Ward5B +0.3347*Ward5D +-1.244*Ward6D +0.00671*ED*Ward6D +0.0352*Ward5A*Ward6D +0.00854*Ward5B*Ward6D +0.8726*Northside +-0.0214*ED*Northside +-0.00121*IC*Northside +0.0355*Ward5B*Northside +-0.0575*Ward5D*Northside +-0.0303*Other +-0.00114*Ward5D*Other +-0.00223*Northside*Other); Num8=Math.Exp(-24.1009 +0.4183*ED +0.3342*IC +0.0209*Ward4D +1.5828*Ward4K +2.1419*Ward5A +-0.061*Ward4K*Ward5A +-0.7781*Ward5B +0.000437*Ward4D*Ward5B +-

	Fitted Multinomial Logistic Regression models by ward of departure (Emergency Patients)
	0.4944*Ward5D +0.7478*Ward6D +-0.0244*ED*Ward6D +-0.0594*Ward5A*Ward6D +0.0249*Ward5B*Ward6D +-1.5345*Northside +0.0187*ED*Northside +-0.0152*IC*Northside +0.028*Ward5B*Northside +0.0325*Ward5D*Northside +0.1328*Other +-0.00315*Ward5D*Other +-0.0121*Northside*Other); Num10=Math.Exp(-103.8 +-0.2286*ED +-0.0182*IC +1.1647*Ward4D +-0.3293*Ward4K +1.0876*Ward5A +0.0138*Ward4K*Ward5A +1.9915*Ward5B +-0.0438*Ward4D*Ward5B +0.6534*Ward5D +2.3143*Ward6D +0.00717*ED*Ward6D +-0.0438*Ward5A*Ward6D +-0.044*Ward5B*Ward6D +-0.2205*Northside +0.00203*ED*Northside +0.00479*IC*Northside +0.0295*Ward5B*Northside +-0.0169*Ward5D*Northside +0.7091*Other +-0.0236*Ward5D*Other +-0.00566*Northside*Other); Num11=1/(1 +Num1 +Num2 +Num3 +Num6 +Num7 +Num8 +Num10); if(Location==1){return Num1*Num11;} else if(Location==2){return Num2*Num11;} else if(Location==3){return Num3*Num11;} else if(Location==4){return 0;} else if(Location==5){return 0;} else if(Location==6){return Num6*Num11;} else if(Location==7){return Num7*Num11;} else if(Location==8){return Num8*Num11;} else if(Location==9){return 0;}

	Fitted Multinomial Logistic Regression models by ward of departure (Emergency Patients)
	else if(Location==10){return Num10*Num11;} else{return Num11;}
Ward 5B	Num1=Math.Exp(-41.2994 +0.5407*IC +-0.1851*Other +3.2985*Ward5B +0.1595*Ward5D +0.9699*Ward6D +-2.6031*Ward4D +-0.1263*Ward5B*Ward6D +0.1065*Ward6D*Ward4D); Num2=Math.Exp(-40.2378 +-0.164*IC +-0.1094*Other +0.1685*Ward5B +-0.0233*Ward5D +1.4801*Ward6D +1.2753*Ward4D +-0.0064*Ward5B*Ward6D +-0.0462*Ward6D*Ward4D); Num3=Math.Exp(-25.0483 +0.1638*IC +-0.026*Other +1.4451*Ward5B +-0.1064*Ward5D +0.8796*Ward6D +-0.6623*Ward4D +-0.0513*Ward5B*Ward6D +0.021*Ward6D*Ward4D); Num4=Math.Exp(-26.3134 +-0.1945*IC +-0.1343*Other +1.4149*Ward5B +-0.0557*Ward5D +1.04*Ward6D +-0.8568*Ward4D +-0.0482*Ward5B*Ward6D +0.0244*Ward6D*Ward4D); Num5=Math.Exp(-50.649 +0.1295*IC +-0.1424*Other +1.3026*Ward5B +-0.0398*Ward5D +1.8915*Ward6D +0.4207*Ward4D +-0.0485*Ward5B*Ward6D +-0.0181*Ward6D*Ward4D); Num7=Math.Exp(-31.0874 +-0.3136*IC +-0.0544*Other +0.446*Ward5B +-0.3022*Ward5D +1.3154*Ward6D +1.0993*Ward4D +-0.00486*Ward5B*Ward6D +-0.0495*Ward6D*Ward4D); Num8=Math.Exp(-37.2694 +0.1118*IC +-0.0154*Other +1.7686*Ward5B +-0.172*Ward5D +1.4191*Ward6D +-0.2768*Ward4D +-0.0678*Ward5B*Ward6D +0.0108*Ward6D*Ward4D);

	Fitted Multinomial Logistic Regression models by ward of departure (Emergency Patients)

	`Num10=Math.Exp(-38.6545 +0.1891*IC +-0.0314*Other +1.4395*Ward5B +-0.0593*Ward5D +1.3058*Ward6D +-0.2025*Ward4D +-0.0486*Ward5B*Ward6D +0.00351*Ward6D*Ward4D);`
	
	`Num11=1/(1 +Num1 +Num2 +Num3 +Num4 +Num5 +Num7 +Num8 +Num10);`
	
	`if(Location==1){return Num1*Num11;}`
	`else if(Location==2){return Num2*Num11;}`
	`else if(Location==3){return Num3*Num11;}`
	`else if(Location==4){return Num4*Num11;}`
	`else if(Location==5){return Num5*Num11;}`
	`else if(Location==6){return 0;}`
	`else if(Location==7){return Num7*Num11;}`
	`else if(Location==8){return Num8*Num11;}`
	`else if(Location==9){return 0;}`
	`else if(Location==10){return Num10*Num11;}`
	`else{return Num11;}`

Ward 5D	`Num1=Math.Exp(-7.5182 +-0.0698*Other +-0.0383*Ward5A +0.1548*Ward5D +0.0276*IC);`
	`Num2=Math.Exp(-3.106 +-0.1662*Other +0.0955*Ward5A +0.043*Ward5D +-0.2363*IC);`
	`Num3=Math.Exp(-8.6175 +0.0106*Other +-0.1297*Ward5A +0.3127*Ward5D +-0.1333*IC);`
	`Num5=Math.Exp(0.0766 +-0.0891*Other +-0.1827*Ward5A +0.0381*Ward5D +0.0441*IC);`
	`Num6=Math.Exp(-15.4544 +-0.111*Other +0.2655*Ward5A +0.1941*Ward5D +-0.0344*IC);`
	`Num8=Math.Exp(1.0021 +-0.0866*Other +-0.1131*Ward5A +-0.0511*Ward5D +0.1187*IC);`
	`Num9=Math.Exp(4.1817 +-0.6313*Other +-0.54*Ward5A +0.18*Ward5D +0.4077*IC);`

	Fitted Multinomial Logistic Regression models by ward of departure (Emergency Patients)

	`Num10=Math.Exp(-5.4195 +-0.0136*Other +-0.0119*Ward5A +0.043*Ward5D +0.205*IC);`
	`Num11=1/(1 +Num1 +Num2 +Num3 +Num5 +Num6 +Num8 +Num9 +Num10);`
	
	`if(Location==1){return Num1*Num11;}`
	`else if(Location==2){return Num2*Num11;}`
	`else if(Location==3){return Num3*Num11;}`
	`else if(Location==4){return 0;}`
	`else if(Location==5){return Num5*Num11;}`
	`else if(Location==6){return Num6*Num11;}`
	`else if(Location==7){return 0;}`
	`else if(Location==8){return Num8*Num11;}`
	`else if(Location==9){return Num9*Num11;}`
	`else if(Location==10){return Num10*Num11;}`
	`else{return Num11;}`

Ward 6D	`Num1=Math.Exp(-2.0408 +-0.1226*Other +-0.0216*Ward6D);`
	`Num2=Math.Exp(-5.209 +-0.1845*Other +0.1695*Ward6D);`
	`Num3=Math.Exp(-6.5853 +-0.0738*Other +0.1538*Ward6D);`
	`Num5=Math.Exp(1.74 +-0.1311*Other +-0.1759*Ward6D);`
	`Num6=Math.Exp(-2.0972 +-0.0335*Other +-0.1249*Ward6D);`
	`Num7=Math.Exp(-4.2361 +-0.035*Other +0.0486*Ward6D);`
	`Num10=Math.Exp(-0.7425 +0.0237*Other +-0.1039*Ward6D);`
	`Num11=1/(1 +Num1 +Num2 +Num3 +Num5 +Num6 +Num7 +Num10);`

	Fitted Multinomial Logistic Regression models by ward of departure (Emergency Patients)

	`if(Location==1){return Num1*Num11;}`
	`else if(Location==2){return Num2*Num11;}`
	`else if(Location==3){return Num3*Num11;}`
	`else if(Location==4){return 0;}`
	`else if(Location==5){return Num5*Num11;}`
	`else if(Location==6){return Num6*Num11;}`
	`else if(Location==7){return Num7*Num11;}`
	`else if(Location==8){return 0;}`
	`else if(Location==9){return 0;}`
	`else if(Location==10){return Num10*Num11;}`
	`else{return Num11;}`

Northside	`Num10=Math.Exp(-806.4 +-8.6194*IC +33.4079*Ward4D +18.2874*Ward5A +-`
	`0.6433*Ward4D*Ward5A +-2.3045*Ward5B +3.5748*Ward5D +21.6555*Ward6D +-`
	`1.0326*IC*Ward6D +-0.6889*Ward4D*Ward6D +-19.4019*Northside +1.4505*IC*Northside`
	`+0.3421*Ward5B*Northside +-6.433*Other +0.7233*IC*Other);`
	
	`Num11=1/(1 +Num10);`
	
	`if(Location==1){return 0;}`
	`else if(Location==2){return 0;}`
	`else if(Location==3){return 0;}`
	`else if(Location==4){return 0;}`
	`else if(Location==5){return 0;}`

	Fitted Multinomial Logistic Regression models by ward of departure (Emergency Patients)
	else if(Location==6){return 0;} else if(Location==7){return 0;} else if(Location==8){return 0;} else if(Location==9){return 0;} else if(Location==10){return Num10*Num11;} else{return Num11;}
Other	Num1=Math.Exp(-16.2345 +0.0382*Ward4D +-0.0888*Other +0.3834*Ward5D +0.00698*Ward4K +0.1593*IC +1.1336*ED +-0.0353*Ward5D*ED); Num2=Math.Exp(-7.3461 +0.0432*Ward4D +-0.0878*Other +0.2186*Ward5D +0.0138*Ward4K +-0.0882*IC +0.7623*ED +-0.0257*Ward5D*ED); Num3=Math.Exp(2.654 +-0.1684*Ward4D +-0.00379*Other +0.0167*Ward5D +-0.0462*Ward4K +0.0747*IC +0.2634*ED +-0.00806*Ward5D*ED); Num4=Math.Exp(-5.575 +0.1157*Ward4D +-0.1177*Other +0.0206*Ward5D +-0.00124*Ward4K +0.0986*IC +-0.3071*ED +0.0111*Ward5D*ED); Num5=Math.Exp(-3.9681 +-0.0317*Ward4D +-0.096*Other +0.1283*Ward5D +0.0236*Ward4K +0.0648*IC +0.3072*ED +-0.00965*Ward5D*ED); Num6=Math.Exp(-11.3779 +0.0397*Ward4D +-0.0859*Other +0.3329*Ward5D +0.0288*Ward4K +-0.0257*IC +1.0714*ED +-0.0362*Ward5D*ED);

	Fitted Multinomial Logistic Regression models by ward of departure (Emergency Patients)

```
Num7=Math.Exp(-0.2562 +-0.0751*Ward4D +-0.0801*Other +0.00209*Ward5D +0.0391*Ward4K
+0.1292*IC +0.8127*ED +-0.0298*Ward5D*ED);
Num8=Math.Exp(-2.2676 +-0.0798*Ward4D +-0.0305*Other +0.0805*Ward5D +-0.013*Ward4K
+0.0764*IC +0.4232*ED +-0.0143*Ward5D*ED);

Num9=Math.Exp(2.5029 +-0.0452*Ward4D +0.0808*Other +-0.3051*Ward5D +0.1923*Ward4K +-
0.3909*IC +-2.0728*ED +0.0747*Ward5D*ED);

Num10=Math.Exp(1.0515 +-0.0114*Ward4D +-0.0719*Other +-0.0445*Ward5D +0.0479*Ward4K
+-0.0373*IC +-0.0404*ED +0.00234*Ward5D*ED);

Num11=1/(1 +Num1 +Num2 +Num3 +Num4 +Num5 +Num6 +Num7 +Num8 +Num9 +Num10);

if(Location==1){return Num1*Num11;}
else if(Location==2){return Num2*Num11;}
else if(Location==3){return Num3*Num11;}
else if(Location==4){return Num4*Num11;}
else if(Location==5){return Num5*Num11;}
else if(Location==6){return Num6*Num11;}
else if(Location==7){return Num7*Num11;}
else if(Location==8){return Num8*Num11;}
else if(Location==9){return Num9*Num11;}
```

	Fitted Multinomial Logistic Regression models by ward of departure (Emergency Patients)
	`else if(Location==10){return Num10*Num11;}` `else{return Num11;}`

D.3 MLR models for the Elective Patients

	Fitted Multinomial Logistic Regression models by ward of departure (Elective Patients)
Entry	Not Applicable. Entry location is pre-determined by the elective admissions schedule.

Emergency Department (ED)	`Num2=Math.Exp(-1.0986);`
	`Num3=Math.Exp(0.7472);`
	`Num4=Math.Exp(0.2007);`
	`Num5=Math.Exp(0.636);`
	`Num6=Math.Exp(0.2877);`
	`Num7=Math.Exp(0.3677);`
	`Num8=Math.Exp(-0.1178);`
	`Num10=Math.Exp(0.5754);`
	`Num11=1/(1 +Num2 +Num3 +Num4 +Num5 +Num6 +Num7 +Num8 +Num10);`
	
	`if(Location==1){return 0;}`
	`else if(Location==2){return Num2*Num11;}`
	`else if(Location==3){return Num3*Num11;}`
	`else if(Location==4){return Num4*Num11;}`

	Fitted Multinomial Logistic Regression models by ward of departure (Elective Patients)

	`else if(Location==5){return Num5*Num11;}`
	`else if(Location==6){return Num6*Num11;}`
	`else if(Location==7){return Num7*Num11;}`
	`else if(Location==8){return Num8*Num11;}`
	`else if(Location==9){return 0;}`
	`else if(Location==10){return Num10*Num11;}`
	`else{return Num11;}`

Intensive Care (IC)	`Num3=Math.Exp(-0.0243 +-0.335*Ward5A +0.3173*Ward5B +-0.1454*Other);`
	`Num4=Math.Exp(-17.2985 +0.000485*Ward5A +-0.0265*Ward5B +0.6518*Other);`
	`Num5=Math.Exp(4.6775 +-0.3175*Ward5A +0.2225*Ward5B +-0.1*Other);`
	`Num6=Math.Exp(2.8084 +-0.0456*Ward5A +-0.0187*Ward5B +-0.0537*Other);`
	`Num7=Math.Exp(-11.2662 +-0.352*Ward5A +0.6585*Ward5B +-0.093*Other);`
	`Num8=Math.Exp(-10.5568 +0.206*Ward5A +0.017*Ward5B +0.0419*Other);`
	`Num10=Math.Exp(-19.711 +-0.1084*Ward5A +0.7349*Ward5B +-0.0425*Other);`
	`Num11=1/(1 +Num3 +Num4 +Num5 +Num6 +Num7 +Num8 +Num10);`
	
	`if(Location==1){return 0;}`
	`else if(Location==2){return 0;}`
	`else if(Location==3){return Num3*Num11;}`
	`else if(Location==4){return Num4*Num11;}`
	`else if(Location==5){return Num5*Num11;}`
	`else if(Location==6){return Num6*Num11;}`
	`else if(Location==7){return Num7*Num11;}`

	Fitted Multinomial Logistic Regression models by ward of departure (Elective Patients)
	else if(Location==8){return Num8*Num11;} else if(Location==9){return 0;} else if(Location==10){return Num10*Num11;} else{return Num11;}
Ward 4D	Num2=Math.Exp(4.694 +-0.2468*Ward5A +0.1323*Ward5B +-0.2083*Ward5D); Num5=Math.Exp(-20.3779 +-0.2828*Ward5A +0.986*Ward5B +-0.2262*Ward5D); Num6=Math.Exp(-37.8228 +-0.1258*Ward5A +-0.0698*Ward5B +1.2506*Ward5D); Num7=Math.Exp(4.2453 +-0.00832*Ward5A +0.1704*Ward5B +-0.4266*Ward5D); Num8=Math.Exp(-25.7001 +-0.9624*Ward5A +1.4312*Ward5B +0.1044*Ward5D); Num10=Math.Exp(-13.8597 +0.1528*Ward5A +0.1351*Ward5B +0.1763*Ward5D); Num11=1/(1 +Num2 +Num5 +Num6 +Num7 +Num8 +Num10); if(Location==1){return 0;} else if(Location==2){return Num2*Num11;} else if(Location==3){return 0;} else if(Location==4){return 0;} else if(Location==5){return Num5*Num11;} else if(Location==6){return Num6*Num11;} else if(Location==7){return Num7*Num11;} else if(Location==8){return Num8*Num11;} else if(Location==9){return 0;} else if(Location==10){return Num10*Num11;} else{return Num11;}

	Fitted Multinomial Logistic Regression models by ward of departure (Elective Patients)
Ward 4K	Num2=Math.Exp(-32.8616 +1.0679*Ward5D +-0.4435*Northside); Num6=Math.Exp(112.8 +-5.5697*Ward5D +1.3863*Northside); Num7=Math.Exp(8.7037 +-0.3802*Ward5D +-0.4579*Northside); Num10=Math.Exp(8.1868 +-0.5271*Ward5D +0.2271*Northside); Num11=1/(1 +Num2 +Num6 +Num7 +Num10); if(Location==1){return 0;} else if(Location==2){return Num2*Num11;} else if(Location==3){return 0;} else if(Location==4){return 0;} else if(Location==5){return 0;} else if(Location==6){return Num2*Num11;} else if(Location==7){return Num2*Num11;} else if(Location==8){return 0;} else if(Location==9){return 0;} else if(Location==10){return Num10*Num11;} else{return Num11;}
Ward 5A	Num2=Math.Exp(-2.483 +-0.2113*IC +0.0809*Ward5D +-0.0524*Other); Num3=Math.Exp(-0.729 +-0.3183*IC +0.0278*Ward5D +-0.1751*Other); Num4=Math.Exp(-15.0357 +-0.5156*IC +0.1573*Ward5D +0.3539*Other); Num6=Math.Exp(-1.1305 +0.1895*IC +-0.0456*Ward5D +-0.1496*Other); Num7=Math.Exp(-2.7808 +0.7667*IC +-0.2025*Ward5D +-0.2774*Other); Num8=Math.Exp(13.899 +1.0024*IC +-1.1295*Ward5D +-0.00508*Other);

