Robust Optimisation of Operating Theatre Schedules

Elizabeth Louise Rowse

School of Mathematics

Cardiff University

A thesis submitted for the degree of

Doctor of Philosophy

June 2015

Summary

Hospitals in the UK are increasingly having to cancel a large proportion of elective operations due to the unavailability of beds on hospital wards for post-operative recovery. The availability of post-operative beds is therefore critical to the scheduling of surgical procedures and the throughput of patients in a hospital. The focus of this research is to investigate, via data-driven modelling, systematic reasons for the unavailability of beds and to demonstrate how the Master Surgery Schedule (MSS) can be constructed using Operational Research techniques to minimise the number of cancellations of elective operations.

Statistical analysis of data provided by the University Hospital of Wales, Cardiff was performed, providing information on patient demand and length of stay distributions. A two-stage modelling process was developed to construct and simulate an MSS that minimises the number of cancellations. The first stage involves a novel set partitioning based optimisation model that incorporates operating room and bed constraints. The second stage simulates the resulting optimal schedule to provide measures on how well the schedule would perform if implemented. The results from this two-stage model provide insights into when best to schedule surgical specialties and how best the beds are distributed between wards.

Two optimisation under uncertainty techniques are then employed to incorporate the uncertainty associated with the bed requirements into the optimisation process. A robust optimisation (RO) approach that uses protection functions in each bed constraint is developed. Investigations into varying levels of protection are performed in order to gain insight into the so called 'price of robustness'. Results show that MSSs that are constructed from protecting more of the uncertainty result in fewer cancellations and a smaller probability of requiring more beds than are available.

The deterministic optimisation model is then extended to become a scenario-based optimisation model in which more scenarios of bed requirement are incorporated into a single optimisation model. Results show that as more scenarios are included, a more robust schedule is generated and fewer cancellations are expected.

Results from the different approaches are compared to assess the benefits of using RO techniques. Future research directions following from this work are discussed, including the construction of the MSS based on sub-specialties and investigation of different working practices within the case study hospital.

Declaration

This work has not been submitted in substance for any other degree or award at this or any other university or place of learning, nor is being submitted concurrently in candidature for any degree or other award.

Signed Date

STATEMENT 1

This thesis is being submitted in partial fulfillment of the requirements for the degree of PhD.

Signed Date

STATEMENT 2

This thesis is the result of my own independent work/investigation, except where otherwise stated. Other sources are acknowledged by explicit references. The views expressed are my own.

Signed Date

STATEMENT 3

I hereby give consent for my thesis, if accepted, to be available for photocopying and for inter-library loan, and for the title and summary to be made available to outside organisations.

Signed Date

Acknowledgements

First and foremost, I would like to thank my supervisors, Professor Paul Harper, Dr Rhyd Lewis, Dr Jonathan Thompson and Dr Janet Williams for their encouragement, interest and support throughout this research. Thank you for your guidance and for all of the opportunities you have given me.

I am grateful to Cardiff and Vale University Health Board for funding this research and for the provision of data for this project. I am particularly grateful to Dr Tony Turley and Mr Andrew Nelson for their continued interest in this research, and I would also wish to thank John Meredith for his patience in responding to my many data requests.

I wish to sincerely thank Professor David Ryan, Department of Engineering Science, University of Auckland, for the invaluable discussions that we had at the early stages of the development of the set partitioning model. I still can't quite believe that I travelled half-way around the world, following that coincidental meeting in Lancaster! I will always fondly remember David's and Ruth's generosity and hospitality during my visit to Waiheke Island.

I am also very grateful to Associate Professor Andrew Mason and Dr Mike O'Sullivan for their guidance and useful discussions when visiting the University of Auckland.

Finally, I would like to extend my thanks to all of my family and friends. To Dr Gillard for tempting me to Cardiff at the careers fair all those years ago. To the Maths Massive for our well deserved tea breaks, conference adventures and euphoria in CBUK – it wouldn't have been the same without you! Last, but by no means least, to my parents who have tried to keep me sane and who are always there for a 'chat' to make sure I do a 'proper job'.

Contents

Publications and Presentations

Publications

Rowse, E., Lewis, R., Harper, P.R. and Thompson, J. (2015). Applying Set Partitioning Methods in the Construction of Operating Theatre Schedues. *4th International Conference on Theory and Practice in Modern Computing*, Las Palmas de Gran Canaria, Spain.

Rowse, E., Lewis, R., Harper, P.R. and Thompson, J. (2014). Set Partitioning Methods for Scheduling: an Application to Operating Theatres. *XLVI SBPO Conference of the Brazilian Operational Research Society*, Salvador, Brazil.

Rowse, E. and Harper, P. R. (2014). Big Data Analysis for Resource-Constrained Surgical Scheduling. *SAS Global Forum*; Washington, D.C.

Castillo, A. and Rowse, E. (2014). SCOR14 – PhD Student Conference. *Inside OR*, July 2014, no. 523, p22.

Rowse, E., Harper, P.R., Lewis, R., and Thompson, J. Scenario-Based Optimisation for the Construction of the Master Surgery Schedule [In preparation].

Rowse, E., Harper, P.R., Lewis, R., and Thompson, J. Robust Optimisation of Operating Theatre Schedules [In preparation].

Conference contributions and Presentations

Rowse, E., Harper, P.R., Lewis, R., and Thompson, J. (2014). Set Partitioning Methods for Robust Scheduling: an Application to Operating Theatres, *PATAT*, York.

Rowse, E., Harper, P.R., Lewis, R., and Thompson, J. (2014). Set Partitioning Methods for Robust Scheduling: an Application to Operating Theatres, *ORAHS*, Lisbon, Portugal.

Rowse, E., Harper, P.R., Lewis, R., and Thompson, J. (2014). Set Partitioning Methods for Robust Scheduling: an Application to Operating Theatres, *SCOR 2014*, Nottingham.

Rowse, E., Harper, P.R., Lewis, R., and Thompson, J. (2014). Set Partitioning Methods for Robust Scheduling: an Application to Operating Theatres, *Developing Mathematical Models Applied to Healthcare Seminar*, Cardiff.

Rowse, E., Harper, P.R., Lewis, R., and Thompson, J. (2014). Big Data Analysis for Resource-Constrained Surgical Scheduling, *SAS Global Forum*, Washington, D.C.

Rowse, E., Harper, P.R., Lewis, R., and Thompson, J. (2013). Set Partitioning Methods for Robust Scheduling: an Application to Operating Theatres, *OR55*, Exeter.

Rowse, E., Harper, P.R., Lewis, R., and Thompson, J. (2013). Set Partitioning Methods for Robust Scheduling: an Application to Operating Theatres, *Wales Mathematics Colloquium*, Gregynog.

Research Visits

Department of Industrial Engeineering, Pontifical Catholic University of Rio de Janeiro. Rio de Janeiro, Brazil, April 2015.

Health Services and Outcomes Research (HSOR) unit, National Healthcare Group. Singapore, March 2013.

Department of Engineering Science, University of Auckland. Auckland,
January – March 2013.

List of Tables

List of Figures

List of Abbreviations

CaV UHB	Cardiff and Vale University Health Board
CCU	Critical Care Unit
CEPOD	Confidential Enquiry into Patient Outcome and Death
GUB	Generalised Upper Bound
LoS	Length of Stay
MIP	Mixed Integer Programming
MSS	Master Surgery Schedule
NHS	National Health Service
RO	Robust Optimisation
RTT	Referral to Treatment
SCP	Sampled Convex Program
SNCP	Sampled Non-Convex Program
SP	Stochastic Programming
SPP	Set Partitioning Problem
SSSU	Short Stay Surgical Unit
TAL	Theatre Admissions Lounge
UHW	University Hospital of Wales

List of Notation

Notation	Indicies	Definition
m		Number of rows in A matrix
n		Number of plans (columns) in the A and B matrices
p		Number of wards
q		Number of days
s		Number of specialties
A		Matrix of Generalised Upper Bound constraints
		and operating theatre constraints.
a_{ij}	$i = s+1, \ldots, m$	Indicates if specialty is scheduled in
	$j = 1, \ldots, n$	operating theatre session i in plan j.
B		Matrix of bed constraints
$b_{kj}^{(l)}$	$k = 1, \ldots, p$	Bed requirement in ward k on day l for plan j.
	$l = 1, \ldots, q$	
	$j = 1, \ldots, n$	
w_{kv}	$k = 1, \ldots, p$	Indicates if bed sharing is allowed between
	$v = 1, \ldots, p+1$	wards k and v.
$d_k^{(l)}$	$k = 1, \ldots, p$	Number of beds available on ward k on day l.
	$l = 1, \ldots, q$	
x_j	$j = 1, \ldots, n$	Decision variable: plan j is chosen or not.
$z_{kv}^{(l)}$	$k = 1, \ldots, p$	Decision variable: number of beds transferred
	$v = 1, \ldots, p+1$	between wards k and v on day l.
	$l = 1, \ldots, q$	
J_k	$k = 1, \ldots, p$	Set of uncertain coefficients for ward k
$\tilde{b}_{kj}^{(l)}$	$k = 1, \ldots, p$	Random variable of bed requirement in ward k
	$l = 1, \ldots, q$	on day l for plan j.
	$j = 1, \ldots, n$	
$\hat{b}_{kj}^{(l)}$	$k = 1, \ldots, p$	Midpoint of interval of bed requirement in ward k
	$l = 1, \ldots, q$	on day l for plan j.
	$j = 1, \ldots, n$	

Notation	**Indicies**	**Definition**
Γ_k	$k = 1, \ldots, p$	Protection level applied to ward k
$p_{kj}^{(l)}$	$k = 1, \ldots, p$ $l = 1, \ldots, q$ $j = 1, \ldots, n$	Dual decision variable in robust counterpart
$q_k^{(l)}$	$k = 1, \ldots, p$ $l = 1, \ldots, q$ $j = 1, \ldots, n$	Dual decision variable in robust counterpart
N		Number of scenarios

Chapter 1

Introduction

Operating theatre scheduling is widely accepted to be a very complex process due to the demand for hospital resources and the impact it has on the running of the entire hospital. This thesis investigates the problems associated with operating theatre scheduling, and investigates optimisation methods that tackle these problems, with the aim of reducing the number of cancellations of elective operations. A novel approach to the scheduling process that incorporates the demand on downstream hospital resources, specifically post-operative beds, is developed and investigated.

This chapter introduces the background and associated challenges involved with the scheduling of operating theatres. The research aims are presented, with an overview of the structure of the thesis given to set out how these aims will be met.

1.1 Motivation

1.1.1 Operating Theatre Scheduling

The operating theatre department has been described as one of the major areas within a hospital with respect to its running costs and impact on other departments in the hospital. Indeed, it is said to be 'the engine that drives the hospital' [20]. Surgical suites have very high costs associated with their function, with staff costs forming the majority of the running costs [111]. Operating theatres have also been found to be the source of almost 70% of hospital admissions [17], with surgical patients providing a significant proportion of the demand on other hospital departments, both before and after surgery [112]. Having such an impact on other hospital resources requires careful planning in order to ensure the smooth running of the hospital within tight resource and budgetary constraints.