<table>
<tr><th></th><th>Fitted Multinomial Logistic Regression models by ward of departure (Elective Patients)</th></tr>
<tr><td></td><td><pre>Num10=Math.Exp(2.8823 +-0.1342*IC +-0.1554*Ward5D +-0.0338*Other);
Num11=1/(1 +Num2 +Num3 +Num4 +Num6 +Num7 +Num8 +Num10);

if(Location==1){return 0;}
else if(Location==2){return Num2*Num11;}
else if(Location==3){return Num3*Num11;}
else if(Location==4){return Num4*Num11;}
else if(Location==5){return 0;}
else if(Location==6){return Num6*Num11;}
else if(Location==7){return Num7*Num11;}
else if(Location==8){return Num8*Num11;}
else if(Location==9){return 0;}
else if(Location==10){return Num10*Num11;}
else{return Num11;}</pre></td></tr>
<tr><td>Ward 5B</td><td><pre>Num2=Math.Exp(-2.6303 +-0.2471*IC +0.1066*Ward5B +-0.0278*Ward5D +-0.0956*Other);
Num3=Math.Exp(1.8069 +0.5178*IC +0.0645*Ward5B +-0.477*Ward5D +-0.1594*Other);
Num4=Math.Exp(-2.2164 +-0.2348*IC +0.9444*Ward5B +-1.0798*Ward5D +-0.2494*Other);
Num5=Math.Exp(-2.761 +-0.2655*IC +0.1381*Ward5B +-0.1191*Ward5D +-0.00206*Other);
Num7=Math.Exp(-3.8224 +-0.4196*IC +0.3719*Ward5B +-0.2786*Ward5D +-0.1319*Other);
Num8=Math.Exp(33.6531 +-1.336*IC +-0.5659*Ward5B +-0.3199*Ward5D +-1.1736*Other);
Num10=Math.Exp(-5.4036 +-0.2176*IC +0.3295*Ward5B +-0.1806*Ward5D +-0.07*Other);
Num11=1/(1 +Num2 +Num3 +Num4 +Num5 +Num7 +Num8 +Num10);</pre></td></tr>
</table>

	Fitted Multinomial Logistic Regression models by ward of departure (Elective Patients)

	`if(Location==1){return 0;}`
	`else if(Location==2){return Num2*Num11;}`
	`else if(Location==3){return Num3*Num11;}`
	`else if(Location==4){return Num4*Num11;}`
	`else if(Location==5){return Num5*Num11;}`
	`else if(Location==6){return 0;}`
	`else if(Location==7){return Num7*Num11;}`
	`else if(Location==8){return Num8*Num11;}`
	`else if(Location==9){return 0;}`
	`else if(Location==10){return Num10*Num11;}`
	`else{return Num11;}`

Ward 5D	`Num2=Math.Exp(-19.3482 +0.3326*Ward4K +0.6064*Northside);`
	`Num3=Math.Exp(-7.1686 +0.3512*Ward4K +-0.3334*Northside);`
	`Num6=Math.Exp(-0.9916 +0.0737*Ward4K +-0.6058*Northside);`
	`Num8=Math.Exp(3.1519 +-0.3294*Ward4K +-0.609*Northside);`
	`Num10=Math.Exp(-3.0849 +0.1599*Ward4K +-0.1713*Northside);`
	`Num11=1/(1 +Num2 +Num3 +Num6 +Num8 +Num10);`
	
	`if(Location==1){return 0;}`
	`else if(Location==2){return Num2*Num11;}`
	`else if(Location==3){return Num3*Num11;}`
	`else if(Location==4){return 0;}`
	`else if(Location==5){return 0;}`

	Fitted Multinomial Logistic Regression models by ward of departure (Elective Patients)

	`else if(Location==6){return Num6*Num11;}`
	`else if(Location==7){return 0;}`
	`else if(Location==8){return Num8*Num11;}`
	`else if(Location==9){return 0;}`
	`else if(Location==10){return Num10*Num11;}`
	`else{return Num11;}`

Ward 6D	`Num5=Math.Exp(13.2326 +-0.3103*Ward6D +-0.4115*Ward5A +0.1555*ED);`
	`Num6=Math.Exp(-11.9763 +0.5757*Ward6D +-0.3412*Ward5A +0.0369*ED);`
	`Num7=Math.Exp(-12.1527 +0.43*Ward6D +-0.00372*Ward5A +-0.191*ED);`
	`Num10=Math.Exp(-12.8545 +-0.0506*Ward6D +0.5369*Ward5A +-0.2921*ED);`
	`Num11=1/(1 +Num5 +Num6 +Num7 +Num10);`
	
	`if(Location==1){return 0;}`
	`else if(Location==2){return 0;}`
	`else if(Location==3){return 0;}`
	`else if(Location==4){return 0;}`
	`else if(Location==5){return Num5*Num11;}`
	`else if(Location==6){return Num6*Num11;}`
	`else if(Location==7){return Num7*Num11;}`
	`else if(Location==8){return 0;}`
	`else if(Location==9){return 0;}`
	`else if(Location==10){return Num10*Num11;}`
	`else{return Num11;}`

	Fitted Multinomial Logistic Regression models by ward of departure (Elective Patients)
Northside	Num11=1; if(Location==1){return 0;} else if(Location==2){return 0;} else if(Location==3){return 0;} else if(Location==4){return 0;} else if(Location==5){return 0;} else if(Location==6){return 0;} else if(Location==7){return 0;} else if(Location==8){return 0;} else if(Location==9){return 0;} else if(Location==10){return 0;} else{return Num11;}
Other	Num1=Math.Exp(-3.3644 +-3.2767*Other +-0.1186*Ward5A +0.21*Ward5B +0.00179*Ward5A*Ward5B +3.0313*IC +-0.3481*Ward4D +-0.1077*Ward5B*IC +0.177*Ward5D +-0.0752*Ward4K +0.1207*Ward6D +0.1019*Other*Ward5B +1.7318*Northside +-0.0521*Ward5B*Northside); Num2=Math.Exp(-32.1102 +0.3476*Other +1.0311*Ward5A +1.3743*Ward5B +-0.0388*Ward5A*Ward5B +0.22*IC +-0.0991*Ward4D +-0.0174*Ward5B*IC +0.0612*Ward5D +-0.0532*Ward4K +-0.0348*Ward6D +-0.0154*Other*Ward5B +0.1263*Northside +-0.00195*Ward5B*Northside);

	Fitted Multinomial Logistic Regression models by ward of departure (Elective Patients)
	Num3=Math.Exp(-24.0151 +0.4441*Other +1.0863*Ward5A +1.1525*Ward5B +-0.0377*Ward5A*Ward5B +-1.1529*IC +-0.2291*Ward4D +0.0393*Ward5B*IC +0.0435*Ward5D +-0.1032*Ward4K +-0.0764*Ward6D +-0.0202*Other*Ward5B +0.5051*Northside +-0.0164*Ward5B*Northside); Num4=Math.Exp(4.1552 +-0.1881*Other +0.0546*Ward5A +-0.1537*Ward5B +-0.00473*Ward5A*Ward5B +-0.1344*IC +-0.0563*Ward4D +0.0139*Ward5B*IC +0.00934*Ward5D +-0.0537*Ward4K +-0.0301*Ward6D +0.00409*Other*Ward5B +0.00341*Northside +0.00662*Ward5B*Northside); Num5=Math.Exp(-37.8795 +0.1339*Other +1.0209*Ward5A +1.6499*Ward5B +-0.0454*Ward5A*Ward5B +0.2958*IC +-0.0648*Ward4D +-0.0106*Ward5B*IC +0.0277*Ward5D +-0.0798*Ward4K +0.0374*Ward6D +-0.00641*Other*Ward5B +0.896*Northside +-0.029*Ward5B*Northside); Num6=Math.Exp(-37.6681 +0.1805*Other +1.176*Ward5A +1.5415*Ward5B +-0.0431*Ward5A*Ward5B +0.6189*IC +-0.1044*Ward4D +-0.0235*Ward5B*IC +-0.0163*Ward5D +-0.0508*Ward4K +0.00578*Ward6D +-0.0057*Other*Ward5B +0.9139*Northside +-0.03*Ward5B*Northside); Num7=Math.Exp(-2.1014 +0.2438*Other +0.3863*Ward5A +0.44*Ward5B +-0.0152*Ward5A*Ward5B +-0.8123*IC +-0.1075*Ward4D +0.0295*Ward5B*IC +-0.1538*Ward5D

	Fitted Multinomial Logistic Regression models by ward of departure (Elective Patients)
	+0.0143*Ward4K +-0.073*Ward6D +-0.0149*Other*Ward5B +0.4879*Northside +-0.0135*Ward5B*Northside); Num8=Math.Exp(2.5699 +0.1957*Other +0.1536*Ward5A +0.1007*Ward5B +-0.00399*Ward5A*Ward5B +0.0302*IC +-0.1616*Ward4D +0.00357*Ward5B*IC +-0.0282*Ward5D +-0.0697*Ward4K +-0.1249*Ward6D +-0.00995*Other*Ward5B +-0.244*Northside +0.0136*Ward5B*Northside); Num10=Math.Exp(-10.9359 +-0.4616*Other +0.7108*Ward5A +0.3454*Ward5B +-0.025*Ward5A*Ward5B +0.2327*IC +0.00713*Ward4D +-0.00918*Ward5B*IC +0.1227*Ward5D +-0.00145*Ward4K +-0.0714*Ward6D +0.0136*Other*Ward5B +-0.3774*Northside +0.0139*Ward5B*Northside); Num11=1/(1 +Num1 +Num2 +Num3 +Num4 +Num5 +Num6 +Num7 +Num8 +Num10); if(Location==1){return Num1*Num11;} else if(Location==2){return Num2*Num11;} else if(Location==3){return Num3*Num11;} else if(Location==4){return Num4*Num11;} else if(Location==5){return Num5*Num11;} else if(Location==6){return Num6*Num11;} else if(Location==7){return Num7*Num11;} else if(Location==8){return Num8*Num11;}

	Fitted Multinomial Logistic Regression models by ward of departure (Elective Patients)

	`else if(Location==9){return 0;}`
	`else if(Location==10){return Num10*Num11;}`
	`else{return Num11;}`

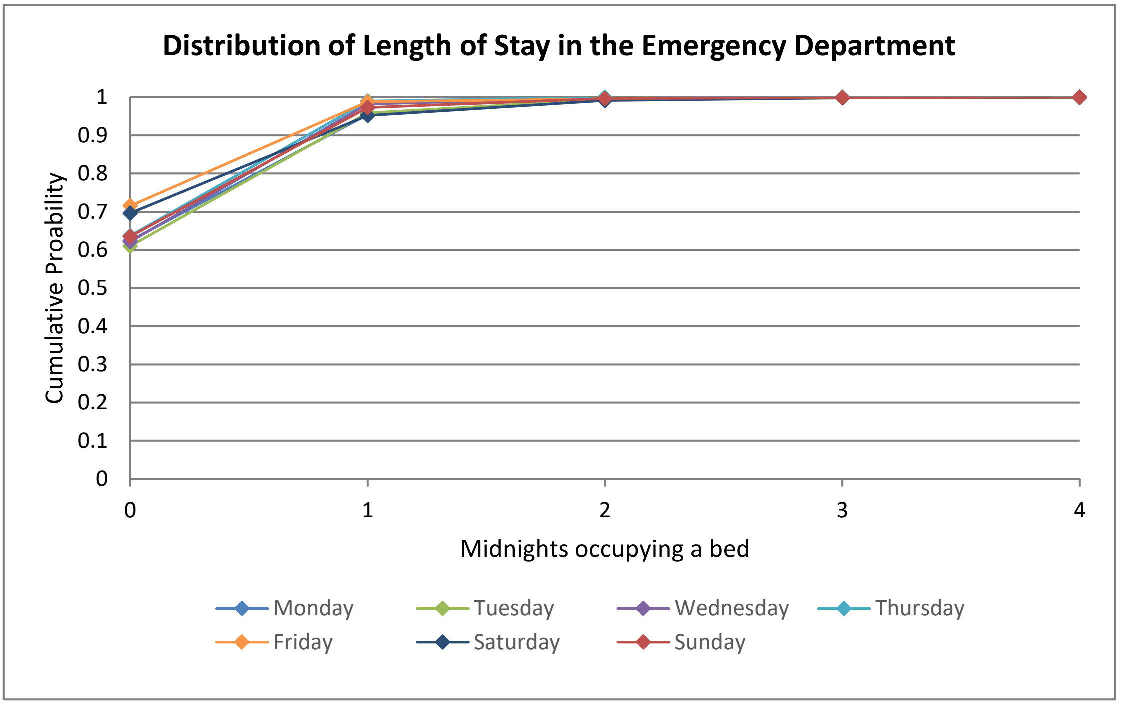
Appendix E
Length of Stay Distributions

Length of stay is disaggregated by the weekday of admission to the ward. Therefore, seven distributions are required to draw a realisations of LOS for each patient type. The rationale for disaggregating by weekday of admission is discussed in Section 4.4.3 (Modelling Length of Stay). Each empirical distribution is represented in the ODES through its cumulative distribution function, from which realisations are drawn using the Inversion Method. For further details of this method, see Section C.2.4 (Algorithms).

In this appendix, the length of stay distributions used in the ODES are charted and tabulated. Although the full distributions populate the tables, the charts may be right-truncated to improve their readability near the vertical axis. To avoid excessively long tables, only the changes in the values of the ECDFs are included, therefore the first column of each table (number of midnights) may not be continuous.

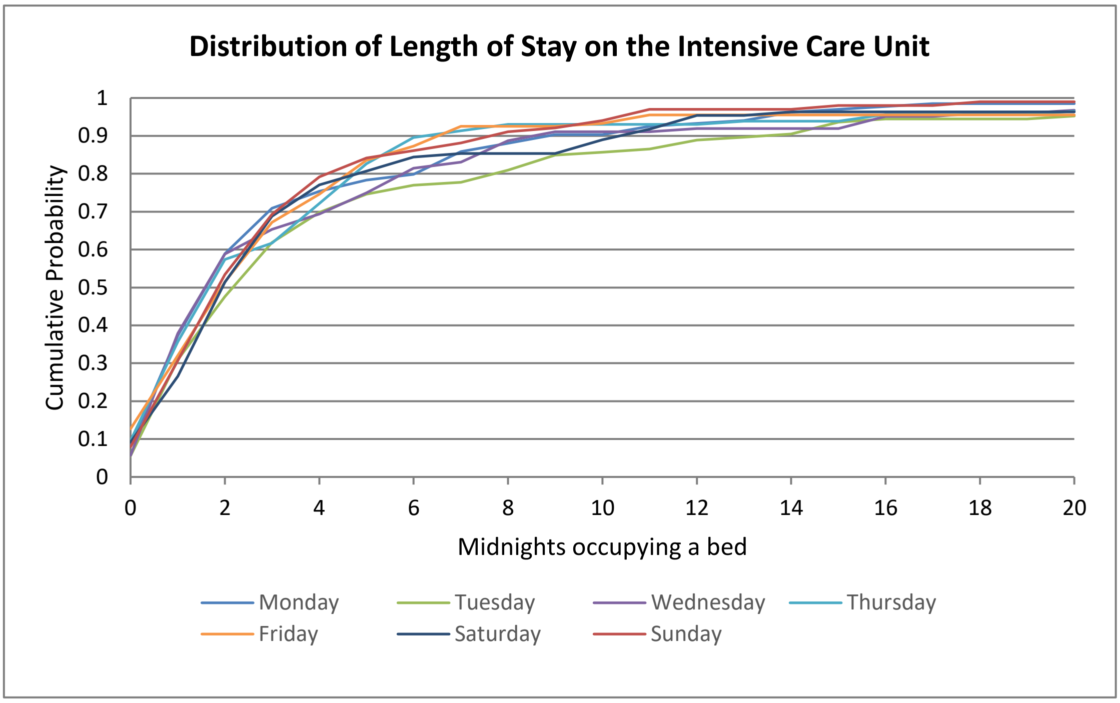
E.1 Length of Stay Distributions for the Emergency Patients

Emergency Department



	Cumulative Distribution of Length-of-Stay						
Midnights	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
0	0.624	0.610	0.622	0.637	0.715	0.696	0.635
1	0.954	0.959	0.982	0.989	0.988	0.952	0.973
2	0.995	0.994	0.998	0.999	0.995	0.992	0.996
3	1.000	1.000	1.000	1.000	0.999	0.998	0.999
4					1.000	1.000	1.000

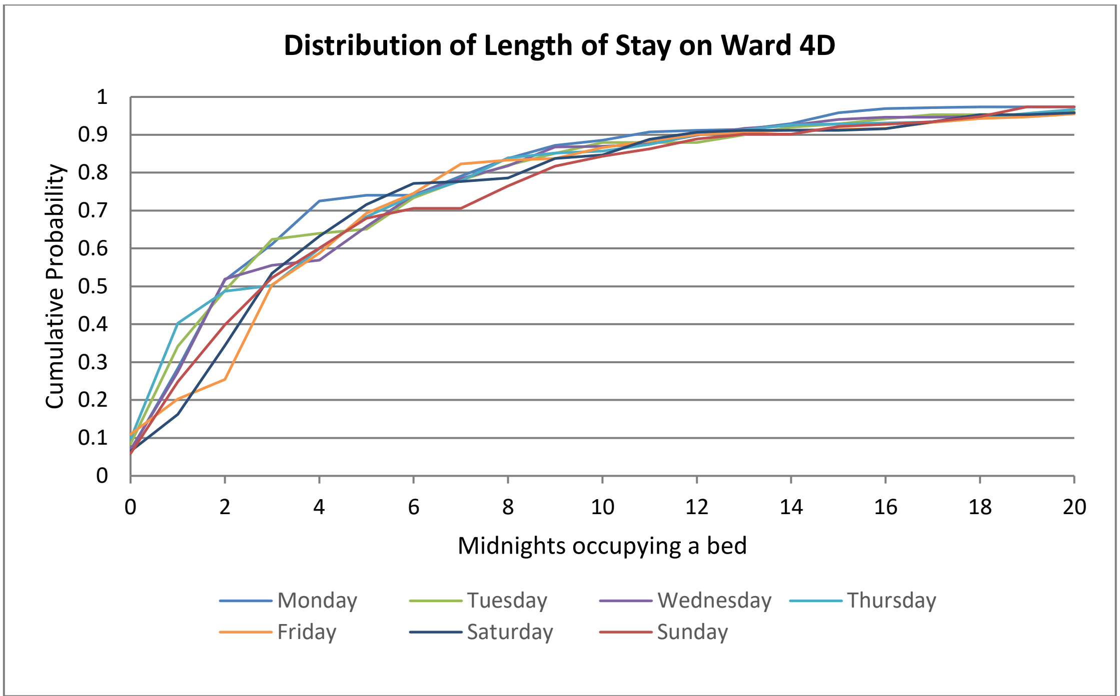
Intensive Care Unit



	Cumulative Distribution of Length-of-Stay						
Midnights	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
0	0.082	0.056	0.056	0.096	0.127	0.092	0.079
1	0.373	0.310	0.379	0.357	0.321	0.266	0.307
2	0.590	0.476	0.589	0.574	0.515	0.514	0.535
3	0.709	0.619	0.653	0.617	0.672	0.688	0.693
4	0.754	0.698	0.694	0.722	0.746	0.771	0.792
5	0.784	0.746	0.750	0.826	0.836	0.807	0.842
6	0.799	0.770	0.815	0.896	0.873	0.844	0.861
7	0.858	0.778	0.831	0.913	0.925	0.853	0.881
8	0.881	0.810	0.887	0.930	0.925	0.853	0.911
9	0.903	0.849	0.911	0.930	0.925	0.853	0.921
10	0.903	0.857	0.911	0.930	0.933	0.890	0.941
11	0.925	0.865	0.911	0.930	0.955	0.917	0.970
12	0.933	0.889	0.919	0.930	0.955	0.954	0.970
13	0.940	0.897	0.919	0.939	0.955	0.954	0.970
14	0.963	0.905	0.919	0.939	0.955	0.963	0.970
15	0.970	0.937	0.919	0.939	0.955	0.963	0.980
16	0.978	0.944	0.952	0.957	0.955	0.963	0.980
17	0.985	0.944	0.952	0.957	0.955	0.963	0.980
18	0.985	0.944	0.960	0.957	0.955	0.963	0.990
20	0.985	0.952	0.968	0.965	0.955	0.963	0.990
21	0.985	0.952	0.968	0.965	0.970	0.972	1.000
22	0.985	0.960	0.968	0.965	0.970	0.972	
23	0.985	0.968	0.968	0.965	0.970	0.972	
24	0.985	0.976	0.968	0.965	0.970	0.972	
25	0.985	0.976	0.968	0.974	0.970	0.972	
26	0.985	0.976	0.976	0.974	0.970	0.982	
27	0.985	0.984	0.976	0.983	0.970	0.982	
29	0.985	0.992	0.976	0.983	0.970	0.982	
31	0.985	0.992	0.976	0.983	0.985	0.982	
32	0.985	0.992	0.976	0.983	0.993	0.982	

	Cumulative Distribution of Length-of-Stay						
Midnights	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
33	0.985	0.992	0.984	0.983	0.993	0.982	
37	0.985	0.992	0.984	0.983	0.993	0.991	
38	1.000	0.992	0.984	0.983	0.993	0.991	
40		0.992	0.984	0.991	0.993	0.991	
45		0.992	0.984	1.000	0.993	0.991	
49		0.992	0.992		0.993	0.991	
51		1.000	0.992		0.993	0.991	
62			0.992		1.000	0.991	
76			1.000			0.991	
79						1.000	

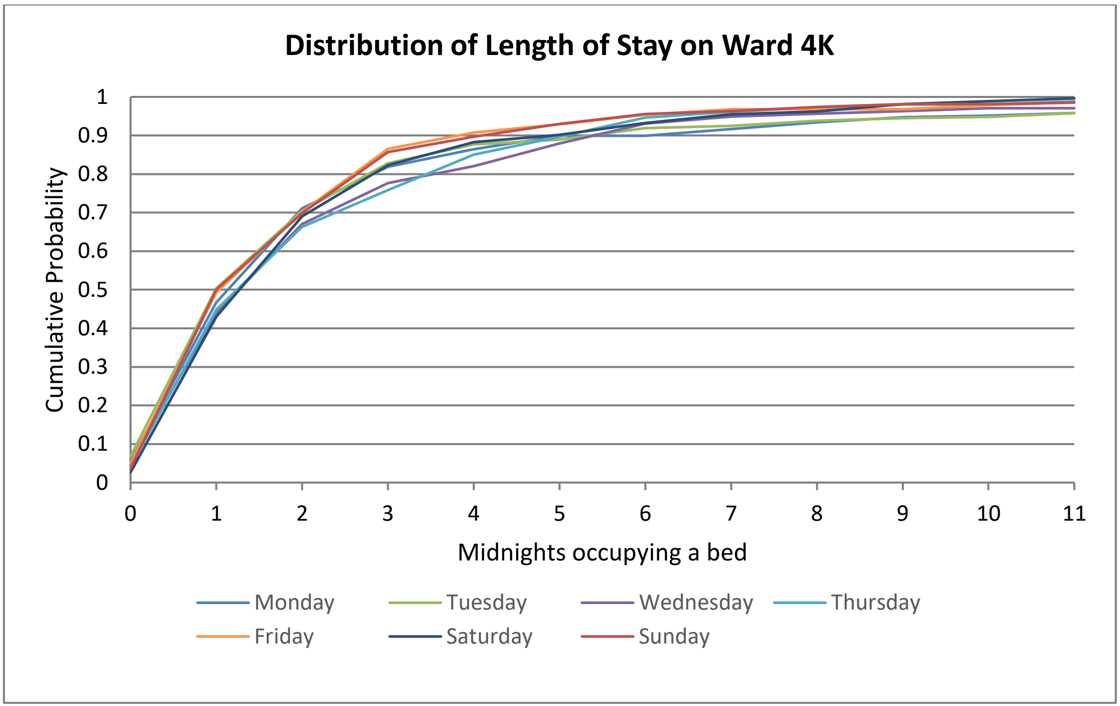
Ward 4D



	Cumulative Distribution of Length-of-Stay						
Midnights	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
0	0.066	0.084	0.069	0.097	0.110	0.065	0.059
1	0.284	0.342	0.275	0.403	0.203	0.163	0.248
2	0.516	0.489	0.520	0.487	0.254	0.344	0.399
3	0.611	0.624	0.555	0.503	0.504	0.535	0.523
4	0.725	0.640	0.569	0.599	0.587	0.633	0.601
5	0.741	0.651	0.658	0.684	0.693	0.716	0.680
6	0.741	0.733	0.741	0.740	0.746	0.772	0.706
7	0.791	0.780	0.785	0.778	0.823	0.777	0.706
8	0.837	0.820	0.818	0.839	0.833	0.786	0.765
9	0.873	0.851	0.868	0.852	0.837	0.837	0.817
10	0.886	0.880	0.870	0.857	0.866	0.847	0.843
11	0.908	0.880	0.875	0.875	0.884	0.888	0.863
12	0.912	0.880	0.899	0.901	0.900	0.907	0.889
13	0.914	0.900	0.917	0.913	0.907	0.912	0.902
14	0.930	0.920	0.925	0.926	0.913	0.912	0.902
15	0.958	0.929	0.941	0.929	0.915	0.912	0.922
16	0.969	0.942	0.947	0.931	0.917	0.916	0.928
17	0.971	0.953	0.947	0.934	0.933	0.935	0.935
18	0.974	0.953	0.949	0.944	0.943	0.953	0.948
19	0.974	0.953	0.953	0.957	0.947	0.953	0.974
20	0.974	0.960	0.957	0.967	0.955	0.958	0.974
21	0.974	0.962	0.968	0.972	0.959	0.967	0.974
22	0.974	0.962	0.974	0.972	0.959	0.967	0.974
23	0.980	0.967	0.974	0.972	0.959	0.967	0.987
24	0.982	0.969	0.976	0.972	0.963	0.977	0.987
25	0.982	0.971	0.976	0.982	0.967	0.977	0.993
26	0.982	0.971	0.978	0.982	0.972	0.977	1.000
27	0.982	0.976	0.982	0.987	0.976	0.977	
28	0.982	0.978	0.986	0.987	0.978	0.977	
29	0.987	0.980	0.986	0.990	0.978	0.977	

	Cumulative Distribution of Length-of-Stay						
Midnights	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
30	0.991	0.982	0.986	0.990	0.978	0.977	
31	0.991	0.984	0.986	0.990	0.984	0.981	
32	0.991	0.984	0.986	0.990	0.986	0.981	
33	0.991	0.984	0.988	0.990	0.986	0.981	
34	0.991	0.987	0.988	0.990	0.986	0.986	
35	0.991	0.987	0.988	0.992	0.986	0.986	
36	0.991	0.987	0.988	0.997	0.986	0.986	
37	0.991	0.987	0.992	0.997	0.986	0.986	
38	0.996	0.987	0.994	0.997	0.986	0.986	
39	0.996	0.987	0.994	0.997	0.986	0.991	
40	0.996	0.987	0.996	0.997	0.988	0.991	
41	0.996	0.987	0.996	0.997	0.990	0.991	
42	0.996	0.987	0.996	0.997	0.992	0.991	
43	0.996	0.987	0.998	0.997	0.992	0.991	
45	0.996	0.989	0.998	0.997	0.992	0.991	
46	0.998	0.989	0.998	0.997	0.994	0.995	
47	0.998	0.989	0.998	1.000	0.994	0.995	
50	0.998	0.991	1.000		0.994	0.995	
54	0.998	0.991			0.996	0.995	
55	0.998	0.993			0.996	0.995	
58	0.998	0.996			0.996	0.995	
60	0.998	0.996			0.998	0.995	
75	0.998	0.998			0.998	0.995	
84	1.000	0.998			0.998	0.995	
93		0.998			0.998	1.000	
132		1.000			0.998		
172					1.000		

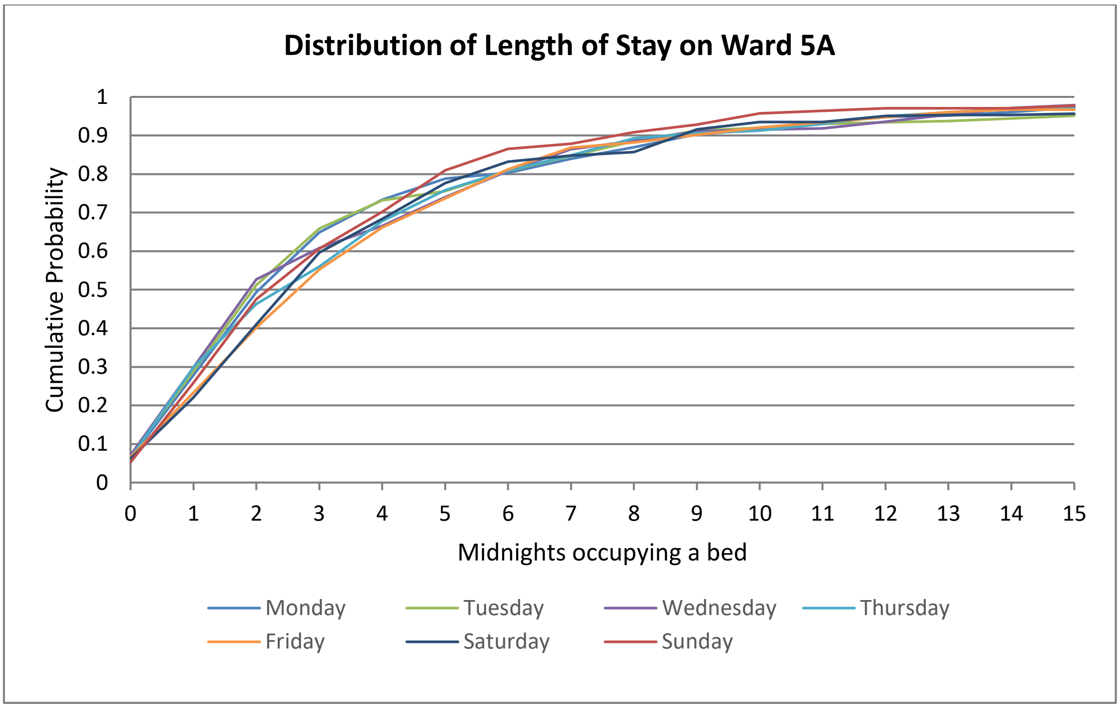
Ward 4K



	Cumulative Distribution of Length-of-Stay						
Midnights	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
0	0.056	0.068	0.051	0.042	0.050	0.026	0.037
1	0.467	0.503	0.440	0.448	0.493	0.430	0.500
2	0.711	0.705	0.670	0.663	0.702	0.691	0.699
3	0.819	0.828	0.777	0.759	0.865	0.823	0.857
4	0.864	0.877	0.821	0.851	0.908	0.883	0.897
5	0.899	0.890	0.879	0.897	0.929	0.902	0.930
6	0.899	0.919	0.930	0.946	0.954	0.932	0.956
7	0.916	0.925	0.949	0.962	0.968	0.955	0.963
8	0.934	0.938	0.956	0.973	0.968	0.962	0.974
9	0.948	0.945	0.963	0.981	0.968	0.981	0.982
10	0.951	0.948	0.971	0.981	0.979	0.989	0.982
11	0.958	0.958	0.971	0.989	0.986	0.996	0.985
12	0.958	0.964	0.974	0.989	0.989	0.996	0.989
13	0.962	0.964	0.974	0.992	0.989	0.996	0.989
14	0.972	0.964	0.974	0.992	0.993	0.996	0.989
15	0.979	0.974	0.978	0.992	0.993	0.996	0.996
16	0.983	0.974	0.978	0.992	0.993	0.996	0.996
17	0.983	0.977	0.978	0.992	0.993	0.996	1.000
18	0.983	0.977	0.982	0.992	0.996	0.996	
19	0.986	0.977	0.982	0.992	0.996	0.996	
20	0.986	0.977	0.982	0.996	0.996	0.996	
21	0.986	0.981	0.982	0.996	0.996	0.996	
28	0.986	0.987	0.982	0.996	0.996	0.996	
29	0.986	0.990	0.985	0.996	0.996	0.996	
31	0.986	0.990	0.989	0.996	0.996	0.996	
34	0.986	0.990	0.993	0.996	0.996	0.996	
36	0.986	0.994	0.993	0.996	0.996	0.996	
37	0.990	0.994	0.996	0.996	0.996	0.996	
41	0.990	0.994	1.000	0.996	0.996	0.996	
42	0.990	0.997		0.996	0.996	0.996	

	Cumulative Distribution of Length-of-Stay						
Midnights	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
43	0.993	0.997		0.996	0.996	0.996	
44	0.993	0.997		1.000	0.996	0.996	
50	0.997	0.997			0.996	0.996	
52	0.997	1.000			0.996	0.996	
59	0.997				0.996	1.000	
71	1.000				0.996		
150					1.000		

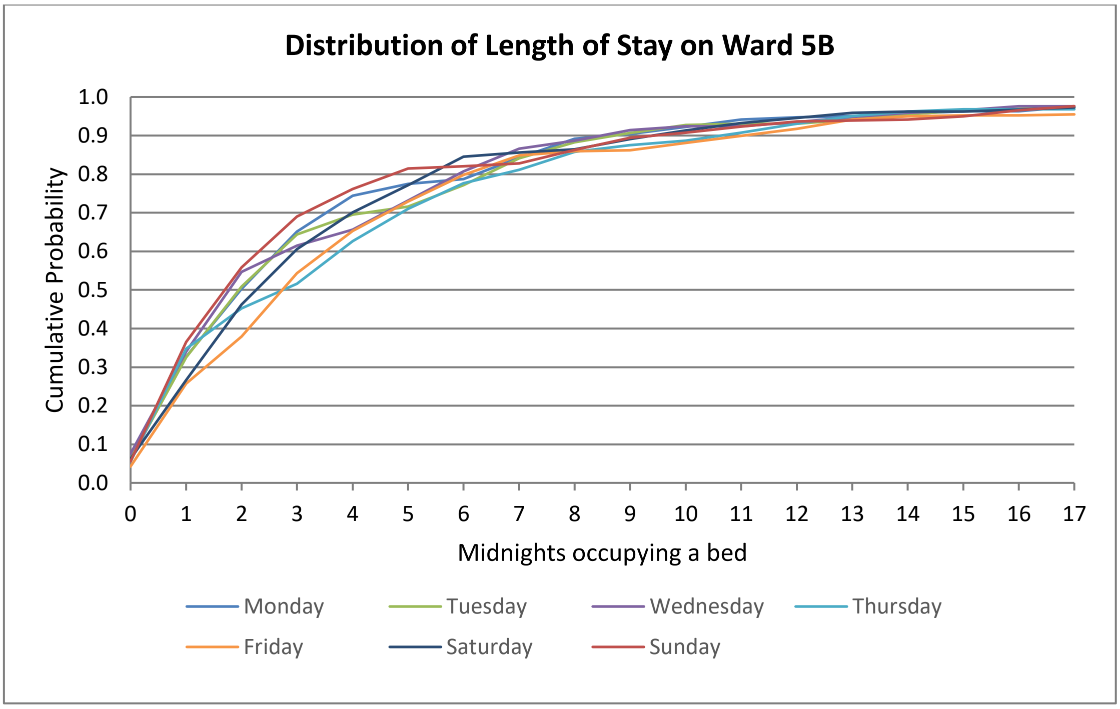
Ward 5A



	Cumulative Distribution of Length-of-Stay						
Midnights	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
0	0.064	0.070	0.071	0.064	0.066	0.062	0.052
1	0.279	0.289	0.299	0.299	0.234	0.220	0.259
2	0.494	0.512	0.527	0.463	0.401	0.410	0.475
3	0.648	0.659	0.609	0.560	0.553	0.596	0.607
4	0.733	0.732	0.665	0.678	0.661	0.683	0.702
5	0.788	0.756	0.740	0.758	0.737	0.776	0.810
6	0.803	0.808	0.808	0.809	0.813	0.832	0.866
7	0.839	0.847	0.865	0.849	0.868	0.848	0.879
8	0.870	0.885	0.886	0.893	0.882	0.857	0.908
9	0.903	0.913	0.911	0.906	0.901	0.916	0.928
10	0.915	0.920	0.915	0.913	0.921	0.935	0.957
11	0.933	0.930	0.918	0.930	0.934	0.935	0.964
12	0.948	0.934	0.936	0.950	0.947	0.950	0.970
13	0.952	0.937	0.954	0.960	0.961	0.953	0.970
14	0.961	0.944	0.972	0.970	0.967	0.953	0.970
15	0.973	0.951	0.979	0.973	0.967	0.957	0.977
16	0.976	0.958	0.986	0.973	0.967	0.957	0.984
17	0.976	0.969	0.986	0.973	0.974	0.957	0.984
18	0.976	0.969	0.986	0.977	0.980	0.960	0.987
19	0.976	0.969	0.993	0.993	0.980	0.960	0.987
20	0.976	0.972	0.993	0.993	0.984	0.963	0.987
21	0.979	0.976	0.993	0.993	0.984	0.966	0.987
22	0.982	0.976	1.000	0.993	0.984	0.966	0.987
23	0.985	0.976		0.993	0.984	0.966	0.987
24	0.985	0.976		0.993	0.990	0.966	0.990
25	0.985	0.976		0.993	0.993	0.966	0.990
26	0.985	0.976		0.993	0.993	0.972	0.990
27	0.985	0.979		0.993	0.993	0.975	0.990
28	0.985	0.979		0.993	0.993	0.978	0.990