Operating theatres are dedicated to the provision of surgery for a number of surgical specialties, where operations are carried out in the theatres in blocks of time that are allocated to a specific specialty. The schedule that specifies these allocations is known as the Master Surgery Schedule (MSS), and is often a weekly timetable that varies very little from week to week. Surgery takes place on two types of patients: emergency and elective patients. Emergency patients are unplanned patients who require surgery as soon as possible, whereas elective patients are planned in advance and can be categorised into day cases and inpatients. Only inpatients require a stay in a hospital ward bed for post-operative recovery, whereas day cases leave hospital on the day of surgery and recover at home. Clearly, inpatients require the use of more hospital resources, and so more planning for this type of surgery is required. This research is focused on the scheduling of elective inpatients, with the impact of emergency patients on hospital resources being accounted for in the scheduling process. Schedules for the day case operating theatres could also be readily accommodated into the methods presented in this thesis, however due to the absence of post-operative bed requirements for day cases, these schedules are considered more trivial to construct in comparison with inpatient MSSs.

1.1.2 Associated Problems

There are a number of problems that have been identified with the scheduling of operating theatres, as identified in government publications and from discussions with hospital staff. A major factor to consider when scheduling operations is the availability of resources required for an operation to take place. An operation requires vital equipment, a variety of consumables, and a range of staff to be present, including surgeons, anaesthetists, nurses and technicians.

As well as ensuring the availability of resources for operations, the arrival of emergency patients also needs to be recognised. Emergency patients require treatment much more urgently than elective patients, so have priority for the use of operating theatres, often resulting in the cancellation of elective operations. Although it is not known with certainty when emergency patients may arrive, certain measures can be taken in order to minimise the adverse knock-on affects for elective patients. Such measures include having a dedicated operating theatre for emergency patients, thus allowing elective operations to continue in other operating theatres [93], and to consider the number of beds on wards that may be required by emergency patients, rather than being available for elective patients.

Both the unavailability of hospital resources and the occurrence of emergency patients can result in the cancellation of elective inpatient operations. This is the main problem that the case study hospital would like to address through this research: to mimimise the number of cancellations of elective surgery through careful scheduling of the operating theatres. Cancellations can upset the flow of patients through the hospital and negatively affect the quality of the patient experience – a key target set by the Welsh government [158]. A recent audit of the operating theatre services in Cardiff and Vale University Health Board (CaV UHB) reported that 'cancellations due to lack of beds was identified as a common problem' [157]. Recovery facilities in the hospital, such as high dependency beds and the surgical wards, were found to cause 'bottlenecks' in the system, particularly for accessing the Critical Care Unit (CCU).

One recommendation from the Welsh Audit Office report, particularly relevant to this research, is the need for 'modelling bed capacity against service reconfiguration to ensure bed availability does not cause cancellations' [157]. Other UK and Welsh Government targets aim to ensure the operating theatres are utilised fully in order to be cost effective, meet waiting time targets and create a more positive patient experience [158].

1.2 Research Objectives

As outlined above, this research is primarily concerned with the investigation of the construction of the MSS for operating theatres and its impact on the demand for beds on hospital wards. Analysis of data provided by CaV UHB concerning the University Hospital of Wales (UHW), Cardiff, will help explore the relationship between the operating theatres and beds on wards, as well as informing models to be developed.

The research will employ statistical and operational research techniques to provide a framework for the tactical level of hospital planning, in which an MSS can be constructed that is robust to the uncertainty associated with the post-operative bed requirements of surgical patients. It is also intended that the developed scheduling approaches will reduce the number of cancellations of elective surgery, which is currently such an important problem across the National Health Service (NHS) in the UK.

The research presented here is applicable to any hospital in which elective inpatient surgery is performed, however, the results relate specifically to the case study hospital. The insights gained and methodologies developed could be extended to any hospital in Wales, or indeed any similar surgical services in the UK or the world, due to the generic nature of the model formulations.

The aims of this research can be summarised by the following objectives:

- Investigate the relationship between the MSS and the resultant bed demand on surgical wards.
- Understand the factors, if any, that affect why cancellations of elective operations occur, and identify whether they occur more frequently on particular wards.
- Develop optimisation models to construct an MSS that satisfy constraints on both the operating theatres and bed availability on wards.
- Evaluate robust optimisation techniques for the construction of the MSS that incorporate the uncertainty associated with post-operative bed requirements.

1.3 Thesis Overview

This thesis aims to address the research objectives outlined above, and is structured such that the background to the problem is introduced and discussed in relation to the case study hospital in the first three chapters. The research into the development of the scheduling models, including their verification and validation, is then covered in the next five chapters, with final conclusions and recommendations for future research being presented in Chapter 9.

A more detailed summary of the remainder of the thesis is as follows:

- Chapter 2 provides a review of the literature on operating theatre scheduling. Key publications in the field are reviewed, and an overview of techniques used in previous studies is presented;
- Chapter 3 describes the case study hospital and its current process of scheduling operations. Relevant data is analysed to provide context to the problem and for inputs of the models to be developed;
- Chapter 4 introduces a deterministic model for the construction of the MSS. A literature review of the set partitioning optimisation problem is provided

in order to demonstrate how the method can be applied to this scheduling problem;

- Chapter 5 evaluates the deterministic scheduling model when applied to the case study hospital. A variety of 'what-if' scenarios are used to investigate the effect of changing parameters within the model;

- Chapter 6 presents the research area of optimisation under uncertainty. A literature review is used to introduce methods that can be used to incorporate uncertainty associated with the post-operative bed requirements within optimisation models. Two techniques have been identified that have potential in this scheduling application – robust counterpart optimisation and scenario-based optimisation;

- Chapter 7 develops a robust counterpart optimisation model for the construction of the MSS. This approach is considered to be particularly attractive for hospital decision makers due to the ability to specify their preferences for protecting the MSS from uncertainty in the bed requirements;

- Chapter 8 extends the deterministic model developed in Chapter 4 into a scenario-based optimisation model. This is an alternative, data-driven method to incorporating uncertainty into the optimisation process. A number of experiments are performed to assess the effectiveness of the model and a comparison is drawn with the robust counterpart optimisation model;

- Chapter 9 draws together the conclusions of this research, evaluates the effectiveness of the models, and suggests possible directions for future research.

Chapter 2

Operating Theatre Scheduling Literature Review

This chapter provides a review of the academic literature on the scheduling of operating theatres via the use of operational research techniques. An overview of the variety of techniques employed for each stage of the scheduling process is given to provide perspective. The need for further research into operating theatre scheduling, as highlighted in the literature reviewed, is also described.

2.1 Introduction to Operating Theatre Scheduling

The issue of operating theatre planning and scheduling has been, and remains, an active area of academic research. Magerlein and Martin [112] published the first extensive review paper on operating theatre scheduling in 1978. More recently Cardoen et al. [48] and Guerriero and Guido [90] have published reviews on this topic. The review by Cardoen et al. classifies papers into a diverse range and well defined areas of research.

Many NHS driven initiatives have been introduced over the years that focus on the smooth running of the operating theatres in hospitals. For example, The Productive Operating Theatre programme [127] focused on changes and improvements that can be made to ensure value and efficiency of the operating theatres, staff performance, and safety and reliability of patient care.

Cardoen et al. [48] define operating theatre planning to be concerned with the supply and demand (i.e. capacity decisions) of the surgery department, and operating theatre scheduling to be the construction of a timetable that specifies the ordering and allotted times for surgeries.

Operating theatre planning and scheduling is an interesting and challenging problem due to the large number of variable factors that can affect operations. The main uncertainties related to scheduling operations, according to Van Oostrum et al. [154], are the stochastic duration of surgical procedures, personnel availability, the no-show of patients and the occurrence of emergency surgical procedures. Gerchak et al. [87] found that the durations of elective surgeries vary according to the complexity of the surgical procedures and the surgeons themselves.

In addition to improving the efficiency of hospital resources, the improved scheduling of operating theatres also aims to provide patients with a better quality of care. Archer and Macario [9] discuss the ever increasing pressure on hospitals to deliver quality care at low cost. They suggest which areas need to be improved concerning the operating theatre, and note that improving scheduling efficiency is a positive way forward in tackling these quality and efficiency problems.

Strum et al. [146] discuss how operating theatre efficiency is related to its utilisation, and suggest strategies on how to increase the utilisation. Santibanez et al. [140] discuss possible benefits of a systematic approach to surgery scheduling. These include:

- Increased efficiency of the operating theatre;
- Increased patient throughput of the operating theatre;
- Lower wait times for both patients and hospital staff.

2.2 Stages of Operating Theatre Scheduling

A number of different stages in the scheduling of operating theatres have been identified in the literature. Blake and Carter [38] present a conceptual framework for operating theatre scheduling that is split into three levels of decision making: strategic, tactical and operational. Strategic level planning involves long term decisions, typically performed annually. Tactical level decisions relate to a medium term, quarterly planning horizon, whereas the operational planning level involves day-to-day decisions on the running of the operating theatres. Actual, known patients are considered at the operational level, whereas tactical and strategic level planning deals with expected patients. Cardoen et al. [48] comment that the boundaries between these levels are hard to define, and authors of other publications in the field do not tend to use consistent definitions.

The strategic, tactical and operational levels of operating theatre scheduling have been identified in the literature as case mix planning, the construction of the MSS, and elective patient scheduling respectively [19, 49]. Decisions made at each stage form a hierarchy for operating theatre scheduling, meaning that the outputs of a higher stage can be used as inputs to inform the next lower stage. The stages of operating theatre scheduling are shown in Figure 2.1, and discussed further in the following sections.

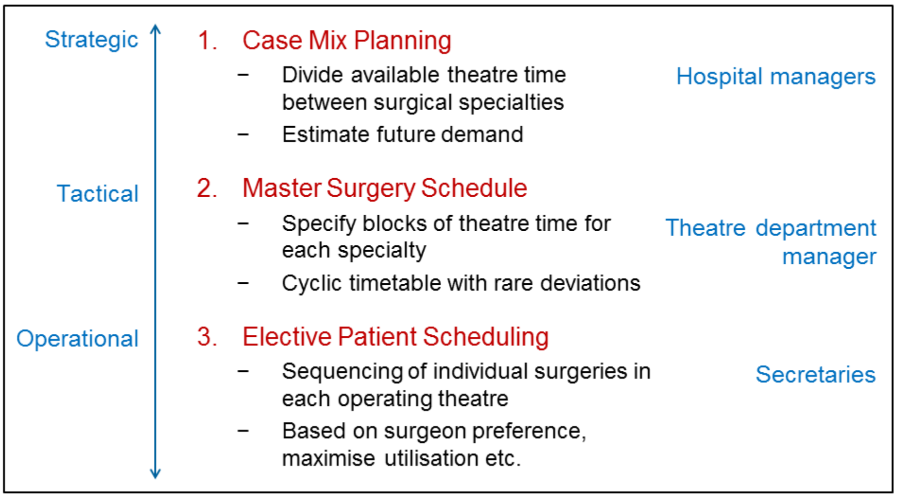


Figure 2.1: Stages of Operating Theatre Scheduling

2.2.1 Case Mix Planning

Case mix planning is performed at the highest level of operating theatre planning, and is used for strategic purposes. Senior management are interested at this stage of planning since decisions on committing resources such as money, staff and theatre time are made. Case mix planning is usually done on an annual basis [140].