	Cumulative Distribution of Length-of-Stay						
Midnights	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
30	0.985	0.979		0.993	0.993	0.978	0.993
33	0.985	0.979		0.997	0.997	0.978	0.993
34	0.985	0.983		0.997	0.997	0.978	0.993
35	0.985	0.990		0.997	0.997	0.978	0.993
36	0.988	0.990		0.997	0.997	0.978	0.993
38	0.988	0.990		0.997	0.997	0.984	0.993
40	0.988	0.990		0.997	0.997	0.991	0.993
42	0.988	0.993		0.997	0.997	0.991	0.993
44	0.988	0.993		0.997	0.997	0.994	0.993
50	0.988	0.997		0.997	0.997	0.994	0.997
56	0.994	0.997		0.997	0.997	0.994	0.997
59	0.994	0.997		0.997	0.997	0.994	1.000
60	0.997	0.997		0.997	0.997	0.994	
66	0.997	0.997		0.997	1.000	0.994	
81	1.000	0.997		0.997		0.994	
87		0.997		1.000		0.994	
116		0.997				0.997	
171		1.000				0.997	
216						1.000	

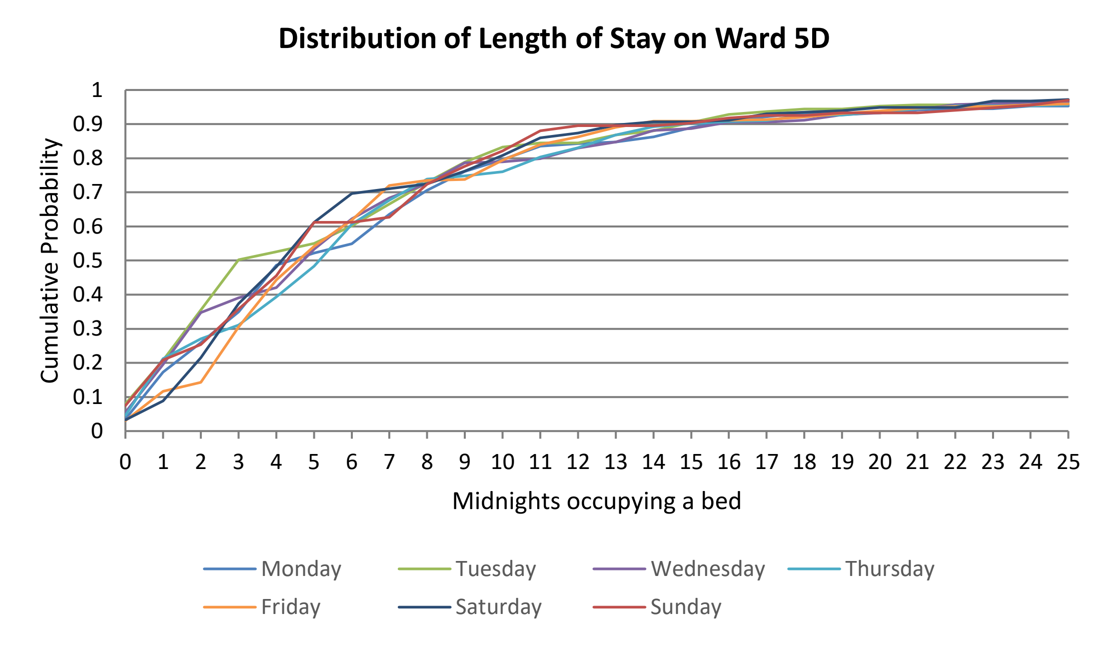
Ward 5B



	Cumulative Distribution of Length-of-Stay						
Midnights	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
0	0.062	0.062	0.076	0.055	0.042	0.063	0.056
1	0.327	0.325	0.340	0.348	0.257	0.266	0.365
2	0.503	0.509	0.546	0.452	0.379	0.462	0.558
3	0.651	0.644	0.615	0.516	0.544	0.606	0.690
4	0.744	0.696	0.656	0.626	0.653	0.701	0.762
5	0.775	0.716	0.732	0.710	0.729	0.772	0.815
6	0.787	0.772	0.808	0.777	0.798	0.845	0.820
7	0.843	0.841	0.866	0.812	0.849	0.856	0.828
8	0.892	0.882	0.887	0.858	0.859	0.864	0.862
9	0.904	0.907	0.914	0.875	0.862	0.891	0.894
10	0.923	0.927	0.924	0.887	0.881	0.913	0.907
11	0.941	0.931	0.924	0.907	0.899	0.932	0.923
12	0.948	0.931	0.935	0.930	0.918	0.946	0.937
13	0.951	0.941	0.948	0.951	0.942	0.959	0.939
14	0.957	0.955	0.959	0.962	0.950	0.962	0.942
15	0.963	0.965	0.966	0.968	0.952	0.962	0.950
16	0.963	0.969	0.976	0.968	0.952	0.967	0.966
17	0.975	0.972	0.976	0.968	0.955	0.973	0.976
18	0.978	0.972	0.976	0.971	0.960	0.981	0.979
19	0.978	0.972	0.976	0.980	0.963	0.989	0.984
20	0.978	0.976	0.979	0.983	0.968	0.992	0.984
21	0.978	0.979	0.979	0.986	0.968	0.992	0.984
22	0.978	0.979	0.983	0.988	0.968	0.992	0.984
23	0.981	0.983	0.990	0.988	0.971	0.992	0.989
24	0.981	0.983	0.990	0.991	0.971	0.992	0.992
25	0.981	0.983	0.990	0.994	0.973	0.992	0.997
26	0.981	0.983	0.993	0.994	0.976	0.992	0.997
27	0.981	0.983	0.997	0.994	0.979	0.992	0.997
28	0.988	0.983	0.997	0.994	0.979	0.992	0.997
30	0.988	0.986	0.997	0.994	0.979	0.992	0.997

	Cumulative Distribution of Length-of-Stay						
Midnights	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
32	0.988	0.986	0.997	0.997	0.979	0.995	0.997
33	0.988	0.986	0.997	0.997	0.981	0.995	0.997
34	0.988	0.986	0.997	0.997	0.987	0.995	0.997
35	0.991	0.990	0.997	0.997	0.989	0.995	0.997
36	0.991	0.990	0.997	0.997	0.989	0.995	1.000
38	0.997	0.990	0.997	0.997	0.989	0.995	1.000
39	0.997	0.990	0.997	0.997	0.995	0.995	1.000
41	0.997	0.993	1.000	0.997	0.997	0.997	1.000
42	0.997	0.997	1.000	0.997	0.997	0.997	1.000
49	1.000	0.997	1.000	1.000	0.997	0.997	1.000
52	1.000	0.997	1.000	1.000	0.997	1.000	1.000
56	1.000	0.997	1.000	1.000	1.000	1.000	1.000
70	1.000	1.000	1.000	1.000	1.000	1.000	1.000

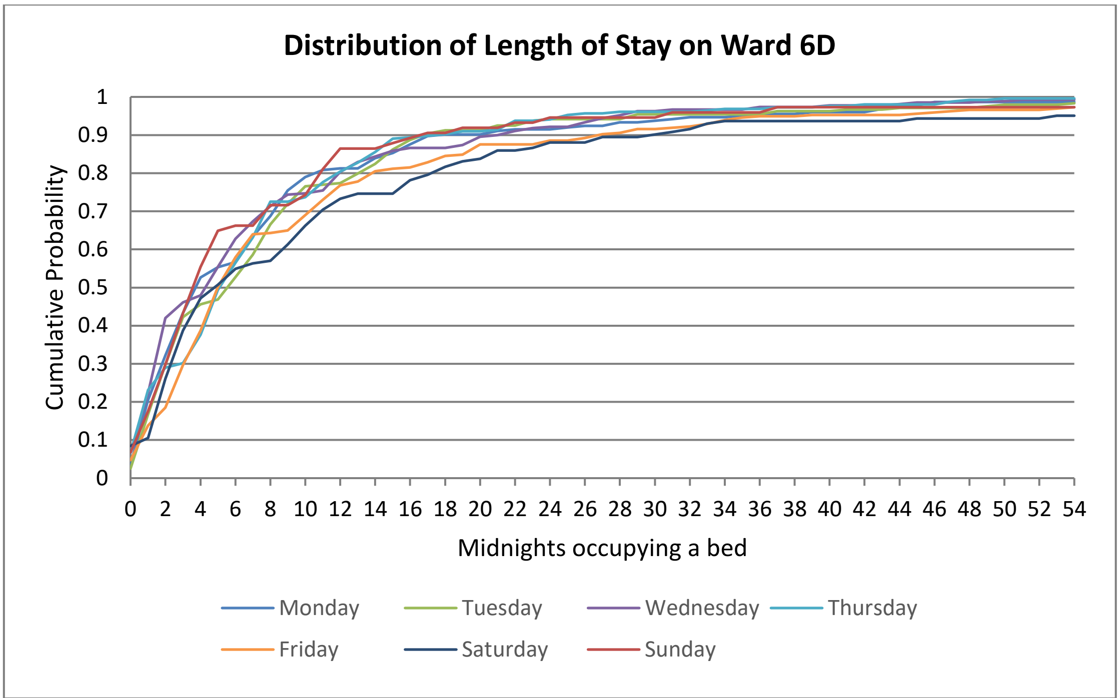
Ward 5D



	Cumulative Distribution of Length-of-Stay						
Midnights	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
0	0.035	0.080	0.055	0.046	0.031	0.033	0.075
1	0.173	0.207	0.195	0.212	0.117	0.089	0.209
2	0.259	0.355	0.348	0.271	0.142	0.215	0.254
3	0.349	0.502	0.390	0.311	0.305	0.374	0.358
4	0.486	0.526	0.421	0.394	0.443	0.481	0.455
5	0.522	0.550	0.534	0.483	0.542	0.612	0.612
6	0.549	0.602	0.622	0.606	0.618	0.696	0.612
7	0.635	0.665	0.683	0.677	0.720	0.710	0.627
8	0.706	0.729	0.726	0.738	0.735	0.724	0.724
9	0.761	0.789	0.787	0.748	0.738	0.762	0.776
10	0.796	0.833	0.790	0.760	0.794	0.808	0.821
11	0.835	0.845	0.799	0.803	0.840	0.860	0.881
12	0.843	0.845	0.829	0.831	0.863	0.874	0.896
13	0.847	0.869	0.848	0.868	0.891	0.897	0.896
14	0.863	0.880	0.881	0.892	0.908	0.907	0.896
15	0.890	0.904	0.887	0.902	0.908	0.907	0.903
16	0.910	0.928	0.905	0.908	0.911	0.911	0.918
17	0.922	0.936	0.905	0.911	0.913	0.930	0.925
18	0.933	0.944	0.912	0.926	0.921	0.935	0.925
19	0.933	0.944	0.927	0.926	0.931	0.939	0.933
20	0.933	0.952	0.933	0.935	0.939	0.949	0.933
21	0.941	0.956	0.942	0.945	0.947	0.949	0.933
22	0.945	0.956	0.957	0.954	0.952	0.949	0.940
23	0.945	0.956	0.960	0.954	0.952	0.967	0.948
24	0.953	0.964	0.963	0.954	0.957	0.967	0.955
25	0.953	0.964	0.966	0.957	0.959	0.972	0.970
26	0.957	0.964	0.976	0.960	0.964	0.977	0.970
27	0.957	0.964	0.976	0.969	0.975	0.977	0.970
28	0.961	0.964	0.979	0.972	0.975	0.977	0.970
29	0.961	0.968	0.979	0.978	0.975	0.977	0.970
30	0.969	0.976	0.982	0.978	0.975	0.977	0.985

	Cumulative Distribution of Length-of-Stay						
Midnights	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
31	0.973	0.980	0.982	0.978	0.977	0.977	0.985
32	0.976	0.980	0.985	0.982	0.980	0.981	0.985
33	0.976	0.980	0.985	0.991	0.985	0.981	0.985
34	0.976	0.980	0.988	0.991	0.987	0.986	0.985
35	0.976	0.980	0.991	0.991	0.987	0.986	0.985
36	0.980	0.984	0.991	0.991	0.987	0.986	0.985
37	0.984	0.988	0.991	0.991	0.987	0.986	0.985
38	0.984	0.992	0.991	0.991	0.992	0.986	0.993
40	0.984	0.992	0.991	0.994	0.992	0.986	0.993
41	0.984	0.992	0.997	0.994	0.992	0.986	0.993
42	0.984	0.992	1.000	0.994	0.995	0.986	0.993
43	0.984	0.996		0.994	0.995	0.991	0.993
45	0.984	0.996		0.994	0.995	0.991	1.000
46	0.988	0.996		0.994	0.995	0.995	
47	0.988	0.996		0.997	0.995	0.995	
50	0.992	0.996		0.997	0.995	0.995	
51	0.996	0.996		0.997	0.995	0.995	
52	1.000	0.996		0.997	0.995	0.995	
57		0.996		1.000	0.995	0.995	
62		0.996			0.995	1.000	
66		0.996			0.997		
68		0.996			1.000		
139		1.000					

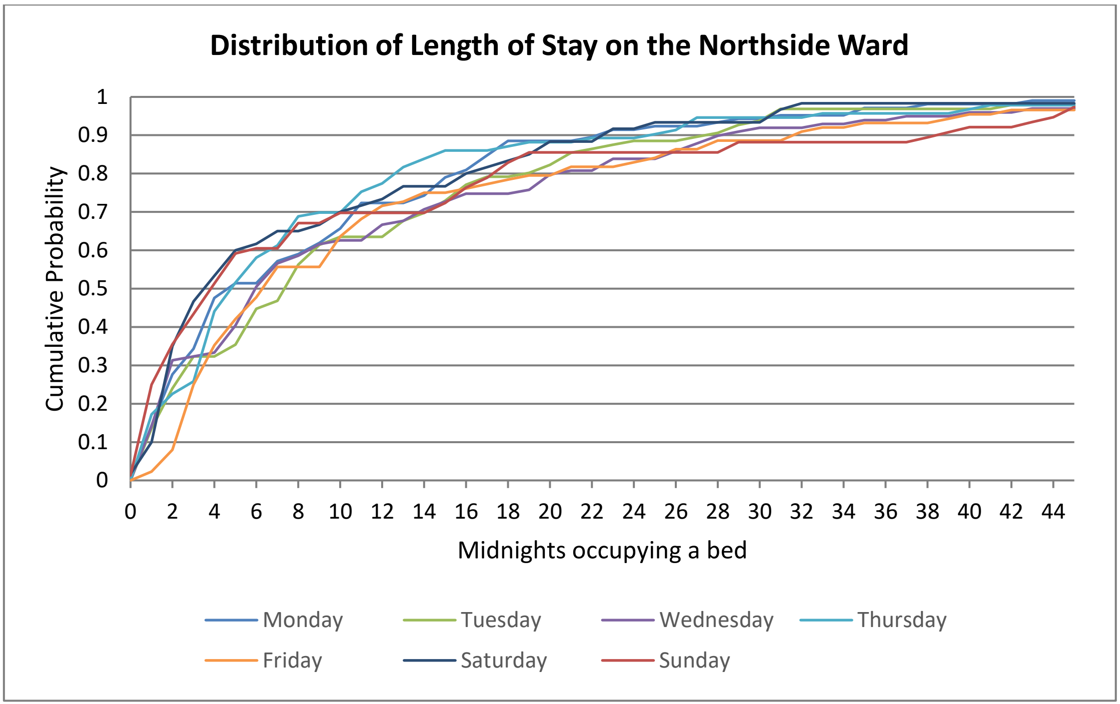
Ward 6D



	Cumulative Distribution of Length-of-Stay						
Midnights	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
0	0.036	0.025	0.067	0.067	0.047	0.085	0.068
1	0.205	0.167	0.223	0.231	0.138	0.106	0.176
2	0.321	0.297	0.420	0.290	0.185	0.261	0.297
3	0.433	0.423	0.461	0.302	0.296	0.387	0.432
4	0.527	0.456	0.480	0.376	0.387	0.472	0.554
5	0.554	0.469	0.554	0.494	0.502	0.507	0.649
6	0.567	0.527	0.628	0.565	0.579	0.549	0.662
7	0.634	0.586	0.673	0.631	0.640	0.563	0.662
8	0.688	0.665	0.714	0.725	0.643	0.570	0.716
9	0.754	0.720	0.743	0.725	0.650	0.613	0.716
10	0.790	0.766	0.747	0.737	0.690	0.662	0.743
11	0.808	0.770	0.755	0.776	0.731	0.704	0.811
12	0.813	0.774	0.803	0.804	0.768	0.732	0.865
13	0.813	0.799	0.829	0.827	0.778	0.746	0.865
14	0.839	0.824	0.844	0.855	0.805	0.746	0.865
15	0.853	0.862	0.859	0.890	0.811	0.746	0.878
16	0.875	0.887	0.866	0.894	0.815	0.782	0.892
17	0.897	0.904	0.866	0.898	0.828	0.796	0.905
18	0.902	0.912	0.866	0.902	0.845	0.817	0.905
19	0.902	0.912	0.874	0.910	0.848	0.831	0.919
20	0.902	0.912	0.896	0.910	0.875	0.838	0.919
21	0.911	0.925	0.900	0.914	0.875	0.859	0.919
22	0.915	0.925	0.911	0.937	0.875	0.859	0.932
23	0.915	0.937	0.918	0.937	0.875	0.866	0.932
24	0.915	0.941	0.922	0.941	0.886	0.880	0.946
25	0.920	0.941	0.922	0.953	0.886	0.880	0.946
26	0.924	0.941	0.933	0.957	0.892	0.880	0.946
27	0.924	0.941	0.944	0.957	0.902	0.894	0.946
28	0.933	0.941	0.952	0.961	0.906	0.894	0.946

	Cumulative Distribution of Length-of-Stay						
Midnights	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
29	0.933	0.954	0.963	0.961	0.916	0.894	0.946
30	0.938	0.954	0.963	0.961	0.916	0.901	0.946
31	0.942	0.954	0.967	0.961	0.919	0.908	0.959
32	0.946	0.954	0.967	0.961	0.923	0.915	0.959
33	0.946	0.954	0.967	0.965	0.929	0.930	0.959
34	0.946	0.954	0.967	0.969	0.939	0.937	0.959
35	0.951	0.954	0.967	0.969	0.946	0.937	0.959
36	0.955	0.954	0.974	0.969	0.949	0.937	0.959
37	0.955	0.962	0.974	0.973	0.949	0.937	0.973
39	0.960	0.962	0.974	0.973	0.953	0.937	0.973
40	0.960	0.962	0.978	0.976	0.953	0.937	0.973
41	0.960	0.967	0.978	0.976	0.953	0.937	0.973
42	0.960	0.967	0.978	0.980	0.953	0.937	0.973
43	0.969	0.967	0.978	0.980	0.953	0.937	0.973
44	0.978	0.971	0.981	0.980	0.953	0.937	0.973
45	0.978	0.971	0.985	0.980	0.956	0.944	0.973
46	0.987	0.971	0.985	0.980	0.960	0.944	0.973
47	0.987	0.971	0.985	0.988	0.963	0.944	0.973
48	0.987	0.971	0.985	0.992	0.966	0.944	0.973
49	0.987	0.975	0.989	0.992	0.966	0.944	0.973
50	0.987	0.979	0.989	0.996	0.966	0.944	0.973
51	0.987	0.979	0.993	0.996	0.966	0.944	0.973
53	0.987	0.979	0.993	0.996	0.970	0.951	0.973
54	0.987	0.983	0.993	0.996	0.973	0.951	0.973
55	0.987	0.983	0.993	0.996	0.980	0.958	0.973
56	0.991	0.983	0.993	0.996	0.980	0.958	0.973
57	0.991	0.992	0.993	0.996	0.980	0.958	0.973
59	0.991	0.992	0.993	0.996	0.983	0.958	0.973
62	0.991	0.992	0.996	0.996	0.987	0.958	0.973
66	0.991	0.992	0.996	0.996	0.990	0.958	0.973
67	0.991	0.992	0.996	0.996	0.990	0.958	0.986
68	0.991	0.992	0.996	0.996	0.990	0.972	1.000
69	0.991	0.992	0.996	0.996	0.993	0.972	
70	0.991	0.992	0.996	0.996	1.000	0.972	
71	0.991	0.996	0.996	0.996		0.979	
74	0.991	0.996	0.996	0.996		0.986	
86	0.991	1.000	0.996	0.996		0.986	
87	0.996		0.996	0.996		0.986	
89	0.996		0.996	0.996		0.993	
117	0.996		0.996	1.000		0.993	
123	0.996		0.996			1.000	
185	1.000		0.996				
208			1.000				

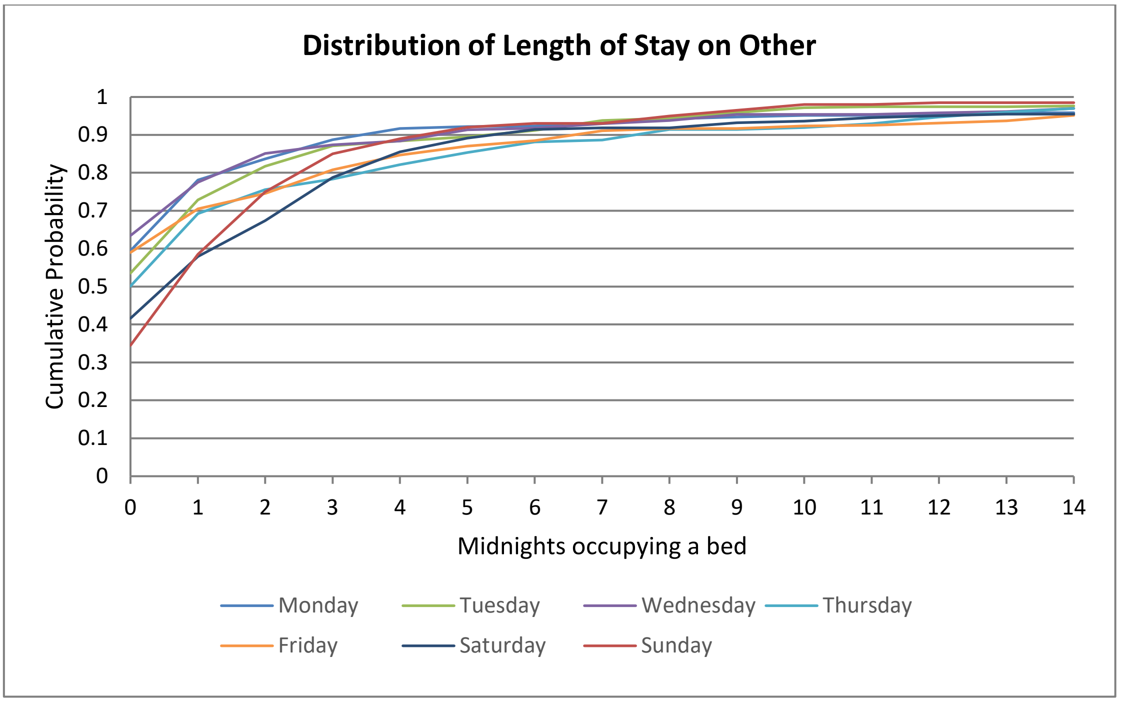
Northside Ward



	Cumulative Distribution of Length-of-Stay						
Midnights	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
0	0.000	0.000	0.000	0.000	0.000	0.017	0.013
1	0.143	0.135	0.141	0.172	0.023	0.100	0.250
2	0.276	0.240	0.313	0.226	0.080	0.350	0.355
3	0.343	0.323	0.323	0.258	0.250	0.467	0.434
4	0.476	0.323	0.333	0.441	0.352	0.533	0.513
5	0.514	0.354	0.404	0.516	0.420	0.600	0.592
6	0.514	0.448	0.505	0.581	0.477	0.617	0.605
7	0.571	0.469	0.566	0.613	0.557	0.650	0.605
8	0.590	0.563	0.586	0.688	0.557	0.650	0.671
9	0.619	0.615	0.616	0.699	0.557	0.667	0.671
10	0.657	0.635	0.626	0.699	0.636	0.700	0.697
11	0.724	0.635	0.626	0.753	0.682	0.717	0.697
12	0.724	0.635	0.667	0.774	0.716	0.733	0.697
13	0.724	0.677	0.677	0.817	0.727	0.767	0.697
14	0.743	0.698	0.707	0.839	0.750	0.767	0.697
15	0.790	0.729	0.727	0.860	0.750	0.767	0.724
16	0.810	0.771	0.747	0.860	0.761	0.800	0.763
17	0.848	0.792	0.747	0.860	0.773	0.817	0.789
18	0.886	0.792	0.747	0.871	0.784	0.833	0.829
19	0.886	0.802	0.758	0.882	0.795	0.850	0.855
20	0.886	0.823	0.798	0.882	0.795	0.883	0.855
21	0.886	0.854	0.808	0.882	0.818	0.883	0.855
22	0.895	0.865	0.808	0.892	0.818	0.883	0.855
23	0.914	0.875	0.838	0.892	0.818	0.917	0.855
24	0.914	0.885	0.838	0.892	0.830	0.917	0.855
25	0.924	0.885	0.838	0.903	0.841	0.933	0.855
26	0.924	0.885	0.859	0.914	0.864	0.933	0.855
27	0.924	0.896	0.879	0.946	0.864	0.933	0.855
28	0.933	0.906	0.899	0.946	0.886	0.933	0.855

	Cumulative Distribution of Length-of-Stay						
Midnights	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
29	0.943	0.927	0.909	0.946	0.886	0.933	0.882
30	0.943	0.938	0.919	0.946	0.886	0.933	0.882
31	0.952	0.969	0.919	0.946	0.886	0.967	0.882
32	0.952	0.969	0.919	0.946	0.909	0.983	0.882
33	0.952	0.969	0.929	0.957	0.920	0.983	0.882
35	0.971	0.969	0.939	0.957	0.932	0.983	0.882
37	0.971	0.969	0.949	0.957	0.932	0.983	0.882
38	0.981	0.969	0.949	0.957	0.932	0.983	0.895
39	0.981	0.969	0.949	0.957	0.943	0.983	0.908
40	0.981	0.969	0.960	0.968	0.955	0.983	0.921
41	0.981	0.969	0.960	0.978	0.955	0.983	0.921
42	0.981	0.979	0.960	0.978	0.966	0.983	0.921
43	0.990	0.979	0.970	0.978	0.966	0.983	0.934
44	0.990	0.979	0.970	0.978	0.966	0.983	0.947
45	0.990	0.979	0.970	0.978	0.966	0.983	0.974
49	0.990	0.979	0.980	0.978	0.977	0.983	0.974
50	1.000	0.979	0.980	0.978	0.977	0.983	0.974
54		0.979	0.990	0.978	0.977	0.983	0.974
57		0.979	0.990	0.989	0.977	0.983	0.974
59		0.990	0.990	0.989	0.977	0.983	0.974
60		0.990	0.990	0.989	0.989	0.983	0.974
65		1.000	0.990	0.989	0.989	0.983	0.987
67			0.990	0.989	0.989	1.000	0.987
71			0.990	0.989	0.989		1.000
75			1.000	0.989	0.989		
98				1.000	0.989		
119					1.000		

Other Ward

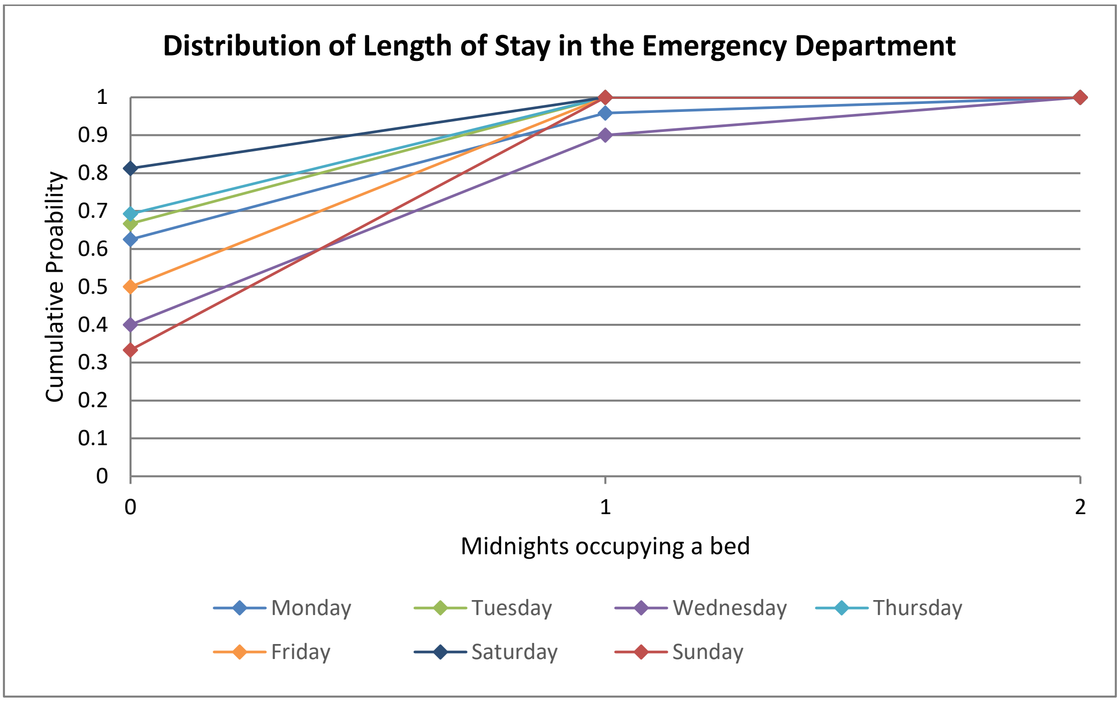


	Cumulative Distribution of Length-of-Stay						
Midnights	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
0	0.594	0.535	0.635	0.501	0.590	0.416	0.345
1	0.781	0.729	0.775	0.693	0.705	0.579	0.585
2	0.836	0.817	0.851	0.756	0.745	0.674	0.750
3	0.887	0.871	0.873	0.783	0.808	0.787	0.850
4	0.917	0.884	0.884	0.821	0.846	0.855	0.890
5	0.922	0.895	0.914	0.854	0.871	0.891	0.920
6	0.924	0.912	0.918	0.882	0.885	0.914	0.930
7	0.931	0.938	0.930	0.887	0.911	0.919	0.930
8	0.942	0.944	0.938	0.914	0.917	0.919	0.950
9	0.947	0.959	0.954	0.914	0.917	0.932	0.965
10	0.952	0.972	0.954	0.919	0.923	0.937	0.980
11	0.952	0.974	0.954	0.929	0.925	0.946	0.980
12	0.956	0.974	0.958	0.947	0.931	0.950	0.985
13	0.959	0.974	0.962	0.960	0.937	0.955	0.985
14	0.959	0.976	0.970	0.970	0.952	0.955	0.985
15	0.959	0.978	0.974	0.972	0.952	0.959	0.990
16	0.961	0.978	0.976	0.972	0.952	0.968	0.990
17	0.968	0.981	0.976	0.975	0.956	0.977	0.990
18	0.972	0.981	0.976	0.977	0.956	0.977	0.995
19	0.972	0.981	0.976	0.980	0.956	0.982	0.995
20	0.972	0.985	0.976	0.982	0.960	0.982	0.995
21	0.975	0.987	0.978	0.985	0.970	0.982	0.995
22	0.977	0.987	0.978	0.987	0.970	0.986	0.995
23	0.979	0.987	0.980	0.987	0.970	0.986	0.995
24	0.984	0.987	0.980	0.987	0.970	0.986	1.000
25	0.984	0.987	0.980	0.987	0.976	0.986	
26	0.984	0.987	0.980	0.987	0.978	0.986	
27	0.984	0.991	0.982	0.990	0.986	0.986	
28	0.986	0.991	0.984	0.990	0.986	0.986	

	Cumulative Distribution of Length-of-Stay						
Midnights	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
29	0.988	0.991	0.984	0.990	0.986	0.986	
30	0.988	0.991	0.988	0.990	0.986	0.986	
31	0.988	0.994	0.988	0.990	0.986	0.991	
32	0.993	0.994	0.988	0.990	0.986	0.991	
33	0.993	0.994	0.990	0.992	0.988	0.991	
35	0.993	0.994	0.990	0.995	0.988	0.991	
36	0.995	0.996	0.994	0.997	0.988	0.991	
38	0.995	0.996	0.994	0.997	0.988	0.995	
39	0.995	0.996	0.994	1.000	0.990	0.995	
40	0.995	0.996	0.994		0.994	0.995	
41	0.995	0.996	0.994		0.996	0.995	
42	0.998	0.996	0.994		0.996	0.995	
43	0.998	0.996	0.996		0.996	0.995	
46	0.998	0.996	0.998		0.996	1.000	
52	0.998	0.996	0.998		0.998		
53	1.000	0.996	0.998		0.998		
74		0.996	0.998		1.000		
83		0.996	1.000				
84		0.998					
157		1.000					

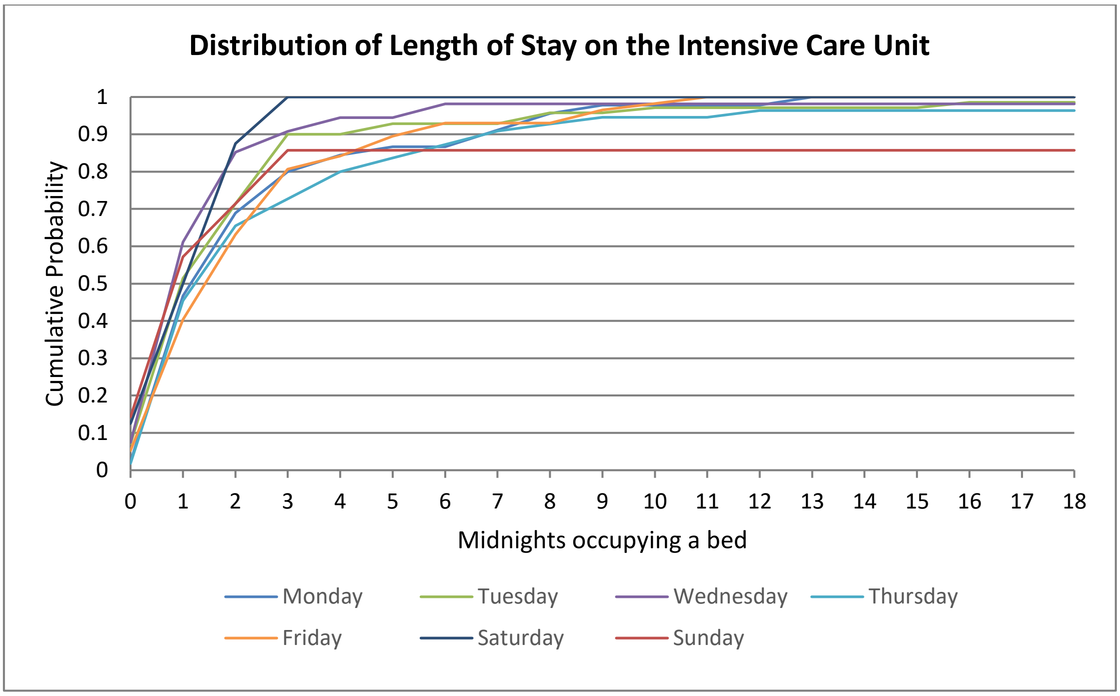
E.2 Length of Stay Distributions for the Elective Patients