During this phase of operating theatre planning, available operating theatre time is divided and assigned to surgeons or specialties. This assignment can be based on different criteria [148], for example, total cases per allocated block (historical utilisation), hospital costs and gains per allocated block (financial criteria), and demand for services (waiting list).

Some surgeons find it hard to reconcile the needs of their own specialty and other specialties, as to how operating theatre time is split between the specialties [140]. An objective of case mix planning is to divide this time as fairly as possible between the specialties. There are, however, factors that affect the proportions of time for each specialty. The waiting times of patients for surgery are a main concern for hospitals so that patient quality of care is maintained. Equity among specialties is desirable [37], and the maximisation of operating theatre utilisation is also sought for financial benefits [63].

Dexter et al. [62] have considered financial criteria and uncertainty of the future workload to determine initial operating theatre allocations for surgical sub-specialties using linear programming. They found that this stage of planning can be performed up to one year in advance in order for management to make strategic decisions. By estimating lower and upper limits on future demand, the authors also showed that the initial allocation of operating theatre time can be performed with only this partial information available.

Trade-offs between cost, throughput of patients and clinical necessity were used by Blake and Carter [39] to determine the case mix of patients within a hospital. Bed availability is considered as a constraint in their linear goal programming approach and their model has been implemented successfully in a large teaching hospital in Canada.

Adan and Vissers [5] formulate a mixed integer programming (MIP) model that identifies the number and mix of patients that must be admitted into a hospital in order to gain the target utilisation of important resources, e.g. the operating theatre or intensive care unit. They consider both inpatients and outpatients, where outpatients are considered as inpatients with a length of stay (LoS) of one day.

Time series analysis has been used to forecast the total number of hours of elective surgery required in the future in order to allocate operating theatre time to each specialty. Dexter et al. [64] found that using the average of the most recent year's total hours of elective surgeries is a valid way of forecasting the future usage of operating theatre time.

2.2.2 Master Surgery Schedule

The tactical level of operating theatre scheduling involves the construction of an MSS. At this stage the number and type of operating theatres are defined, the hours of operating theatre time available is stated and the specialty that has priority in each operating theatre is given [36]. The MSS is a cyclic timetable that usually has a cycle time of one week, requiring that only the surgical procedures of a certain type are scheduled and not the specific procedures of actual patients [154]. Deviations from this cyclic timetable are discouraged by the hospital.

Since this level of planning is the main focus of this research, a more thorough discussion of previous techniques and models used for the construction of the MSS in the literature is given in Section 2.3.

2.2.3 Elective Patient Scheduling

Given a particular MSS, the final stage of operating theatre scheduling is to schedule individual patients for their surgeries. This is at the detailed, operational level of planning and is often performed on a daily basis. For each operating theatre available to a specialty, the patients for that day are scheduled such that various criteria are met [48]. These criteria may relate to the surgeons' preferences (e.g. order of surgeries for clinical reasons), resource availability, maximisation of throughput, efficiency and utilisation of the operating theatre, and minimisation of staff overtime.

Magerlein and Martin [112] define the dichotomy of advance and allocation scheduling. Advance scheduling is when a surgery date is fixed for a specific patient in the future, whereas allocation scheduling sequences a number of surgical cases for a given day by determining the operating theatre and start time of the procedure, assuming that the corresponding patients are ready for surgery in the hospital.

The reviewed literature can be split into two themes: the construction of a schedule for elective patients, and the improvement of existing schedules. These areas are discussed in more detail in the following sections.

Construction of Schedules

Guinet and Chaabane [91] propose that the scheduling of elective patients should be done in two steps. First they assign patients to operating theatres over the planning horizon, and then each operating theatre is scheduled individually. The surgeries in an individual operating theatre are scheduled in such a way that human and material resources are considered, as well as patient hospitalisation date and a surgery deadline, in order to maximise patient satisfaction and resource efficiency. The authors describe a primal-dual heuristic to solve the assignment problem.

A two-step approach for operating theatre scheduling is also used by Jebali et al. [96] to determine the schedule for the next operating day. In the first step, surgical operations are assigned operating theatres with the aim of minimising overtime, undertime and patient waiting time using a mixed integer program. The second step then sequences the operations that have been assigned to each operating theatre in the previous step using a MIP model which sequences the operations to minimise overtime, and by considering the recovery room beds as a bottleneck resource.

Saadouli et al. [139] incorporate the stochastic nature of surgery durations and the availability of post-operative beds for an orthopaedic specialty when generating a schedule for elective patients. An additive slack is given to the duration of each surgery, and a knapsack model is applied to generate daily schedules that maximise operating theatre utilisation. Discrete event simulation is then employed to evaluate the resulting schedules.

Another uncertain aspect associated with the scheduling of operations is the occurrence of emergency patients. Lamiri et al. [101] use a stochastic MIP model to generate a schedule for elective patients, whilst incorporating the possibility of emergency patients occurring over the planning horizon. Experimental results show that running costs can be significantly reduced by using a stochastic model where uncertainty related to emergency surgery is explicitly considered.

Van Houdenhoven et al. [152] evaluate several scenarios in which a bin-packing algorithm is used to optimise the operating theatre case schedule. The planned slack within the schedule is minimised by making use of the portfolio effect for multiple operations with similar variation of duration. Based on data from a large teaching hospital, it was found that this approach could yield a 4.5% increase in operating theatre utilisation.

It has been shown by Dexter et al. [66] that by building planned slack into an operating theatre schedule, the likelihood of operations starting at their scheduled start times can increase. They show that this can be done by calculating the upper prediction bound for the duration of the cases performed later on in the day.

Improving Schedules

As discussed above, there are a number of ways in which operating theatre schedules can be constructed, however, there are also papers in the literature that show that the schedules can be further improved.

Dexter et al. [65] evaluate ten scheduling algorithms that can be used to schedule additional add-on elective cases to the operating theatre schedule. This daily process happens once the operating theatre schedules have been submitted and approved for the next day. Using simulation, the approach was found to increase operating theatre utilisation by performing more operations in the 'open time' of the operating theatre schedules.

Gerchak et al. [87] have also considered how to schedule add-on elective cases to the operating theatre schedule. Stochastic dynamic programming was used to determine how many of the additional requests for add-on cases should be accepted, when the operating theatre capacity is uncertain (due to variable operation duration and unscheduled emergency cases).

2.2.4 All Three Stages

While most publications reviewed are concerned with only one stage of the operating theatre scheduling process, Testi et al. [148] have developed a hierarchical modelling approach to operating theatre scheduling. Their three-phase model integrates all stages discussed above into one model that has been implemented in a surgical department in Genova. The first, case-mix planning phase (referred to by the authors as session planning) is solved using a bin-packing algorithm that distributes available operating theatre time over the surgical specialties. The MSS is then constructed using an integer programming model in which an operating theatre is assigned an amount of time to each surgical specialty. Finally, discrete event simulation is used to model a variety of different sequencing rules for booking inpatients for specific dates, rooms and times. This elective patient scheduling model also considers the downstream ward capacity. A key finding from this research was the consequential reorganisation of the recovery beds into short and long stay areas. This was found to increase the utilisation of the operating theatres and increase throughput of patients.

2.3 Construction of the Master Surgery Schedule

Many factors need to be considered in the creation of an MSS, including the compatibility between operating theatres and the specialties working in them (i.e. ensuring that the correct equipment is in the appropriate operating theatre), the availability of surgeons and whether there are enough post-operative resources, for example critical care beds [140].

A large amount of research has been carried out relating to the construction of the MSS as will be discussed below; however, van Oostrum et al. [153] comment that the impact of this research is very limited in practice. In their paper, the authors discuss the potential problems that might arise when implementing an MSS both for the researcher and healthcare organisation. In relation to these implementation problems, Belien et al. [20] have developed a software package that represents visually the impact of different MSSs on various resources throughout the hospital.

Studies in the literature use different performance measures in order to determine the effectiveness of the operating theatre scheduling procedures. Common performance measures, as categorised by Cardoen et al. [48], include waiting time, patient throughput, operating theatre utilisation, resource levelling, patient deferrals, financial measures and surgeon preferences. These performance measures are discussed in Table 2.1.

Performance Measure	**Aim**	**Examples**
Waiting time	To decrease waiting times for patients and surgeons.	Wullink et al. [161] use discrete event simulation to minimise patient waiting times for emergency surgeries. Denton et al. [58] propose stochastic optimisation models to find sequencing rules that minimise surgeon waiting times.

Patient throughput	To increase the number of patients treated.	Van Berkel and Blake [155], using discrete event simulation, showed that by changing the bed capacity in wards and the amount of operating theatre time available, the throughput of patients increased which also caused waiting times to decrease.
Operating theatre utilisation	To keep the operating theatre running at a desired level of utilisation.	Tyler et al. [151] use simulation to determine the best utilisation of an operating theatre. They find that a utilisation of 85 – 95% allows for uncertainty of operation durations and start times. An in-depth discussion on the pros and cons of over-/under-utilisation is given in van Houdenhoven et al. [152].
Resource levelling	To have smooth use of hospital resources, i.e. no peaks in demand.	Marcon and Dexter [114] consider the levelling of resources in the post anaesthesia care unit and recovery area, as well as within the operating theatre.
Patient deferrals	To minimise the number of patients who are deferred or decline treatment.	Kim and Horowitz [98] study how the number of cancelled surgeries can be reduced by considering the need for post-operative admission to the CCU within the operating theatre schedule. A simulation model is used to model the many pathways of patients to the CCU.

Financial measures	To minimise operating theatre costs.	This performance measure is particularly popular in research that relates to American hospitals. Cardoen et al. [48] believe more research should be done on this as any cost savings can be invested back into solving any of the other problems above. Dexter et al. [60] present a case study that looks at the effect on profit margins when throughput increased, while Dexter et al. [61] also considered the uncertainty in the surgeons' future workloads.

Table 2.1: Common performance measures

A small number of papers in the literature consider the MSS for only one surgical team or only one operating theatre. Vissers et al. [156] use MIP to construct a master timetable for the cardiothoracic surgery department with a four week cycle time. A number of resources such as nursing staff and intensive care beds are considered as constraints in the model.

The majority of papers in the literature concern the construction of the MSS for multiple surgical specialties needing to be assigned to multiple operating theatres. A variety of modelling approaches have been used, however, from the literature review carried out by Cardoen et al. [48], it was found that the most common technique used for operating theatre planning and scheduling is mathematical programming. In particular, MIP was found to be the most commonly used approach. Here we provide a discussion on a selection of papers to illustrate how mathematical programming techniques have been used to construct the MSS.

MIP is used by Blake et al. [36, 37] to produce a schedule that minimises the shortfall between the target and actual assignment of operating theatre time for each surgical group. Their scheduling model has been implemented and used by Mount Sinai hospital in Toronto, Canada. Van Oostrum et al. [154] also use an MIP model to construct an MSS that uses a column generation technique to find a solution. The stochastic nature of the duration of surgical procedures is considered, and planned slack is built into the timetable in order to account for this. Their

MIP model aims to maximise the operating theatre utilisation as well as levelling the subsequent hospital bed requirements. Adan et al. [4] also formulate an MIP model that follows a goal programming approach in order to create an MSS that allows reservation of some operating theatre capacity for emergency patients.

Kuo et al. [100] use integer programming to allocate operating theatre time to multiple operating theatres in order to maximise surgeon revenue in American hospitals; the results of which indicate a 15% increase in revenue. This research, however, relied upon the assumption that there was not a shortage of intensive care beds or nursing staff.

A number of MIP and quadratic programming models for constructing the MSS were proposed by Belien and Demeulemeester [19]. They evaluate these methods by considering the resulting bed occupancy after surgeries, with the aim to level the demand as much as possible. They build a model that minimises the total expected bed shortage with constraints on the demand for operating theatre blocks for each surgical group and on the capacity of the number of available operating theatre blocks each day. Belien et al. [21] subsequently discuss a decision support system for the implementation of these models in a large hospital. They find that the different models provide slightly different schedules, but that it is up to the hospital managers to choose the 'best' schedule.

Less common modelling approaches have also been used to construct an MSS, such as Vanberkel et al. [155] who used a queuing theory approach to build the MSS in such a way that demand on downstream hospital departments is predicted and taken account of in the MSS. Strum et al. [146] use a minimal cost analysis model to assist with optimising subspecialty operating theatre block time allotments. Their model uses estimates of the costs of under- and over- utilisation of operating theatres in order to allocate operating theatre time to surgical subspecialties at minimum cost.

Note that most of the literature reviewed is concerned with finding the MSS for surgery within one hospital. Santibanez et al. [140] consider the more complex problem of allocating operating theatre time to specialties across multiple hospitals. An MIP model is formulated to construct the MSS with two objectives: to reduce the variability in bed utilisation (achieved by minimising the maximum daily bed utilisation), and to maximise the throughput and mix of patients.

2.4 Future Research Opportunities

There are many aspects of operating theatre scheduling that have yet to be considered or expanded upon in the literature, as discussed in the review paper of May et al. [117]. The most relevant include the need to consider the affect of the MSS on other hospital resources, and to take account of the stochastic nature of the length operations and the post-operative LoS.

In order for more factors to be considered in the modelling process, Santibanez et al. [140] suggest that the characteristics of individual surgeons could be included in the model. Since each surgeon performs a different mix of procedures, input of specific demand would create a more hospital specific model. Cardoen et al. [48] believe this would have a larger success rate when scheduling is performed at the surgeon level and not the patient level.

Cardoen et al. [48] recommend that global performance within a hospital could be improved by incorporating other hospital facilities in the scheduling process. Since the operating theatre suite is a main driver of demand in the hospital, the consideration of upstream and/or downstream departments is important. These facilities do not have to be limited to within one hospital.

It is also recommended by Lamiri et al. [101] that more research should be carried out when the stochastic nature of the operating theatre is taken into account, both for the arrival of emergency patients and the duration of surgical procedures. The uncertainty that relates to the availability of resources should also be considered [121, 130].

2.5 Summary

This chapter has provided a preliminary review of the literature that is relevant to the scheduling of operating theatres. A particular focus on the tactical planning stage involving the construction of the MSS has been given. A variety of operational research techniques have been employed for each stage of the scheduling process, however, it is clear that MIP has been most commonly used. Gaps in the existing literature concern the inclusion of up- and down-stream hospital resources, and the consideration of the stochastic nature of operations. The research presented in the subsequent chapters aims to address these issues.

Chapter 3

Description of the Case Study Hospital

This chapter introduces the case study hospital used in this research. In particular, current working practices are described and relevant data analysis is presented. Extensive data analyses relating to theatre and ward activity in the hospital is carried out to provide context and to derive inputs for the developed models.

3.1 The Case Study Hospital: University Hospital of Wales, Cardiff

The case study hospital for this research is UHW, Cardiff. It is the largest hospital within the CaV UHB, and is indeed the largest hospital in Wales. CaV UHB is a teaching health board that has strong links with universities in South Wales, in particular with the School of Medicine, Cardiff University. The health board serves a population of around 500,000 people in Cardiff and the surrounding region of the Vale of Glamorgan.

UHW is the largest hospital in CaV UHB, with an average of 987 beds available for use in the year 2012/13 and an average 88.0% occupancy rate according to figures published by the Welsh Government [1]. It has five tertiary referral centres that offer highly specialist services for cardiothoracic surgery, neurosurgery, transplant surgery, critical care and haematology. Inpatient and day-case surgery is performed in UHW in two locations: inpatient operations in the larger, main theatres and day-case operations in the Short Stay Surgical Unit (SSSU). The SSSU has a dedicated ward for day-case surgical patients, whereas inpatients who have operations in the main theatres stay in beds on one of the surgical wards in

UHW. Inpatient operations are performed in the SSSU on very few occasions, so the impact of this on the main theatres is assumed to be negligible and is therefore not included for purposes of the modelling.

There are 18 surgical specialties that use the main theatres in UHW. A suite of 14 operating theatres is available for use by these specialties, however, some specialties require specialist equipment that is only available in particular theatres. The theatres are utilised by the specialties according to the MSS; the current MSS that is used in UHW for the main theatres is shown in Figure 3.1.

	Theatre 0	Theatre 1	Theatre 2	Theatre 3	Theatre 4	Theatre 5	Theatre 6	Theatre 7	Theatre 8	Theatre 9	Theatre 10	Theatre 11	Theatre 12	Theatre 14
MONDAY AM	Trauma	Scoliosis	ENT	Renal	Oral Surg	CEPOD	Urology	Colorectal	General	Thoracic	Cardiac	Cardiac	Neuro	Neuro
MONDAY PM	Trauma	Scoliosis	ENT	Renal	Oral Surg	CEPOD	Urology	Colorectal	General	Thoracic	Cardiac	Cardiac	Neuro	Neuro
TUESDAY AM	Trauma	Vascular	Ophthalmology	Vascular	Paeds General	CEPOD	Urology	Colorectal	Renal	Thoracic	Cardiac	Cardiac	Neuro	Neuro
TUESDAY PM	Trauma	Vascular	Ophthalmology	Vascular	Paeds General	CEPOD	Urology	Colorectal	Renal	Thoracic	Cardiac	Cardiac	Neuro	Neuro
WEDNESDAY AM	Trauma	Paeds Ortho	ENT	General	Paeds General	CEPOD	Urology	Colorectal	General	Thoracic	Cardiac	Cardiac	Neuro	Neuro
WEDNESDAY PM	Trauma	Paeds Ortho	ENT	General	Paeds General	CEPOD	Urology	Colorectal	General	Thoracic	Cardiac	Cardiac	Neuro	Neuro
THURSDAY AM	Trauma	Trauma	Oral Surg	Vascular	Paeds General	CEPOD	Urology	Colorectal	General	Thoracic	Cardiac	Cardiac	Neuro	Neuro
THURSDAY PM	Trauma	Trauma	Oral Surg	Vascular	Paeds General	CEPOD	Urology	Colorectal	General	Thoracic	Cardiac	Cardiac	Neuro	Neuro
FRIDAY AM	Trauma	Scoliosis	Paeds ENT	Vascular	Paeds General	CEPOD	Urology	Renal	Liver Surgery	Oral Surg	Cardiac	Cardiac	Neuro	Neuro
FRIDAY PM	Trauma	Scoliosis	ENT	Vascular	Paeds General	CEPOD	Urology	Renal	Liver Surgery	Oral Surg	Cardiac	Cardiac	Neuro	Neuro

Figure 3.1: The current MSS used for the main theatres in UHW

As can be seen in Figure 3.1, specialties are assigned to theatres in whole or half-day sessions in which the specified specialty has sole use of the theatre. Morning sessions run from 8.30am to 12.30pm, and afternoon sessions from 1.30pm to 5pm. The construction or monitoring of the MSS is not currently undertaken by any one person within UHW. Senior managers monitor the balance between demand (the number of operations required from emergencies and the elective waiting lists) and activity (the number of operations performed) in the operating theatres and make adjustments to the MSS when required. For example, extra sessions may be given to a specialty that has a particularly long waiting list in order to treat patients and reduce the waiting list. These extra sessions are taken from other specialties that could temporarily cope with a reduced operating theatre time.

Within a week of surgery, elective surgical inpatients are required to attend an appointment in the SSSU outpatient clinic for a pre-operative assessment to determine whether they are medically fit for their planned operation. Some patients are then admitted to a bed in hospital before their elective surgery in order for doctors to monitor them and perform pre-operative tests. Patients who do not require this supervision before an operation can arrive at the hospital on the day of surgery through what is known as the Theatre Admissions Lounge (TAL), thus avoiding the need for a bed before surgery. Just before the scheduled start time for surgery the patient is brought from the ward or TAL to the suite of main operating theatres. Each operating theatre has an adjacent anaesthetic room in which the patient is anaesthetised whilst the theatre is being cleaned and prepared for surgery. Once the operation is complete, the patient is moved to the recovery ward in the operating theatre area where they are closely monitored until the patient is ready to continue their post-operative recovery on one of the surgical wards. Patients who were in a bed before surgery will go back to the same bed after surgery. In some cases, patients will need to recover from surgery in the CCU where they will receive the highest level of care.

Numerous resources are required for operations, including specialist equipment and staff. If these resources are not available, then operations can be cancelled. Medical staff that are required in surgery include the consultant surgeon who will either perform the surgery, or oversee a trainee surgeon, an anaesthestist (two are required for paediatric surgery), and scrub nurses. For a given MSS, these members of staff will be scheduled by their own department in order that the required numbers of each skill-set is present for each surgery. For example, the anaesthetic department schedules the anaesthetists approximately a week in advance of surgery.

Elective inpatient operations are currently scheduled by the consultant surgeons working together with their secretaries. Around three weeks prior to the date of surgery, the secretary generates a list of patients that should be operated on during the session. This is often based on how urgent the patient requires surgery due to their medical needs, and how close the patients are to breaching the Referral to Treatment (RTT) time. The RTT time is the time from when a patient is referred to an outpatient clinic in UHW from a GP, to when they have surgery. The current targets in Wales are a maximum RTT time of 26 weeks for at least 95% of patients, and for those who are not treated within 26 weeks to be treated within 36 weeks from referral [159]. As of January 2015, CaV UHB was the worst performing health board in Wales in terms of the RTT targets, with 81.7% of patients waiting less than 26 weeks for treatment and 93.1% of patients waiting less than 36 weeks [124].

Once the specific patients are chosen for surgery, the consultant surgeon then decides the order of patients on which to operate. This is based on his experience of past operations, the expected duration of surgeries and equipment availability. For example, paediatric surgery is performed on the youngest to oldest children, since younger patients are more likely to have adverse reactions to the anaesthetic or complications in surgery.

On the day before surgery, representatives of each surgical specialty attend bed management meetings in which the number of beds that are available in the hospital is discussed. During this meeting, each specialty puts forward how many beds they require for their planned elective surgeries the next day. Depending on the current capacity of UHW, either all patients are confirmed for surgery, or some or all of the elective surgery has to be cancelled due to a lack of beds on the wards. The ideal level of bed capacity that the managers of UHW prefer to run at is 85%, but in recent years it has experienced very high levels of 95–98% during winter months.

Depending on the outcome of the bed management meeting, the final lists of elective patients are signed off by the consultant surgeon and are submitted to the theatre directorate by 3pm the day before surgery. These theatre lists are then distributed to other departments in the hospital, e.g. blood bank, x-ray and the wards, to ensure that resources and equipment are available at the required times.