Emergency Department



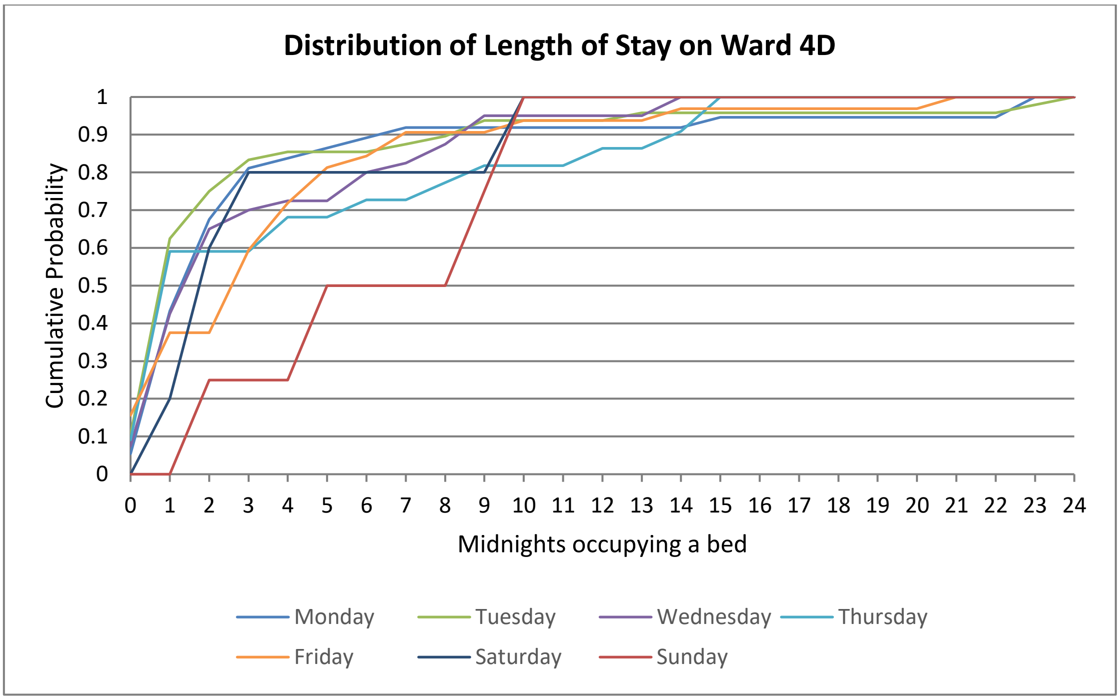
	Cumulative Distribution of Length-of-Stay						
Midnights	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
0	0.625	0.667	0.400	0.692	0.500	0.813	0.333
1	0.958	1.000	0.900	1.000	1.000	1.000	1.000
2	1.000		1.000				

Intensive Care Unit



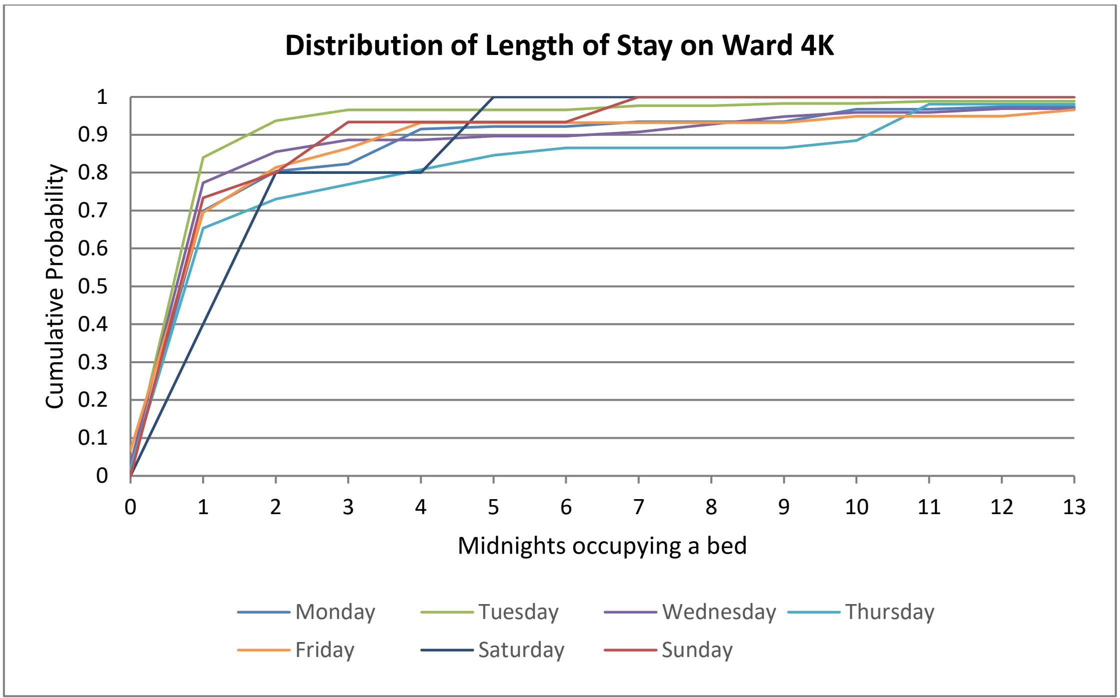
	Cumulative Distribution of Length-of-Stay						
Midnights	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
0	0.022	0.071	0.074	0.018	0.053	0.125	0.143
1	0.467	0.514	0.611	0.455	0.404	0.500	0.571
2	0.689	0.714	0.852	0.655	0.632	0.875	0.714
3	0.800	0.900	0.907	0.727	0.807	1.000	0.857
4	0.844	0.900	0.944	0.800	0.842		0.857
5	0.867	0.929	0.944	0.836	0.895		0.857
6	0.867	0.929	0.981	0.873	0.930		0.857
7	0.911	0.929	0.981	0.909	0.930		0.857
8	0.956	0.957	0.981	0.927	0.930		0.857
9	0.978	0.957	0.981	0.945	0.965		0.857
10	0.978	0.971	0.981	0.945	0.982		0.857
11	0.978	0.971	0.981	0.945	1.000		0.857
12	0.978	0.971	0.981	0.964			0.857
13	1.000	0.971	0.981	0.964			0.857
16		0.986	0.981	0.964			0.857
19		0.986	1.000	0.964			1.000
33		1.000		0.964			
48				0.982			
58				1.000			

Ward 4D



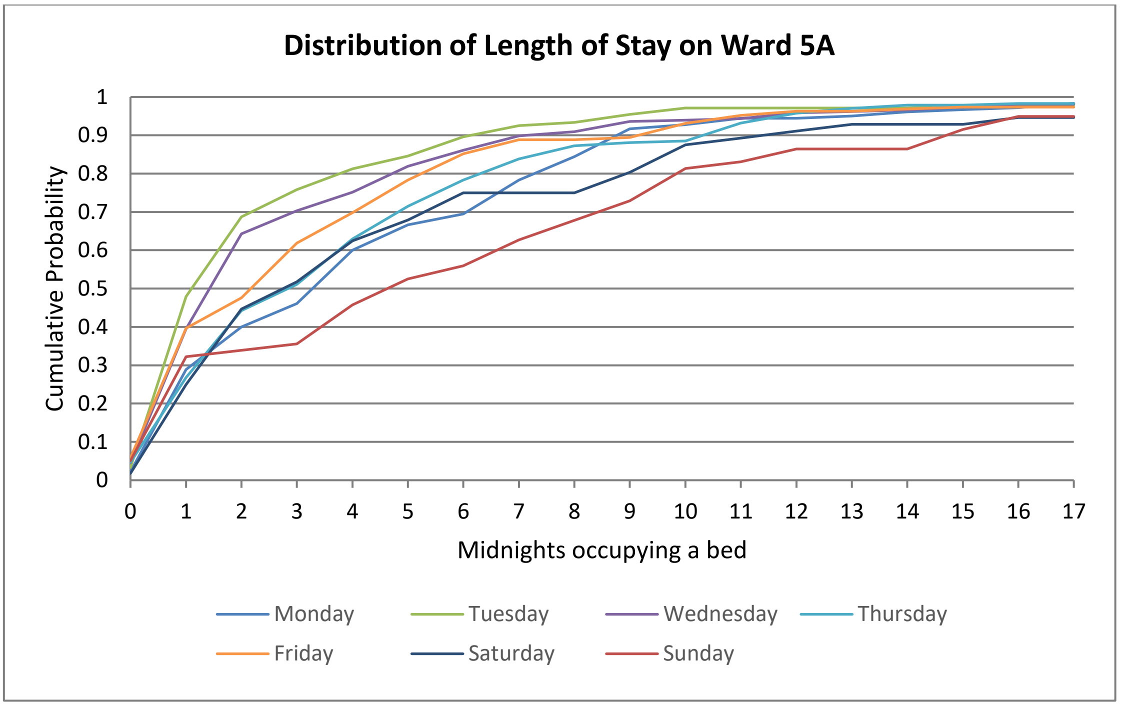
	Cumulative Distribution of Length-of-Stay						
Midnights	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
0	0.054	0.104	0.075	0.091	0.156	0.000	0.000
1	0.432	0.625	0.425	0.591	0.375	0.200	0.000
2	0.676	0.750	0.650	0.591	0.375	0.600	0.250
3	0.811	0.833	0.700	0.591	0.594	0.800	0.250
4	0.838	0.854	0.725	0.682	0.719	0.800	0.250
5	0.865	0.854	0.725	0.682	0.813	0.800	0.500
6	0.892	0.854	0.800	0.727	0.844	0.800	0.500
7	0.919	0.875	0.825	0.727	0.906	0.800	0.500
8	0.919	0.896	0.875	0.773	0.906	0.800	0.500
9	0.919	0.938	0.950	0.818	0.906	0.800	0.750
10	0.919	0.938	0.950	0.818	0.938	1.000	1.000
12	0.919	0.938	0.950	0.864	0.938		
13	0.919	0.958	0.950	0.864	0.938		
14	0.919	0.958	1.000	0.909	0.969		
15	0.946	0.958		1.000	0.969		
21	0.946	0.958			1.000		
23	1.000	0.979					
24		1.000					

Ward 4K



	Cumulative Distribution of Length-of-Stay						
Midnights	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
0	0.013	0.023	0.031	0.019	0.068	0.000	0.000
1	0.699	0.840	0.773	0.654	0.695	0.400	0.733
2	0.804	0.937	0.856	0.731	0.814	0.800	0.800
3	0.824	0.966	0.887	0.769	0.864	0.800	0.933
4	0.915	0.966	0.887	0.808	0.932	0.800	0.933
5	0.922	0.966	0.897	0.846	0.932	1.000	0.933
6	0.922	0.966	0.897	0.865	0.932		0.933
7	0.935	0.977	0.907	0.865	0.932		1.000
8	0.935	0.977	0.928	0.865	0.932		
9	0.935	0.983	0.948	0.865	0.932		
10	0.967	0.983	0.959	0.885	0.949		
11	0.967	0.989	0.959	0.981	0.949		
12	0.974	0.989	0.969	0.981	0.949		
13	0.974	0.989	0.969	0.981	0.966		
14	0.987	0.994	0.979	0.981	0.966		
15	0.993	0.994	0.979	0.981	0.966		
16	0.993	0.994	0.990	0.981	0.966		
18	0.993	0.994	0.990	0.981	0.966		
19	0.993	0.994	0.990	0.981	0.983		
21	0.993	0.994	0.990	0.981	1.000		
25	1.000	0.994	0.990	0.981			
33		0.994	0.990	1.000			
63		1.000	0.990				
72			1.000				

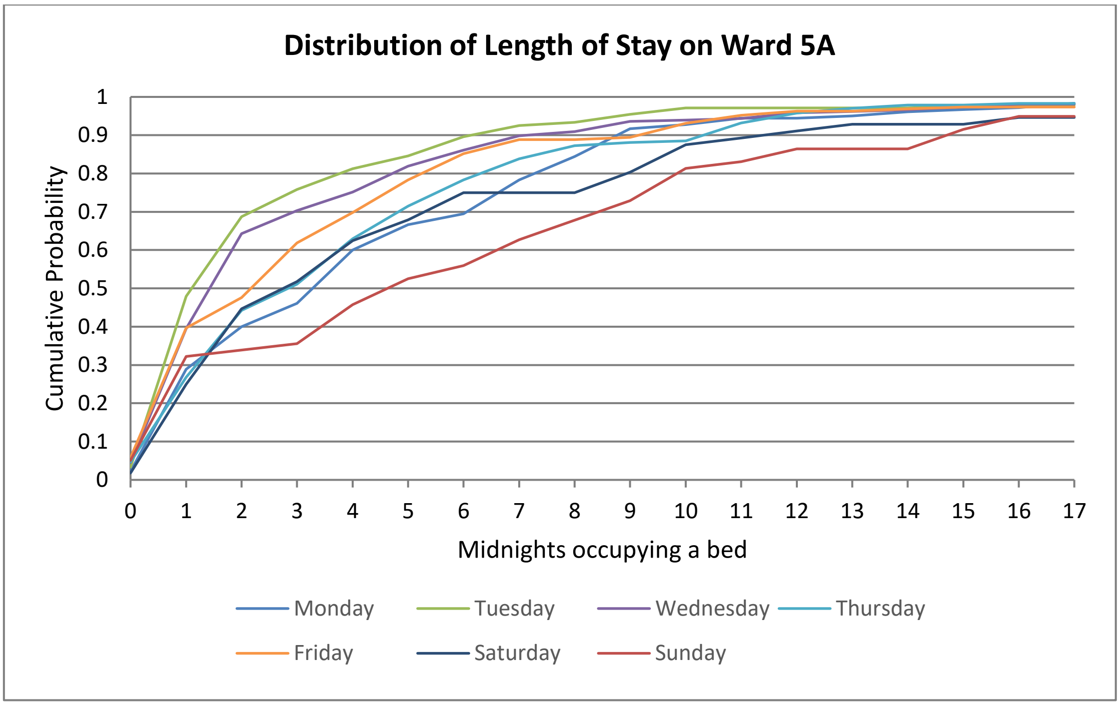
Ward 5A



	Cumulative Distribution of Length-of-Stay						
Midnights	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
0	0.022	0.033	0.045	0.047	0.058	0.018	0.051
1	0.289	0.479	0.395	0.268	0.397	0.250	0.322
2	0.400	0.688	0.643	0.443	0.476	0.446	0.339
3	0.461	0.758	0.703	0.511	0.619	0.518	0.356
4	0.600	0.813	0.752	0.630	0.698	0.625	0.458
5	0.667	0.846	0.820	0.715	0.783	0.679	0.525
6	0.694	0.896	0.861	0.783	0.852	0.750	0.559
7	0.783	0.925	0.898	0.838	0.889	0.750	0.627
8	0.844	0.933	0.910	0.872	0.889	0.750	0.678
9	0.917	0.954	0.936	0.881	0.894	0.804	0.729
10	0.928	0.971	0.940	0.885	0.931	0.875	0.814
11	0.944	0.971	0.944	0.932	0.952	0.893	0.831
12	0.944	0.971	0.959	0.957	0.963	0.911	0.864
13	0.950	0.971	0.962	0.970	0.963	0.929	0.864
14	0.961	0.971	0.966	0.979	0.968	0.929	0.864
15	0.967	0.975	0.974	0.979	0.974	0.929	0.915
16	0.972	0.975	0.981	0.983	0.974	0.946	0.949
17	0.983	0.979	0.981	0.983	0.974	0.946	0.949
18	0.983	0.979	0.981	0.991	0.974	0.964	0.966
19	0.983	0.979	0.992	0.991	0.974	0.964	0.966
20	0.983	0.979	0.992	0.991	0.979	0.982	0.966
21	0.983	0.983	0.996	0.991	0.984	0.982	0.966
22	0.983	0.983	0.996	0.991	0.984	0.982	0.983
24	0.983	0.988	1.000	0.991	0.984	0.982	0.983
25	0.983	0.988		0.991	0.989	0.982	1.000
26	0.983	0.988		0.996	0.989	0.982	
27	0.989	0.988		0.996	0.995	0.982	
28	0.994	0.988		0.996	1.000	0.982	
29	0.994	0.992		0.996		0.982	

	Cumulative Distribution of Length-of-Stay						
Midnights	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
30	1.000	0.996		0.996		1.000	
31		0.996		1.000			
73		1.000					