If some elective operations had to be cancelled as a result of a lack of beds available, the consultant surgeon decides which surgeries will be cancelled. Patients that can wait longer for surgery, based on medical reasons, even if they are close to breaching their RTT target, are cancelled to enable the more medically urgent operations to be carried out. In extreme cases, if a surgery is considered very urgent, the surgery will go ahead and the patient will be put in a bed on a ward that is not their specialty's designated ward. This is known as 'outlying' on another ward and is discouraged as it prevents patients from the other ward being able to be brought in for their surgery.

3.2 Data Provided by UHW

Data relating to both the operating theatres and patient stays on wards in UHW were provided by the data team in the Surgical Support Services directorate of CaV UHB. Two main datasets were provided, with additional datasets supplementing analyses when required. The data provided by CaV UHB includes records for every operation carried out in the health board from 1st April 2009 to 31st March 2013.

Data on operating theatre activity was extracted from the TheatreMan database that is used in CaV UHB. TheatreMan is a software package that is available on desktop computers in each operating theatre allowing staff to record aspects of every surgical procedure performed in real-time, such as patient information, staff present in the operating theatre, and start and end times of surgical procedures [150].

Data relating to patient information and LoS was captured in a separate database and covers the same period from 1st April 2009 to 31st March 2013. Initial data manipulation and processing was required to merge these two datasets into one, master dataset. The datasets were merged using the SAS statistical software package [142], based on a unique patient identifier assigned to each patient in the health board. The master dataset therefore resulted in data on patient information, operating theatre activity and LoS for each patient in UHW for this time period. Subsequent data analysis reported in this chapter is performed on the master dataset.

3.2.1 Data Validation

In order to ensure that the subsequent analysis is performed on clean and accurate data, a number of validation checks were made on the data. Much of the information captured in the TheatreMan database is entered by medical staff in the operating theatre as the operation is being performed. Having to enter the data in such a stressful environment can cause difficulties in ensuring that accurate and complete data is recorded. As such, one validation check performed was whether the timings of surgery, for example, when anaesthetic was administered, surgery start time and surgery end time, were in chronological order. This was an easy test to perform via inspection of the data, however, it was not clear how to determine the true values of data that were not in chronological order, since no-one would be able to remember the exact timings of a past surgery. If theatre activity was found not to be recorded in the correct order, the difference between the first and last time recorded in TheatreMan is assumed to be the duration of surgery from when a patient enters the anaesthetic room to when the patient is moved to the recovery suite.

Merging the two datasets based on the unique patient identifier sometimes resulted in a mis-match of hospital stays and operation dates if a patient had more than one operation during the period for which we have data. These erroneous records were removed from the master dataset if the operation date did not lie between the start and end date of an episode in hospital.

From inspection of the data, there were found to be some outliers in the operation duration and LoS records. For example, several operations were recorded as taking less than 10 minutes, and a patient in the ENT specialty was recorded as being in the hospital for 211 days. Hospital staff were consulted in order to establish whether these types of values were errors in the data, or whether they were indeed possible. All records that were considered erroneous were removed.

3.2.2 Determining Specialties from TheatreMan Data

Surgical patients in UHW are assigned to a surgical specialty, depending on the care required. This is recorded in both the TheatreMan database and the patient information database, however, some discrepancies occur between the two datasets. For the purpose of this study, the surgical specialty specified in the TheatreMan database has been taken to be the surgical specialty in the master dataset, following discussion with hospital staff.

On inspection of the master dataset, it was found that the list of surgical specialties did not entirely match the list of surgical specialties that are named in the UHW MSS (Figure 3.1). In order to assign specialties that match those in the MSS to records in the master dataset, a number of criteria were defined on the specialties named in the TheatreMan dataset. Otherwise it was assumed that the field 'Actual Procedure Specialty' in the TheatreMan dataset was the correct MSS specialty. These criteria were defined with assistance from managerial staff in UHW and are based on factors such as the age of the patient, the theatre in which the surgery was performed, and the OPCS-4 code of the surgical procedure. OPCS-4 codes are set by the Office of Population and Censuses and Surveys to classify surgical interventions and procedures [2]. Each code consists of four characters, with the first character a letter, followed by three numbers. Criteria used for the classification of MSS surgical specialties to records in the master dataset are summarised in Table 3.1.

Specialty	**Criteria used**
CEPOD	If operation was performed in the emergency theatre, or if the patient was assigned to any of the emergency surgery sessions.
Oral	If 'Actual Procedure Specialty' = 'Oral surgery' or 'Maxillio-facial surgery'.
Scoliosis	If 'Actual Procedure Specialty' = 'Trauma' AND the patient is 16 years old or younger.
Vascular	If the OPCS-4 code starts with an L.
All other specialties	'Actual Procedure Specialty' field.

Table 3.1: Criteria used to assign MSS specialties to data

3.3 Demand for Surgeries in UHW

3.3.1 Number of Operations Performed in UHW

Figure 3.2 shows the number of operations carried out in the main theatres and the SSSU in UHW. The TheatreMan database was introduced to the SSSU in 2011, so there is only complete data for the last two years of the data collection period. For the final two years, it can be seen that the majority of operations are performed in the SSSU. The less complex operations performed in the SSSU take less time than the more complex surgeries in main theatres, hence more operations can be performed per year in the SSSU. The number of operations performed in both locations has remained steady from year to year, with an average of 11,657 operations per year carried out in the main theatres.

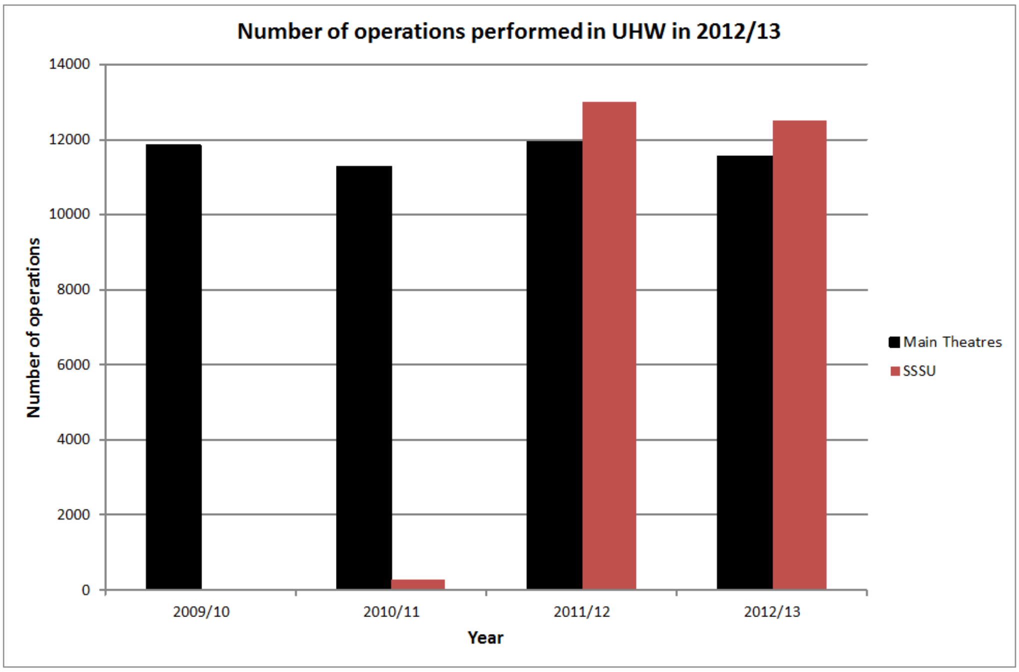


Figure 3.2: Number of operations performed in UHW in 2012/13

Data from the year 2012/13 for the main theatres will be analysed in all subsequent sections because 2012/13 is the most recent data and is considered the most complete and accurate data. Staff in UHW have described how, in times of high capacity on the wards, surgical patients who receive treatment in the main theatres are put in beds on the ward that is dedicated for the SSSU. This is not desirable since it results in the day-case surgeries being cancelled in order to accommodate the inpatients for post-operative recovery, and so is avoided if possible. The SSSU ward is not included in the data analysis or model, since it should not be relied upon for planning purposes and it is intended that the two theatre resources should be managed separately.

3.3.2 Elective and Emergency Operations

The total number of operations performed in the main theatres in UHW in 2012/13 is shown in Figure 3.3.

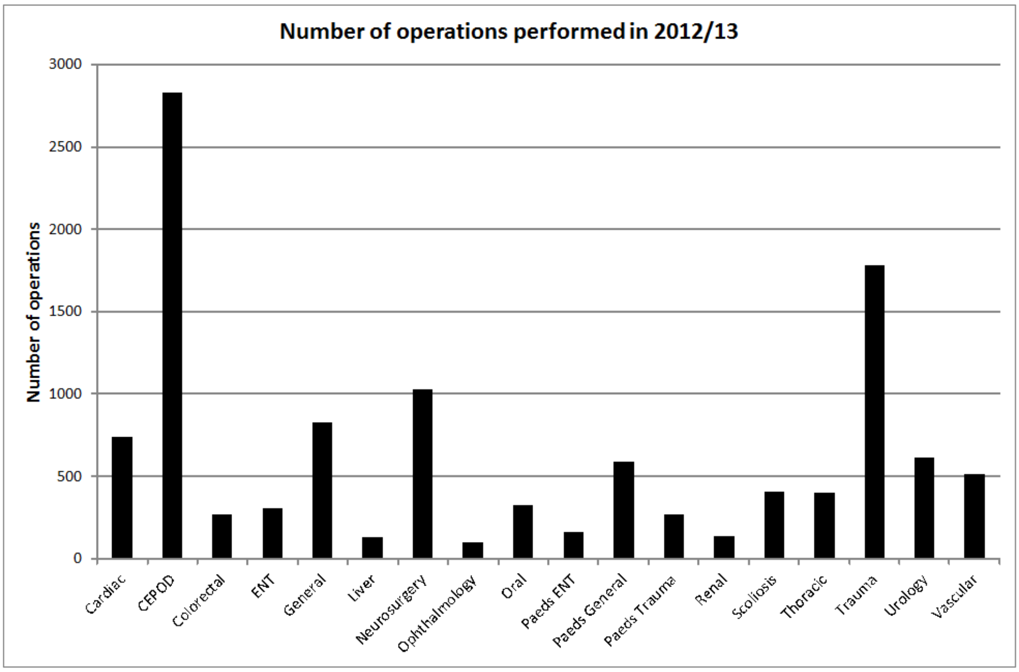


Figure 3.3: Number of operations performed in the main theatres in UHW in 2012/13

As can be seen in Figure 3.3, the CEPOD specialty performed the most operations in the year 2012/13. The CEPOD theatre is reserved for emergency surgery and is treated as a specialty for the construction of the MSS. Patients who are treated in the CEPOD theatre, however, belong to specific specialties that are aligned to their surgical procedure. A more detailed discussion and analysis of the CEPOD theatre and specialty is given in Section 3.3.4. The specialty that has the second highest number of operations is the Trauma specialty, which also carries out operations of an urgent nature. The Ophthalmology specialty performed the least amount of operations in the year 2012/13, which is not surprising since it only has one whole-day session per week in the current MSS used in UHW (Figure 3.1).

Surgery is classified in relation to the urgency of the surgery required. Elective surgery is performed to correct a non-life-threatening condition, and is planned or booked in advance of routine admission to hospital by request from a doctor or patient. Unplanned surgeries of a more urgent nature are classified as either 'urgent' or 'emergency'. Urgent surgery can wait until the patient is medically stable, but should generally be done within 48 hours of the patient being admitted to hospital. Emergency surgery is of the highest priority, which must be performed without delay to save life, limb, or functional capacity. In 2012/13, 72.7% of operations carried out in the main theatres were planned, elective operations, with the remainder being urgent or emergency operations. Figure 3.4 shows the percentages of operations that were classed as planned or emergencies for each specialty in 2012/13.

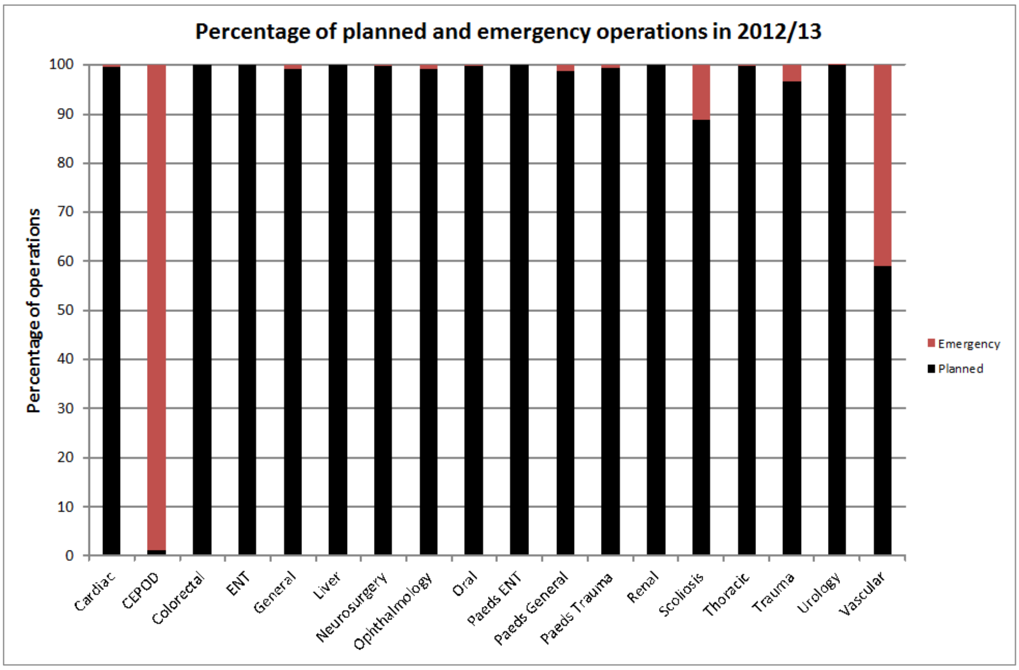


Figure 3.4: Percentage of planned and emergency operations in 2012/13

As can be seen in Figure 3.4, the CEPOD specialty has the highest proportion of emergency operations. The Vascular specialty also has a high proportion of emergency patients due to the urgent nature of the surgical procedures performed by this specialty.

Emergency patients in the CEPOD specialty are put onto one of three emergency lists, each of varying degree of urgency, that acts as a waiting list for emergency surgery. Emergency operations are recorded in the TheatreMan software as either the physical theatre in which the opeartions were performed, or which emergency list the patient was on. It is not possible to know from the data in which theatre the patients on the emergency lists had their operation, however, staff in UHW have advised that they are most likely to be carried out in the emergency CEPOD theatre. Over 95% of emergency operations in 2012/13 were recorded as being carried out in the CEPOD theatre, or were on one of the emergency lists. It was also found from the data that in 2012/13, 45% of emergency operations took place within the normal working hours of the operating theatres. The rest of the emergency operations took place either in extended sessions at the end of the day, or in additional sessions, for example on the weekend or at night.

3.3.3 Number of Operations per Session

The operating theatres have been identified as a driver of demand for many other hospital departments and resources, such as scanning machines and beds on the wards [48]. Particularly relevant to this research is the fact that every inpatient who has an operation requires a bed on a ward. Therefore, it is of interest to investigate how many operations take place in a session for each specialty, in order to determine the scale of demand for beds.

Figure 3.5 shows the distributions of the number of operations per session carried out in 2012/13 for each specialty. As expected, specialties that generally involve more complex, and hence longer operations have fewer operations per session than other specialties that perform less complex operations. For example, more Ophthalmology operations are able to be performed in an operating session than Cardiac operations.

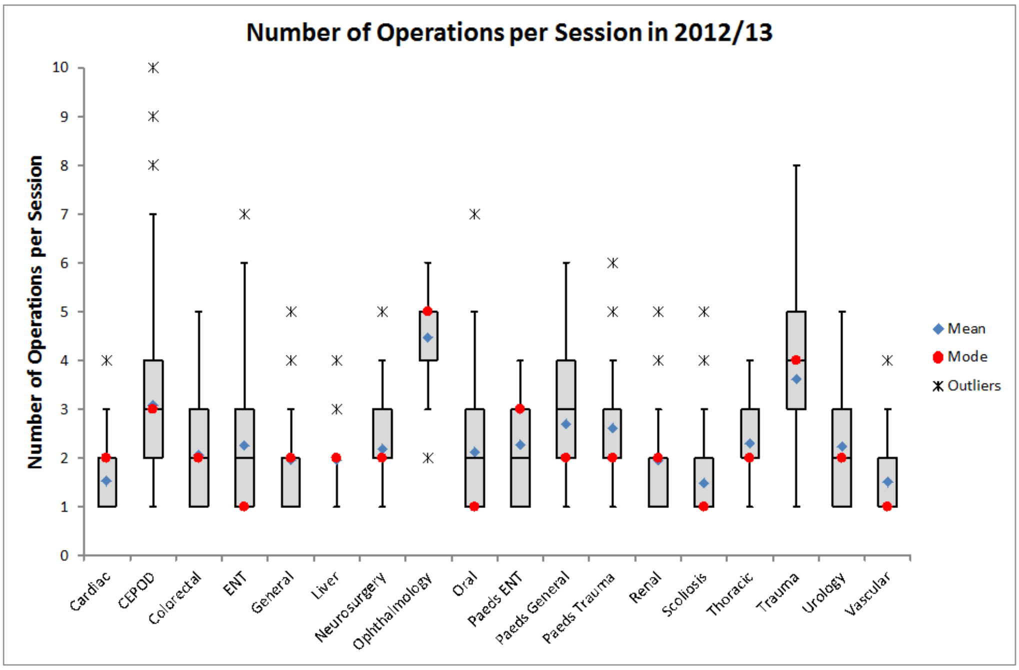


Figure 3.5: Number of observed operations per session in 2012/13

The above analysis of the observed number of operations per session is useful to examine what has happened in the past, however, the number of operations per session used in the model should be independent of any past circumstances that may have affected the number of operations per session. The number of operations per session can be affected by a number of reasons, including the long duration of some surgical procedures causing early starts and/or late finishes, thereby effectively lengthening the session time allocated to a specialty. By using the observed number of operations per session from past data, these problems are inherent in the data. Therefore, it is best not to use the observed number of operations per session from past data for planning purposes.

A method that is independent of these inherent problems in the data is to calculate the number of operations that is possible to perform during a session, given the length of time surgical procedures take. The total time for a patient to occupy an operating theatre is defined to be from the time a patient enters the anaesthetic room until the patient is moved to the recovery suite, and the operating theatre has been prepared for the next patient. The time between a patient being administered anaesthetic to when they leave the operating theatre for recovery is known from the TheatreMan dataset. Two additional lengths of time are required in order to calculate the total time a theatre is occupied by a patient; the time between the patient arriving in the anaesthetic room to when the anaesthesia is administered, and the turnaround time between patients, during which the operating theatre is cleaned and prepared for the next patient. Both of which are not recorded in the TheatreMan dataset.

Following discussions with operating theatre staff, it was agreed that the time taken in the anaesthetic room before the patient is anaesthetised is roughly 10 minutes for each patient. This time can vary depending on the medical needs of the patient, how anxious the patient is, and whether the anaesthetic team are ready to anaesthetise. It will be assumed here that every patient will spend 10 minutes in the anaesthetic room before the anaesthetic is administered.

As part of an audit of the operating theatres in UHW carried out in 2013 [157], it was reported that the time between patients, the turnaround time, ranges between 11 and 27 minutes, with an average of 22 minutes across all specialties. Since data regarding turnaround time is not recorded in the TheatreMan dataset, it will be assumed that a turnaround time of 22 minutes is associated with each operation. The total time that is associated with each patient in theatre is therefore the sum of the time in the anaesthetic room, the time for the surgical procedure, and the turnaround time.

The number of operations per session is calculated in the following way:

$$\text{Number of operations per session} = \frac{\text{Session duration}}{\text{Total duration in theatre per patient}}$$

It is assumed that specialties either have half-day (3.5 hours = 210 minutes) or whole-day (7 hours = 420 minutes) sessions, as specified in the current MSS used in UHW (Figure 3.1). The calculations for the number of operations per session for each specialty, based on the procedure lengths from the observed data from 2012/13, are summarised in Table 3.2. The average procedure duration used in the calculation was found from both emergency and elective surgeries.

Specialty	**Average procedure duration (mins)**	**Total time in theatre (mins)**	**Session duration (mins)**	**Calculated number of operations per session**
Cardiac	331.6	363.6	420	1.2
CEPOD	119.6	151.6	420	2.8
Colorectal	203.1	235.1	210	0.9
ENT	123.6	155.6	210	1.4
General	168.2	200.2	420	2.1
Liver	253.2	285.2	420	1.5
Neurosurgery	184.4	216.4	420	1.9
Ophthalmology	70.6	102.6	210	2.1
Oral	197.1	229.1	210	0.9
Paeds ENT	68.2	100.2	210	2.1
Paeds General	119.6	151.6	210	1.4
Paeds Trauma	90.4	122.4	210	1.7
Renal	172.2	204.2	420	2.1
Scoliosis	90.8	122.8	420	3.4
Thoracic	182.7	214.7	210	1.0
Trauma	109.1	141.1	420	3.0
Urology	160.2	192.2	420	2.2
Vascular	142.8	174.8	210	1.2

Table 3.2: Calculated number of operations per session

Values from Table 3.2 suggest that the specialties that perform operations of a more urgent nature, namely CEPOD and Trauma, are able to perform the most number of operations per session, given the typical length of their surgical procedures. This is in agreement with the observed number of operations per session from the 2012/13 data in Figure 3.5. A comparison of the calculated number of operations with the observed number is shown in Figure 3.6. The calculations suggest that more operations can be performed per session for the Scoliosis specialty than were observed for 2012/13. Figure 3.5 shows that the Ophthalmology and Paediatric General specialties performed a high number of operations per session in 2012/13, however, this is not suggested by the results of the calculations. Higher observed numbers of operations per session could be a result of session overruns, causing more time to be used in the session than was allocated in the MSS. It is not possible to confirm this from the data available, however, discussions with hospital staff have described that sessions often overrun.