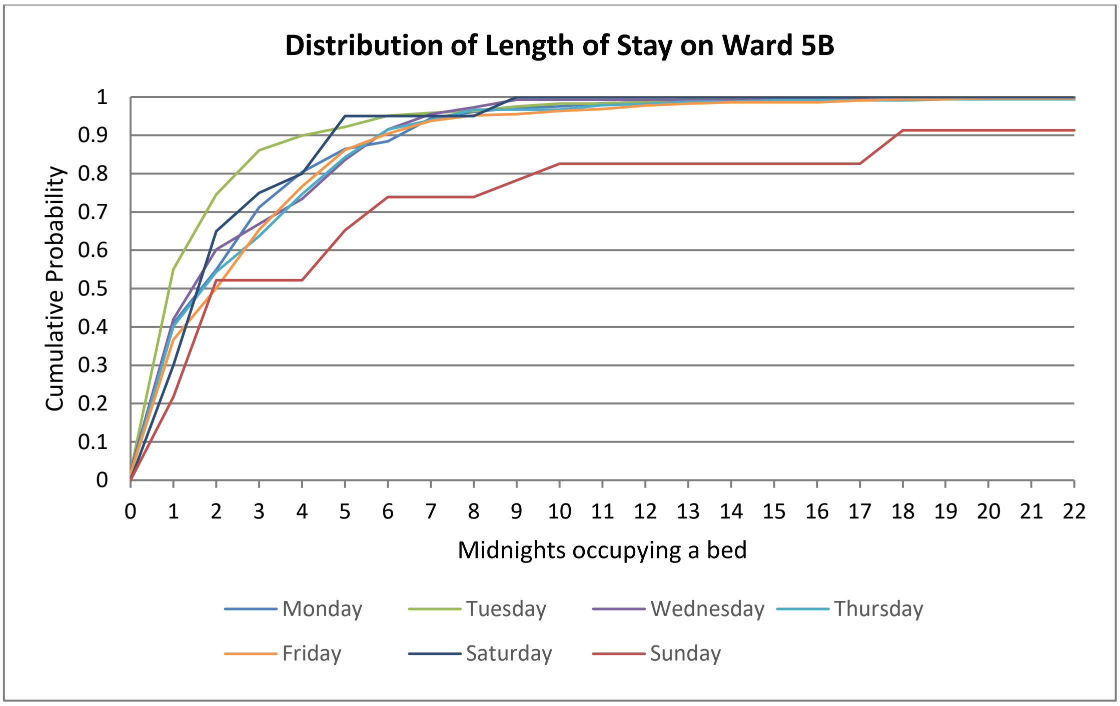
Ward 5A



	Cumulative Distribution of Length-of-Stay						
Midnights	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
0	0.022	0.033	0.045	0.047	0.058	0.018	0.051
1	0.289	0.479	0.395	0.268	0.397	0.250	0.322
2	0.400	0.688	0.643	0.443	0.476	0.446	0.339
3	0.461	0.758	0.703	0.511	0.619	0.518	0.356
4	0.600	0.813	0.752	0.630	0.698	0.625	0.458
5	0.667	0.846	0.820	0.715	0.783	0.679	0.525
6	0.694	0.896	0.861	0.783	0.852	0.750	0.559
7	0.783	0.925	0.898	0.838	0.889	0.750	0.627
8	0.844	0.933	0.910	0.872	0.889	0.750	0.678
9	0.917	0.954	0.936	0.881	0.894	0.804	0.729
10	0.928	0.971	0.940	0.885	0.931	0.875	0.814
11	0.944	0.971	0.944	0.932	0.952	0.893	0.831
12	0.944	0.971	0.959	0.957	0.963	0.911	0.864
13	0.950	0.971	0.962	0.970	0.963	0.929	0.864
14	0.961	0.971	0.966	0.979	0.968	0.929	0.864
15	0.967	0.975	0.974	0.979	0.974	0.929	0.915
16	0.972	0.975	0.981	0.983	0.974	0.946	0.949
17	0.983	0.979	0.981	0.983	0.974	0.946	0.949
18	0.983	0.979	0.981	0.991	0.974	0.964	0.966
19	0.983	0.979	0.992	0.991	0.974	0.964	0.966
20	0.983	0.979	0.992	0.991	0.979	0.982	0.966
21	0.983	0.983	0.996	0.991	0.984	0.982	0.966
22	0.983	0.983	0.996	0.991	0.984	0.982	0.983
24	0.983	0.988	1.000	0.991	0.984	0.982	0.983
25	0.983	0.988		0.991	0.989	0.982	1.000
26	0.983	0.988		0.996	0.989	0.982	
27	0.989	0.988		0.996	0.995	0.982	
28	0.994	0.988		0.996	1.000	0.982	
29	0.994	0.992		0.996		0.982	

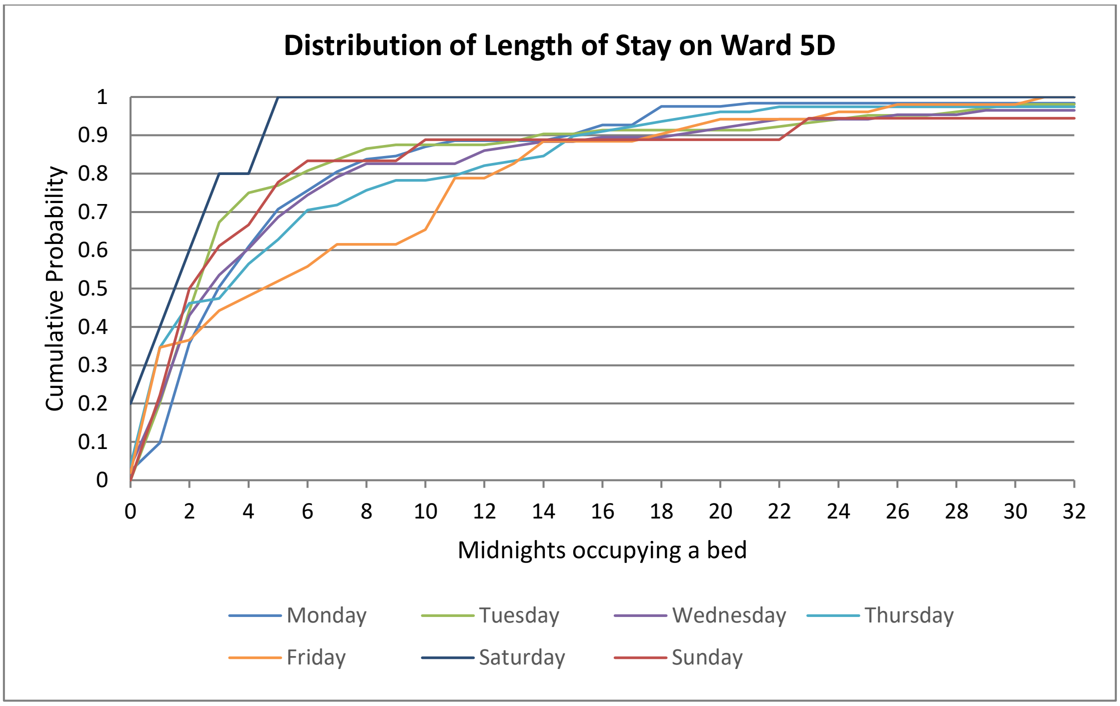
	Cumulative Distribution of Length-of-Stay						
Midnights	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
30	1.000	0.996		0.996		1.000	
31		0.996		1.000			
73		1.000					

Ward 5B



	Cumulative Distribution of Length-of-Stay						
Midnights	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
0	0.019	0.022	0.022	0.015	0.017	0.000	0.000
1	0.409	0.550	0.420	0.402	0.366	0.300	0.217
2	0.550	0.746	0.602	0.544	0.501	0.650	0.522
3	0.713	0.861	0.669	0.637	0.654	0.750	0.522
4	0.804	0.900	0.734	0.746	0.766	0.800	0.522
5	0.865	0.922	0.836	0.843	0.862	0.950	0.652
6	0.884	0.951	0.915	0.915	0.904	0.950	0.739
7	0.945	0.958	0.955	0.940	0.938	0.950	0.739
8	0.961	0.963	0.973	0.967	0.952	0.950	0.739
9	0.972	0.976	0.993	0.967	0.955	1.000	0.783
10	0.975	0.983	0.993	0.967	0.963		0.826
11	0.983	0.983	0.993	0.979	0.969		0.826
12	0.986	0.988	0.993	0.982	0.977		0.826
13	0.986	0.995	0.993	0.988	0.983		0.826
14	0.989	0.995	0.993	0.988	0.986		0.826
15	0.989	0.995	0.995	0.991	0.986		0.826
16	0.992	0.995	0.998	0.991	0.986		0.826
17	0.992	0.995	0.998	0.991	0.992		0.826
18	0.994	0.995	0.998	0.991	0.994		0.913
19	0.994	0.995	0.998	0.994	0.994		0.913
20	0.994	0.995	0.998	0.994	0.997		0.913
22	0.997	0.995	1.000	0.994	0.997		0.913
23	1.000	0.995		0.994	0.997		0.957
25		0.995		0.994	0.997		1.000
26		0.995		0.997	0.997		
29		0.998		1.000	0.997		
32		0.998			1.000		
71		1.000					

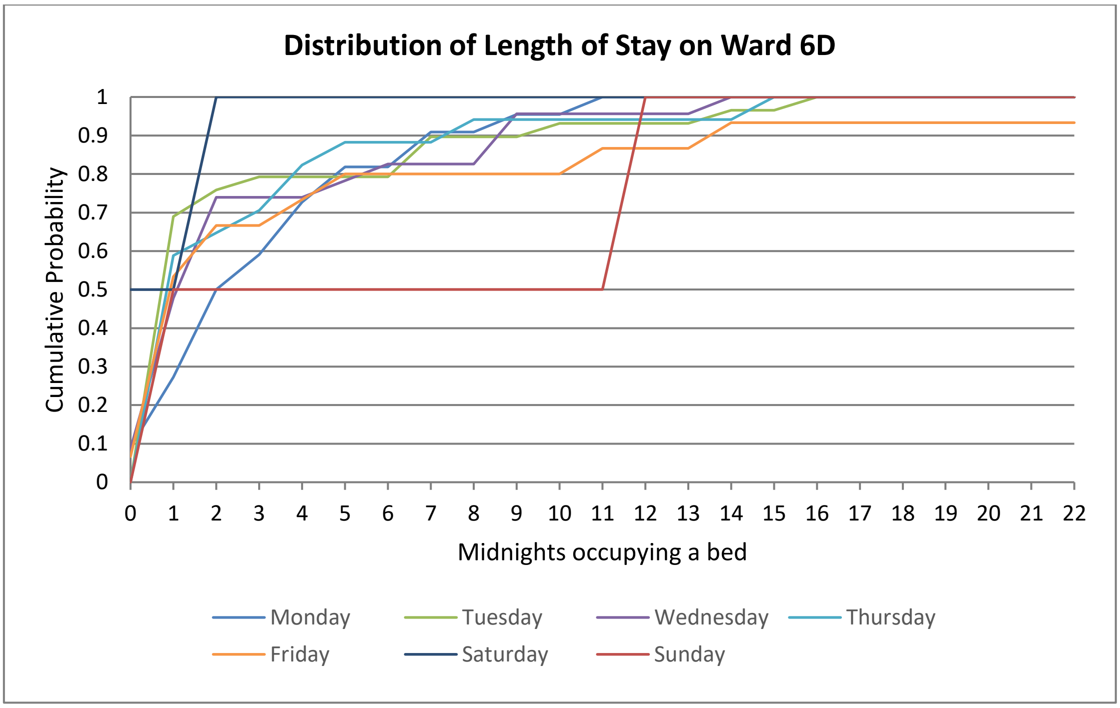
Ward 5D



	Cumulative Distribution of Length-of-Stay						
Midnights	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
0	0.024	0.000	0.035	0.038	0.019	0.200	0.000
1	0.098	0.202	0.209	0.346	0.346	0.400	0.222
2	0.358	0.442	0.430	0.462	0.365	0.600	0.500
3	0.504	0.673	0.535	0.474	0.442	0.800	0.611
4	0.610	0.750	0.605	0.564	0.481	0.800	0.667
5	0.707	0.769	0.686	0.628	0.519	1.000	0.778
6	0.756	0.808	0.744	0.705	0.558		0.833
7	0.805	0.837	0.791	0.718	0.615		0.833
8	0.837	0.865	0.826	0.756	0.615		0.833
9	0.846	0.875	0.826	0.782	0.615		0.833
10	0.870	0.875	0.826	0.782	0.654		0.889
11	0.886	0.875	0.826	0.795	0.788		0.889
12	0.886	0.875	0.860	0.821	0.788		0.889
13	0.886	0.885	0.872	0.833	0.827		0.889
14	0.886	0.904	0.884	0.846	0.885		0.889
15	0.902	0.904	0.884	0.897	0.885		0.889
16	0.927	0.913	0.895	0.910	0.885		0.889
17	0.927	0.913	0.895	0.923	0.885		0.889
18	0.976	0.913	0.895	0.936	0.904		0.889
19	0.976	0.913	0.907	0.949	0.923		0.889
20	0.976	0.913	0.919	0.962	0.942		0.889
21	0.984	0.913	0.930	0.962	0.942		0.889
22	0.984	0.923	0.942	0.974	0.942		0.889
23	0.984	0.933	0.942	0.974	0.942		0.944
24	0.984	0.942	0.942	0.974	0.962		0.944
25	0.984	0.952	0.942	0.974	0.962		0.944
26	0.984	0.952	0.953	0.974	0.981		0.944
28	0.984	0.962	0.953	0.974	0.981		0.944
29	0.984	0.971	0.965	0.974	0.981		0.944

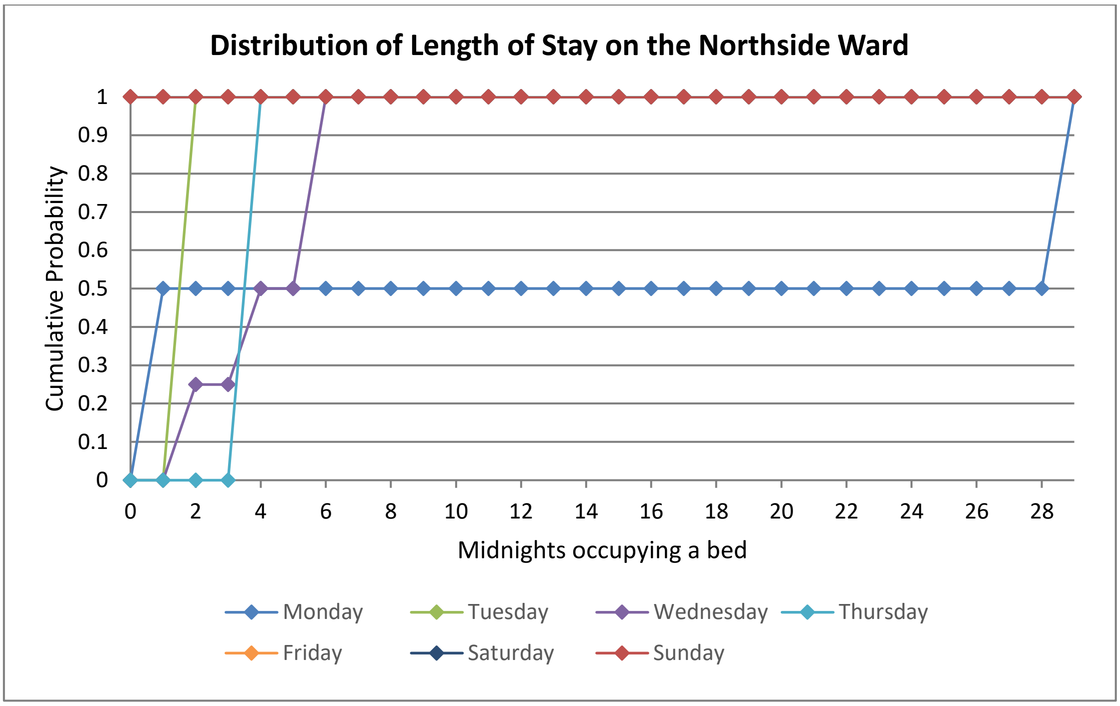
	Cumulative Distribution of Length-of-Stay						
Midnights	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
30	0.984	0.981	0.965	0.974	0.981		0.944
31	0.984	0.981	0.965	0.974	1.000		0.944
33	0.984	0.981	0.965	0.987			1.000
34	0.984	0.990	0.977	0.987			
35	1.000	0.990	0.988	1.000			
41		0.990	1.000				
63		1.000					

Ward 6D



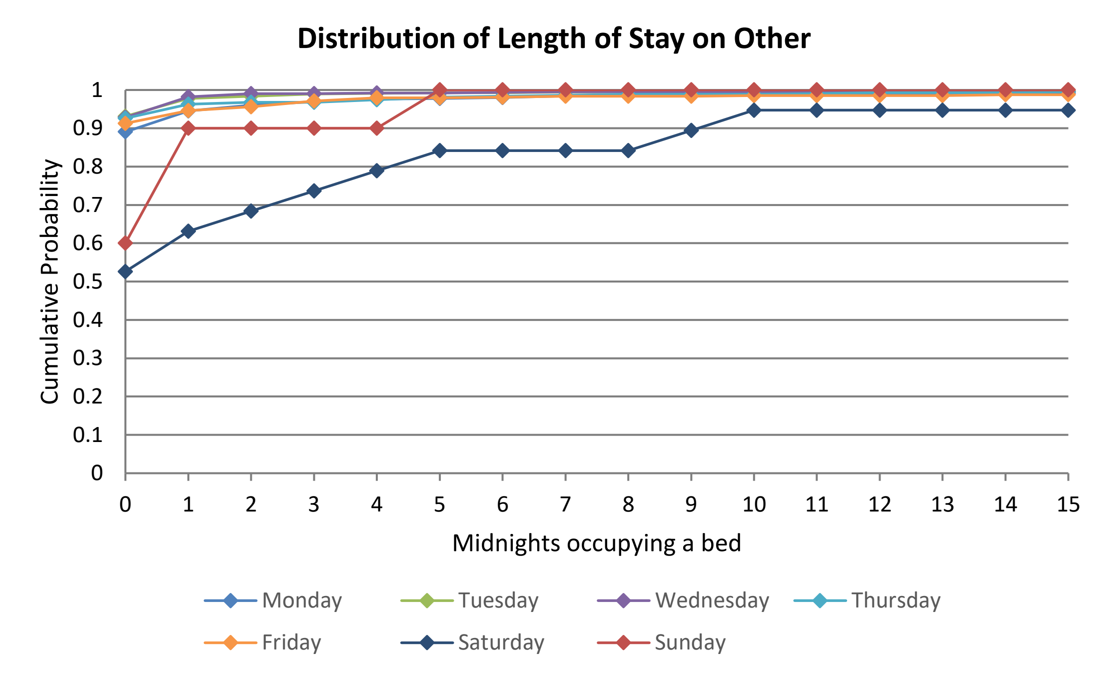
	Cumulative Distribution of Length-of-Stay						
Midnights	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
0	0.091	0.000	0.087	0.000	0.067	0.500	0.000
1	0.273	0.690	0.478	0.588	0.533	0.500	0.500
2	0.500	0.759	0.739	0.647	0.667	1.000	0.500
3	0.591	0.793	0.739	0.706	0.667		0.500
4	0.727	0.793	0.739	0.824	0.733		0.500
5	0.818	0.793	0.783	0.882	0.800		0.500
6	0.818	0.793	0.826	0.882	0.800		0.500
7	0.909	0.897	0.826	0.882	0.800		0.500
8	0.909	0.897	0.826	0.941	0.800		0.500
9	0.955	0.897	0.957	0.941	0.800		0.500
10	0.955	0.931	0.957	0.941	0.800		0.500
11	1.000	0.931	0.957	0.941	0.867		0.500
12		0.931	0.957	0.941	0.867		1.000
14		0.966	1.000	0.941	0.933		
15		0.966		1.000	0.933		
16		1.000			0.933		
180					1.000		

Northside Ward



	Cumulative Distribution of Length-of-Stay						
Midnights	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
0	0.000	0.000	0.000	0.000	1.000	1.000	1.000
1	0.500	0.000	0.000	0.000			
2	0.500	1.000	0.250	0.000			
4	0.500		0.500	1.000			
6	0.500		1.000				
29	1.000						

Other Ward



	Cumulative Distribution of Length-of-Stay						
Midnights	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
0	0.891	0.931	0.928	0.926	0.913	0.526	0.600
1	0.945	0.978	0.982	0.963	0.946	0.632	0.900
2	0.961	0.984	0.990	0.968	0.956	0.684	0.900
3	0.969	0.990	0.990	0.968	0.971	0.737	0.900
4	0.976	0.991	0.992	0.975	0.979	0.789	0.900
5	0.978	0.993	0.992	0.982	0.979	0.842	1.000
6	0.980	0.993	0.993	0.984	0.981	0.842	
7	0.985	0.994	0.995	0.986	0.983	0.842	
8	0.987	0.996	0.995	0.988	0.983	0.842	
9	0.987	0.996	0.995	0.988	0.983	0.895	
10	0.987	0.996	0.995	0.988	0.985	0.947	
11	0.987	0.996	0.995	0.991	0.985	0.947	
12	0.987	0.996	0.995	0.993	0.985	0.947	
14	0.989	0.996	0.995	0.993	0.988	0.947	
15	0.991	0.997	0.995	0.995	0.988	0.947	
16	0.993	0.999	0.995	0.995	0.988	1.000	
17	0.993	1.000	0.995	0.995	0.988		
18	0.993		0.995	0.995	0.990		
19	0.993		0.995	0.998	0.990		
21	0.993		0.995	0.998	0.994		
23	0.993		0.997	0.998	0.994		
29	0.996		0.997	0.998	0.994		
30	0.996		0.998	0.998	0.994		
31	0.996		0.998	0.998	0.996		
40	0.996		0.998	0.998	0.998		
41	0.996		0.998	1.000	0.998		
43	0.998		0.998		0.998		
46	0.998		0.998		1.000		
47	0.998		1.000				

	Cumulative Distribution of Length-of-Stay						
Midnights	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
92	1.000						

References

A

Adan, I., Bekkers, J., Dellaert, N., Vissers, J. and Yu, X. (2009) 'Patient mix optimisation and stochastic resource requirements: A case study in cardiothoracic surgery planning', *Health Care Management Science,* 12(2), pp. 129-141.