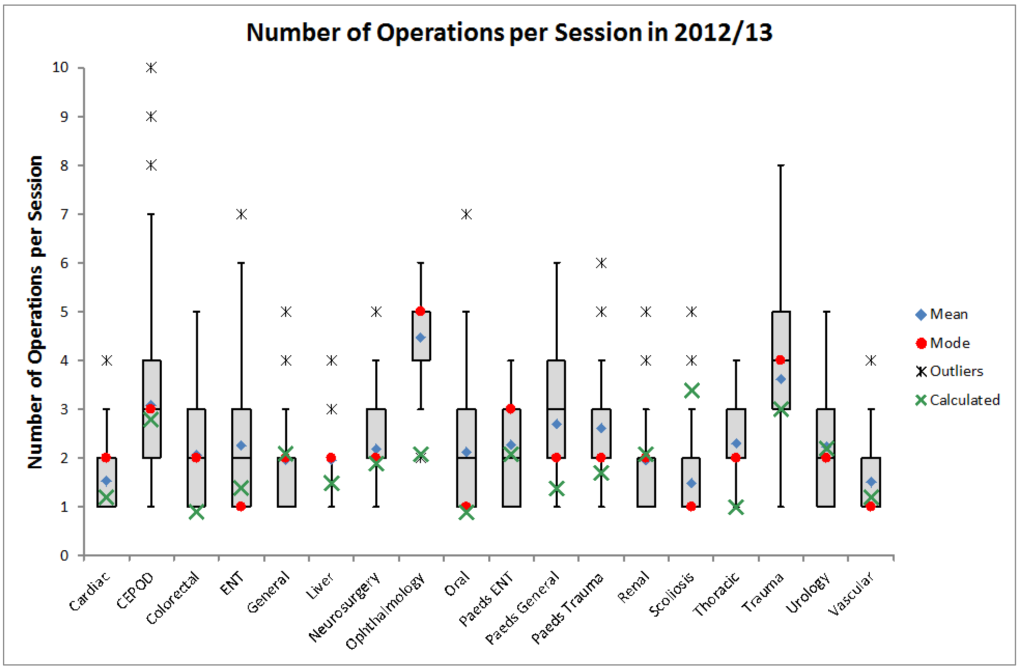


Figure 3.6: Number of observed and calculated operations per session in 2012/13

3.3.4 Theatres Used by Specialties

The main theatre suite in UHW comprises 14 operating theatres that are located along one corridor, allowing for a centralised point of contact for surgical staff and equipment. The theatres are numbered 0 to 14, however, the number 13 is omitted due to the superstitious connotations with the number. There are a number of dedicated theatres that certain specialties have sole use of which is reflected in the MSS (Figure 3.1). These theatres are:

- Theatre 0: Trauma
- Theatre 5: CEPOD
- Theatre 9: Thoracic
- Theatres 10 and 11: Cardiac
- Theatres 12 and 14: Neurosurgery

The remaining theatres can, in theory, be used by any other specialty for elective operations, however due to access to specialist equipment, similar specialties tend to prefer to be allocated to operating theatres so that they are in the same or adjacent theatres.

Of the 14 main theatres, only one theatre (Theatre 5) is used solely for emergency cases and is referred to as the CEPOD theatre. This theatre was first introduced after the 1990 review of the peri-operative care of surgical patients by the National Confidential Enquiry into Patient Outcome and Death (NCEPOD) [125]. The CEPOD theatre is a dedicated, staffed emergency operating theatre available 24 hours/day, 7 days/week. No elective patients are scheduled to have operations in the CEPOD theatre, however, if the need for an elective operation becomes more urgent whilst the patient is in hospital, then the patient may be put on the emergency list to have surgery sooner than planned.

The number and corresponding cumulative percentage of operations that were performed in the CEPOD theatre by each specialty in the year 2012/13 is given in Table 3.3. It can be seen that eight specialties account for over 90% of the operations carried out in this theatre. The remaining specialties use the emergency theatre very rarely.

Specialty	**Number of operations in the CEPOD theatre**	**Percentage of patients**	**Cumulative percentage**
General	1143	38.8	38.8
Neurosurgery	407	13.8	52.6
Paeds General	317	10.8	63.4
Oral	212	7.2	70.5
Vascular	212	7.2	77.7
Paeds Trauma	157	5.3	83.1
Trauma	123	4.2	87.2
ENT	91	3.1	90.3
Renal	72	2.4	92.8
Urology	52	1.8	94.5
Colorectal	43	1.5	96.0
Cardiac	33	1.1	97.1
Paeds ENT	24	0.8	97.9
Scoliosis	24	0.8	98.7
Ophthalmology	18	0.6	99.4
Liver	16	0.5	99.9
Thoracic	3	0.1	100.0
Total =	2947		

Table 3.3: Specialties that used the CEPOD theatre in 2012/13

Figure 3.7 shows the proportion of operations for each specialty that took place in each theatre for the year 2012/13. Recall from Section 3.3.2 that emergency operations are recorded on emergency theatre list (EM1, EM2 or EM3), not the physical theatre in which surgery took place. Shaded in blue are the theatres in which at least 90% of the operations took place for each specialty. The boxes with a red border indicate which theatre each specialty is actually assigned to in the current UHW MSS.

	EM 1	EM 2	EM 3	0	1	2	3	4	5	6	7	8	9	10	11	12	14
Cardiac													0.1	0.4	0.5		
CEPOD	0.5	0.2							0.3								
Colorectal											0.2	0.8					
ENT						1.0											
General					0.1		0.5				0.2	0.2					
Liver											0.6	0.4					
Neurosurgery																0.5	0.5
Ophthalmology						1.0											
Oral surgery						0.3		0.4					0.2				
Paeds ENT						1.0											
Paeds General								1.0									
Paeds Trauma					1.0												
Renal							0.1				0.9						
Scoliosis				0.7	0.3												
Thoracic													1.0				
Trauma				0.9	0.1												
Urology										1.0							
Vascular	0.2	0.1					0.5	0.1	0.2								

Figure 3.7: Theatres used by each specialty in 2012/13

As can be seen in Figure 3.7, most specialties have used more than one theatre in 2012/13, i.e. there is not a one-to-one mapping of specialty to theatre. The majority of operations for most specialties were performed in the theatre(s) allocated to them in the MSS. Specialties that do not follow this trend include:

- Colorectal – The majority of Colorectal operations were performed in Theatre 8, rather than Theatre 7. Colorectal is very closely related to the General surgical specialty that was mainly assigned to Theatre 8 in the current UHW MSS.

- Liver – The majority of Liver operations were carried out in Theatre 7, rather than Theatre 8. Liver is closely related to the Colorectal specialty that was assigned to Theatre 7 in the MSS.

- Renal – Not many operations have been performed in Theatre 8, perhaps due to nature of transplant surgery and the availability of organs may not have coincided with the scheduled time in Theatre 9.

- The majority of Scoliosis operations were carried out in the Trauma theatre (Theatre 0). Scoliosis is a paediatric sub-specialty of Trauma.

- Due to the urgent nature of Vascular operations, patients are often put on the emergency lists and most likely performed in the CEPOD theatre (Theatre 5).

3.4 Surgical Wards in UHW

3.4.1 Wards Used by Specialties

There are 17 physical wards in UHW that are used by surgical specialties. Some specialties are assigned to multiple wards, so for simplicity in the data analysis and model, these wards will be collated to form 'combined' wards. The combined wards that will be analysed and used in the model are listed in Table 3.4, and the specialties that are assigned to each ward are specified.

Ward	**Number of beds**	**Specialties using each ward**
Paediatric	28	Paeds ENT, Paeds General, Paeds Trauma
ENT/Oral	19	ENT, Opthalmology, Oral
Vascular	38	Vascular
Trauma	83	Trauma
Renal	20	Renal
General/Liver	76	General, Liver
Urology	19	Urology
Colorectal	20	Colorectal
Cardiothoracic	50	Cardiac, Thoracic
Neurosurgery	53	Neurosurgery
Critical Care	27	General, Neurosurgery, Trauma, Vascular

Table 3.4: Surgical wards used in data analysis

The CCU, where patients receive specialist care, is also analysed and included in the model. More analysis for the CCU is given in Section 3.4.2. Other high dependency wards that are specialty specific, such as the Neurosurgery High Dependency Unit and the Cardiac Intensive Care Unit, are incorporated into the Neurosurgery and Cardiothoracic combined wards respectively. This is because these high dependency wards are managed by the specialties themselves.

The proportion of patients that were in each of the combined wards in 2012/13 is given in Figure 3.8. The wards in which at least 80% of the patients from each specialty were in in 2012/13 are shaded in blue. The wards with a red border refer to the wards that each specialty has been assigned to.

	Paediatric	ENT/ Oral	Vascular	Trauma	Renal	General/ Liver	Urology	Colorectal	Cardio- thoracic	Neuro surgery	Critical Care
Cardiac									1.0		
CEPOD	0.2		0.1			0.4				0.2	0.1
Colorectal						0.2		0.8			
ENT		0.4				0.2	0.4				
General						0.6		0.1			0.3
Liver						0.9					0.1
Neurosurgery										0.9	0.1
Ophthalmology	1.0										
Oral surgery	0.2	0.4				0.1	0.3				
Paeds ENT	0.9					0.1					
Paeds General	1.0										
Paeds Trauma	0.9					0.1					
Renal					0.7	0.3					
Scoliosis	1.0										
Thoracic									1.0		
Trauma			0.1	0.4		0.5					
Urology						0.6	0.1				0.1
Vascular			0.7			0.2					0.1

Figure 3.8: Proportions of patients on each ward in 2012/13

It can be seen from Figure 3.8 the for the majority of specialties, patients are on a ward that is related to their specialty. There are some exceptions, including the ENT and Oral specialties for which some of their patients were on the General/Liver or Urology wards in 2012/13. All of the patients in the Ophthalmology specialty, for which we have data, are sent to the Paediatric ward. From discussions with hospital staff, this was explained by the fact that the majority of adult patients are able to leave hospital on the same day after an ophthalmic operation, and hence have a LoS of zero days. Children are often admitted after an opthalmic operation in order to monitor them over night, and so must be sent to the Paediatric ward. Around 10% of the children from the Paediatric ENT and Trauma specialties were recorded as being on the adult General/Liver ward. It is unclear why this occurred, since there is a strict rule that child patients must be on the Paediatric ward. The Trauma, Urology and Vascular specialties also have patients outlying on the General/Liver ward.

The data obtained from UHW refers to the first destination of each patient after surgery. It does not include any subsequent ward that a patient may have been moved to during their post-operative care. This may explain why there appears to be some patients that were outlying on wards that were not assigned to their surgical specialty.

Overall, the issue of outlying patients on a different specialty's ward can be seen from the data to have occurred quite often in 2012/13. In particular, there were a lot of outliers on the General/Liver ward. This illustrates the high demand for beds in UHW as reported in the media and explained by hospital staff. It is hospital policy that a bed on a ward is never left empty if demand for a bed exists and should be used by a patient if medically safe. The problem of outlying patients is also a reinforcing issue, since if beds are taken on a specialty's ward by outlying patients, then the patients who should be on this ward will be forced to outlie on another ward, hence exacerbating the situation. It is also possible for patients to move beds if a space on their specialty's ward becomes available for them to continue their post-operative recovery on the correct ward, however, this was not captured in the data available.