Adan, I. and Vissers, J. (2002) 'Patient mix optimisation in hospital admission planning: a case study', *International journal of operations & production management,* 22(4), pp. 445-461.

Audit Commission (2003) *Bed Management: Review of National Findings* (ISBN 186240 442 9. Available at: http://www.audit-commission.gov.uk/nationalstudies/health/other/Pages/bedmanagement.aspx#downloads (Accessed: 03/05/2012).

Aydt, H., Turner, S. J., Cai, W., Low, M. Y. H., Lendermann, P., Gan, B. P. and Ayani, R. (2008) 'Preventive what-if analysis in symbiotic simulation', *Proceedings of the 40th Conference on Winter Simulation*, Miami, Florida. 1516883: Winter Simulation Conference, 750-758.

B

Bagust, A., Place, M. and Posnett, J. (1999) 'Dynamics of bed use in accommodating emergency admissions: stochastic simulation model', *BMJ: British Medical Journal (International Edition),* 319(7203), pp. 155-158.

Bahrani, S., Tchemeube, R. B., Mouttham, A. and Amyot, D. (2013) 'Real-time simulations to support operational decision making in healthcare', *Proceedings of the 2013 Summer Computer Simulation Conference*, Toronto, Ontario, Canada. 2557754: Society for Modeling & Simulation International, 1-7.

Barnoon, S. and Wolfe, H. (1968) 'Scheduling a multiple operating room system: a simulation approach', *Health Services Research,* 3(4), pp. 272.

Beliën, J. and Demeulemeester, E. (2007) 'Building cyclic master surgery schedules with leveled resulting bed occupancy', *European Journal of Operational Research,* 176(2), pp. 1185-1204.

Beliën, J., Demeulemeester, E. and Cardoen, B. (2008) 'A decision support system for cyclic master surgery scheduling with multiple objectives', *Journal of Scheduling,* 12(2), pp. 147.

Broyles, J. R., Cochran, J. K. and Montgomery, D. C. (2010) 'A statistical Markov chain approximation of transient hospital inpatient inventory', *European Journal of Operational Research,* 207(3), pp. 1645-1657.

C

Chow, V. S., Puterman, M. L., Salehirad, N., Huang, W. and Atkins, D. (2011) 'Reducing Surgical Ward Congestion Through Improved Surgical Scheduling and Uncapacitated Simulation', *Production and Operations Management,* 20(3), pp. 418-430.

Connors, M. M. (1970) 'A stochastic elective admissions scheduling algorithm', *Health Services Research,* 5(4), pp. 308.

D

Davis, W. J. (1998) 'On-Line Simulation: Need and Evolving Research Requirements', *Handbook of Simulation: Principles, Methodology, Advances, Applications, and Practice*, pp. 465-516.

de Bruin, A. M., Bekker, R., Van Zanten, L. and Koole, G. (2010) 'Dimensioning hospital wards using the Erlang loss model', *Annals of Operations Research,* 178(1), pp. 23-43.

Devroye, L. (1986) 'Sample-based non-uniform random variate generation'. *Proceedings of the 18th conference on Winter simulation*: ACM, 260-265.

E

El-Darzi, E., Vasilakis, C., Chaussalet, T. and Millard, P. H. (1998) 'A simulation modelling approach to evaluating length of stay, occupancy, emptiness and bed blocking in a hospital geriatric department', *Health Care Management Science,* 1(2), pp. 143-149.

England, W. and Roberts, S. D. (1978) 'Applications of computer simulation in health care', *Proceedings of the 10th conference on Winter simulation - Volume 2*, Miami Beach, FL. 807598: IEEE Computer Society Press, 665-677.

Esogbue, A. O. and Singh, A. J. (1976) 'A stochastic model for an optimal priority bed distribution problem in a hospital ward', *Operations Research*, pp. 884-898.

Espinoza, C., Pascual, J., Ramis, F., Bórquez, D. and Sepúlveda, J. A. (2014) 'Real-time simulation as a way to improve daily operations in an emergency room', *Proceedings of the 2014 Winter Simulation Conference*, Savannah, Georgia. 2694033: IEEE Press, 1445-1456.

Everett, J. E. (2002) 'A Decision Support Simulation Model for the Management of an Elective Surgery Waiting System', *Health Care Management Science,* 5(2), pp. 89-95.

F

Fetter, R. B. and Thompson, J. D. (1965) 'THE SIMULATION OF HOSPITAL SYSTEMS', *Operations Research,* 13(5), pp. 689.

Fetter, R. B. and Thompson, J. D. (1969) 'A decision model for the design and operation of a progressive patient care hospital', *Medical Care*, pp. 450-462.

Fletcher, A. and Worthington, D. (2009) 'What is a ‘generic’ hospital model?—a comparison of ‘generic’ and ‘specific’ hospital models of emergency patient flows', *Health Care Management Science,* 12(4), pp. 374-391.

Fone, D., Hollinghurst, S., Temple, M., Round, A., Lester, N., Weightman, A., Roberts, K., Coyle, E., Bevan, G. and Palmer, S. (2003) 'Systematic review of the use and value of computer simulation modelling in population health and health care delivery', *Journal of Public Health,* 25(4), pp. 325-335.

G

Gallivan, S. and Utley, M. (2005) 'Modelling admissions booking of elective in-patients into a treatment centre', *IMA Journal of Management Mathematics,* 16(3), pp. 305-315.

Gemmel, P. and Van Dierdonck, R. (1999) 'Admission scheduling in acute care hospitals: does the practice fit with the theory?', *International journal of operations & production management,* 19(9), pp. 863-878.

Goldman, J., Knappenberger, H. A. and Eller, J. (1968) 'Evaluating bed allocation policy with computer simulation', *Health Services Research,* 3(2), pp. 119.

Groot, P. M. A. (1993) *Decision support for admission planning under multiple resource constraints.* Doctoral degree, Eindhoven University of Tehnology, Eindhoven.

Gupta, A. K., Sivakumar, A. I. and Sarawgi, S. (2002) 'Scheduling & control: shop floor scheduling with simulation based proactive decision support'. *Proceedings of the 34th conference on Winter simulation: exploring new frontiers*: Winter Simulation Conference, 1897-1902.

Günal , M. M. (2008) *Simulation modelling for understanding performance in healthcare.* Doctor of Philosophy, Lancaster University, Lancaster.

Günal, M. M. and Pidd, M. (2010) 'Discrete event simulation for performance modelling in health care: a review of the literature', *Journal of Simulation,* 4(1), pp. 42-51.

H

Hanisch, A., Tolujew, J., Richter, K. and Schulze, T. (2003) 'Online simulation of pedestrian flow in public buildings'. *Simulation Conference, 2003. Proceedings of the 2003 Winter*, 7-10 Dec. 2003, 1635-1641 vol.2.

Hanisch, A., Tolujew, J. and Schulze, T. (2005) 'Initialization of online simulation models', *Proceedings of the 37th conference on Winter simulation*, Orlando, Florida. 1163021: Winter Simulation Conference, 1795-1803.

Hans, E. W., Van Houdenhoven, M. and Hulshof, P. J. (2012) 'A framework for healthcare planning and control', *Handbook of healthcare system scheduling*: Springer, pp. 303-320.

Harper, P. and Shahani, A. (2002) 'Modelling for the planning and management of bed capacities in hospitals', *Journal of the operational research society*, pp. 11-18.

Harper, P. R. (2002) 'A Framework for Operational Modelling of Hospital Resources', *Health Care Management Science,* 5(3), pp. 165-173.

Helm, J. E., AhmadBeygi, S. and Van Oyen, M. P. (2009) 'The flexible patient flow simulation framework'. *IIE Annual Conference. Proceedings*: Institute of Industrial and Systems Engineers (IISE), 803.

Helm, J. E., AhmadBeygi, S. and Van Oyen, M. P. (2011) 'Design and analysis of hospital admission control for operational effectiveness', *Production and Operations Management,* 20(3), pp. 359-374.

Helm, J. E. and Van Oyen, M. P. (2014) 'Design and Optimization Methods for Elective Hospital Admissions', *Operations Research,* 62(6), pp. 1265-1282.

Hill, J. M., Surdu, J., Vaglia, J. and Pooch, U. W. (2001) 'Synchronized Simulations in Planning Systems', *SIMULATION SERIES,* 33(4), pp. 69-75.

Hoot, N. R., LeBlanc, L. J., Jones, I., Levin, S. R., Zhou, C., Gadd, C. S. and Aronsky, D. (2008) 'Forecasting emergency department crowding: a discrete event simulation', *Annals of emergency medicine,* 52(2), pp. 116-125.

J

Jun, J. B., Jacobson, S. H. and Swisher, J. R. (1999) 'Application of discrete-event simulation in health care clinics: A survey', *The Journal of the Operational Research Society,* 50(2), pp. 109-123.

K

Klein, R. W., Dittus, R. S., Roberts, S. D. and Wilson, J. R. (1993) 'Simulation modeling and health-care decision making', *Medical decision making,* 13(4), pp. 347-354.

Knuth, D. E. (1998) *The art of computer programming.* 3rd ed. edn. Reading, Mass. ; Harlow: Addison-Wesley, 1997.

Koestler, D. C., Ombao, H. and Bender, J. (2013) 'Ensemble-based methods for forecasting census in hospital units', *BMC Medical Research Methodology,* 13, pp. 67-67.

Kolesar, P. (1970) 'A Markovian model for hospital admission scheduling', *Management Science*, pp. 384-396.

Kusters, R. J. and Groot, P. M. A. (1996) 'Modelling resource availability in general hospitals design and implementation of a decision support model', *European Journal of Operational Research,* 88(3), pp. 428-445.

L

Lakshmi, C. and Iyer, S. A. (2013) 'Application of queueing theory in health care: A literature review', *Operations research for health care,* 2(1), pp. 25-39.

Littig, S. J. and Isken, M. W. (2007) 'Short term hospital occupancy prediction', *Health Care Management Science,* 10(1), pp. 47-66.

Low, M. Y., Turner, S. J., Ling, D., Peng, H. L., Lendermann, P., Chan, L. P. and Buckley, S. (2007) 'Symbiotic simulation for business process re-engineering in high-tech manufacturing and service networks'. *Simulation Conference, 2007 Winter*: IEEE, 568-576.

M

Magerlein, J. M. and Martin, J. B. (1978) 'Surgical demand scheduling: a review', *Health Services Research,* 13(4), pp. 418.

Marmor, Y. N., Wasserkrug, S., Zeltyn, S., Mesika, Y., Greenshpan, O., Carmeli, B., Shtub, A. and Mandelbaum, A. (2009) 'Toward simulation-based real-time decision-support systems for emergency departments'. *Simulation Conference (WSC), Proceedings of the 2009 Winter*: IEEE, 2042-2053.

Mazur, F., Chrobok, R., Hafstein, S. F., Pottmeier, A. and Schreckenberg, M. (2004) 'Future of Traffic Information: Online-Simulation of a Large Scale Freeway Network'. *ICWI*, 665-672.

McClean, S. and Millard, P. H. (1995) 'A decision support system for bed-occupancy management and planning hospitals', *Mathematical Medicine and Biology,* 12(3-4), pp. 249-257.

Mechanic, D., McAlpine, D. D. and Olfson, M. (1998) 'Changing patterns of psychiatric inpatient care in the united states, 1988-1994', *Archives of General Psychiatry,* 55(9), pp. 785-791.

Milsum, J. H., Turban, E. and Vertinsky, I. (1973) 'Hospital admission systems: Their evaluation and management', *Management Science,* 19(6), pp. 646-646.

Monks, T., Currie, C. S. M., Onggo, B. S., Robinson, S., Kunc, M. and Taylor, S. J. E. (2018) 'Strengthening the reporting of empirical simulation studies: Introducing the STRESS guidelines', *Journal of Simulation*, pp. 1-13.

Monks, T., Worthington, D., Allen, M., Pitt, M., Stein, K. and James, M. A. (2016) 'A modelling tool for capacity planning in acute and community stroke services', *BMC health services research,* 16(1), pp. 530.

Mousavi, A., Komashie, A. and Tavakoli, S. (2011) 'Simulation-based real-time performance monitoring (simmon): A platform for manufacturing and healthcare systems'. *Simulation Conference (WSC), Proceedings of the 2011 Winter*: IEEE, 600-611.

N

Nance, R. E. (1996) 'A history of discrete event simulation programming languages', in Thomas J. Bergin, Jr. & Richard G. Gibson, Jr. (eds.) *History of programming languages---II*: ACM, pp. 369-427.

NHS England (1988-2017) 'Statistics; Bed Availability and Occupancy Data - Overnight'. Available at: https://www.england.nhs.uk/statistics/statistical-work-areas/bed-availability-and-occupancy/bed-data-overnight/ (Accessed 22/01/2017).

NHS Improvement (2016a) *NHS providers working hard, but still under pressure | NHS Improvement*. London: NHS Improvement. Available at: https://improvement.nhs.uk/news-alerts/nhs-providers-working-hard-still-under-pressure/ (Accessed: 22/01/2017.

NHS Improvement (2016b) *Rapid improvement guide: setting the clinical criteria and expected date of discharge*. Rapid improvement guides for urgent and emergency care. London: NHS Improvement. Available at: https://improvement.nhs.uk/uploads/documents/expected-date-of-discharge-and-clinical-criteria-RIG.pdf (Accessed: 30/08/2017.

O

Ou, L., Chen, J., Young, L., Santiano, N., Baramy, L.-S. and Hillman, K. (2011) 'Effective discharge planning–timely assignment of an estimated date of discharge', *Australian Health Review,* 35(3), pp. 357-363.

P

Pidd, M. (2009) *Tools for Thinking: Modelling in Management Science.* 3rd edn.: Wiley.

Potoradi, J., Boon, O. S., Mason, S. J., Fowler, J. W. and Pfund, M. E. (2002) 'Semiconductor manufacturing: using simulation-based scheduling to maximize demand fulfillment in a semiconductor assembly facility', *Proceedings of the 34th conference on Winter simulation: exploring new frontiers*, San Diego, California. 1030728: Winter Simulation Conference, 1857-1861.

R

Rais, A. and Viana, A. (2011) 'Operations research in healthcare: a survey', *International transactions in operational research,* 18(1), pp. 1-31.

Robinson, G. H., Wing, P. and Davis, L. E. (1968) 'Computer simulation of hospital patient scheduling systems', *Health Services Research,* 3(2), pp. 130.

S

Shtatland, E. S., Kleinman, K. and Cain, E. M. (2003) 'Stepwise methods in using SAS PROC LOGISITIC and SAS ENTERPRISE MINER for prediction'. *Statistics and Data Analysis*. *SAS Users Group International 28*: SAS Institute.

Smith, W. G. and Solomon Jr, M. B. (1966) 'A simulation of hospital admission policy', *Communications of the ACM,* 9(5), pp. 362-365.

T

Tan, K. W., Lau, H. C. and Lee, F. C. Y. (2013) 'Improving patient length-of-stay in emergency department through dynamic queue management'.

V

van Essen, J. T., Bosch, J. M., Hans, E. W., van Houdenhoven, M. and Hurink, J. L. (2012) 'Improve OR-schedule to reduce number of required beds'.

Vanberkel, P. T., Boucherie, R. J., Hans, E. W., Hurink, J. L., van Lent, W. A. M. and van Harten, W. H. (2011) 'An exact approach for relating recovering surgical patient workload to the master surgical schedule', *Journal of the Operational Research Society,* 62(10), pp. 1851-1860.

Y

Yi, P., George, S. K., Paul, J. A. and Lin, L. (2010) 'Hospital capacity planning for disaster emergency management', *Socio-Economic Planning Sciences,* 44(3), pp. 151-160.