3.4.2 Critical Care Unit

The CCU is available to every medical and surgical specialty when a patient requires the very highest level of care. Patients are often cared for on a one-to-one basis with nurses and specialist life-saving equipment, resulting in very high running costs for this ward. There are 27 beds in the CCU and are given as a priority to emergency patients over elective surgical patients. Patients in the CCU are categorised into two levels based on the level of care they require: level 3 patients require the most care and are intubated, whereas level 2 patients are not intubated. There are typically 17 beds available for level 3 and 10 for level 2, however, the number of beds for each level can be altered for the demand.

A separate dataset was provided by UHW on the activity of surgical patients in the CCU. The number of CCU admissions from surgical specialties in 2012/13 is given in Table 3.5.

Specialty	Number of CCU admissions	Cumulative percentage
General	327	62.9
Neurosurgery	80	78.3
Oral	34	84.8
Urology	30	90.6
Vascular	20	94.4
Trauma	18	97.9
Cardiothoracic	6	99.0
ENT	4	99.8
Ophthalmology	1	100.0
Total =	520	

Table 3.5: Number surgical admissions to the CCU in 2012/13

Only six specialties account for over 95% of the CCU admissions from surgical specialties, with General surgery accounting for the majority of the admissions in 2012/13. The daily bed count of surgical patients in the CCU throughout 2012/13 is shown in Figure 3.9. The bed count fluctuates around a mean of 7.1 throughout the year. Surgical patients accounted for between 7.4% and 48.1% of the admissions to the CCU on any one day in 2013.

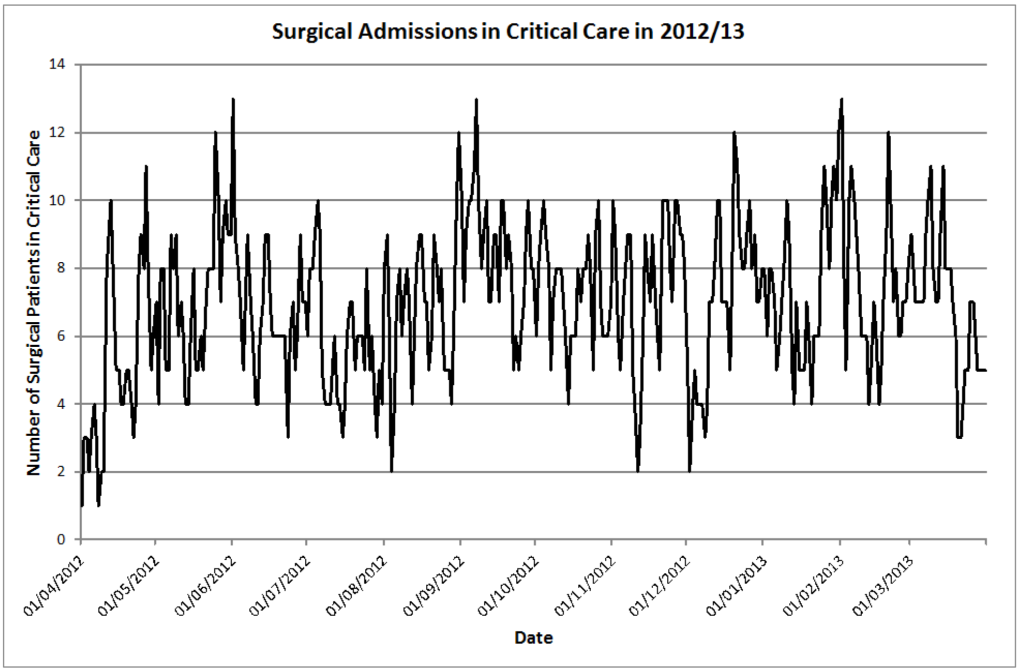


Figure 3.9: Daily bed count of surgical admissions in the CCU in 2012/13

The LoS of patients in the CCU in 2012/13 is shown in Figure 3.10. The data exhibits typical characteristics of LoS distributions, being skewed to the right with the majority of patients having a shorter LoS and a few patients having a very long LoS. CCU beds experience a very high demand from both medical and surgical specialties for their most ill patients, so as soon as a patient is well enough to leave the CCU, they will be sent to their specialty's ward to continue their recovery. This explains why the majority of patients are in the CCU for up to two days. The mortality rate for the CCU in 2012/13 was 10%.

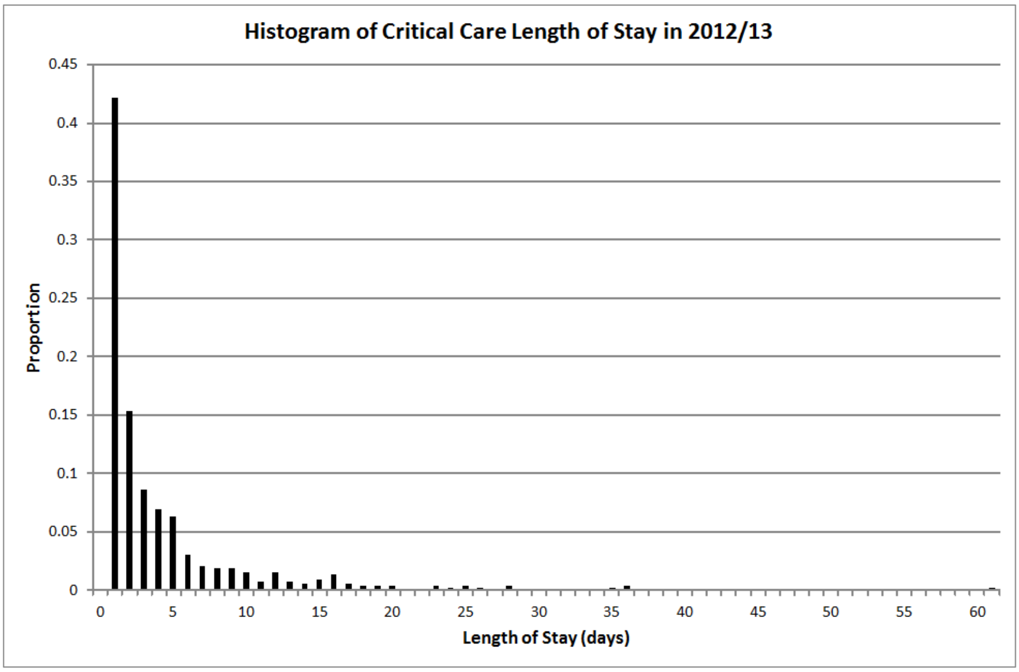


Figure 3.10: Length of stay in the CCU in 2012/13

3.5 Length of Stay Data Analysis

Between a patient's admission to and discharge from hospital, there are many periods of time that are of interest to this study. A hospital spell is defined to be between each admission to and discharge from hospital. A spell may be spilt into one or more episode; an episode is the time spent under a particular specialty in the hospital. This can either be a medical or surgical specialty. The episode of interest in this study is the episode when patients are assigned to a surgical specialty and have an operation. This surgical episode can be split into two defined time periods; 'pre-op' is the time spent under a surgical specialty before the operation has taken place, and 'post-op' is the time spent under the same surgical specialty after the operation. These periods of time are represented in Figure 3.11, and calculated in the master dataset as fractions of days, then rounded to the nearest whole day.

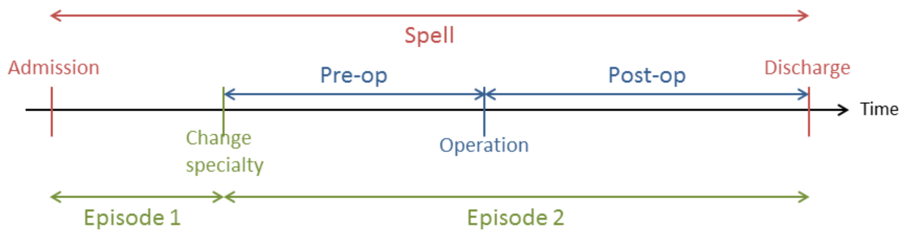


Figure 3.11: Spells and episodes in hospital

Patient LoS can be used in analysis for both strategic and operational purposes within the hospital. The most common statistic that is reported on LoS is the average LoS, however, due to the high variability of LoS data, this perhaps is not the best estimate. LoS distributions can vary across different patient demographics and the environment in which the patient is treated, and will be discussed in more detail in the following sections.

3.5.1 Pre-Operative Length of Stay

Distributions of the pre-operative LoS for each surgical specialty in 2012/13 are shown as box and and whisker plots in Figure 3.12.

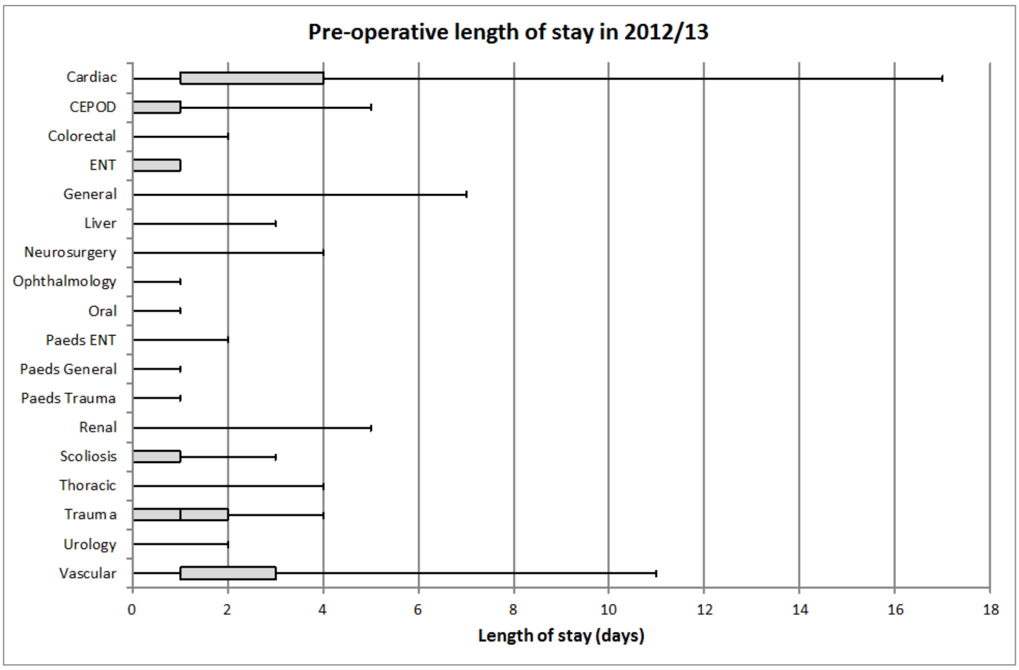


Figure 3.12: Pre-operative length of stay for all specialties in 2012/13

The pre-operative LoS varies between specialties, with the majority of specialties having pre-operative LoSs of less than 6 days in 2012/13. Patients had a pre-operative LoS longer than 6 days in three specialties: Cardiac, General and Vascular surgery. All specialties had a median pre-operative LoS of 1 day, except for the CEPOD, Paediatric Trauma and Scoliosis specialties that had a median pre-operative LoS of zero days, i.e. they were admitted to hospital on the day of surgery.

3.5.2 Post-Operative Length of Stay

Distributions of the post-operative LoS for each surgical specialty in 2012/13 are shown as box and whisker plots in Figure 3.13.

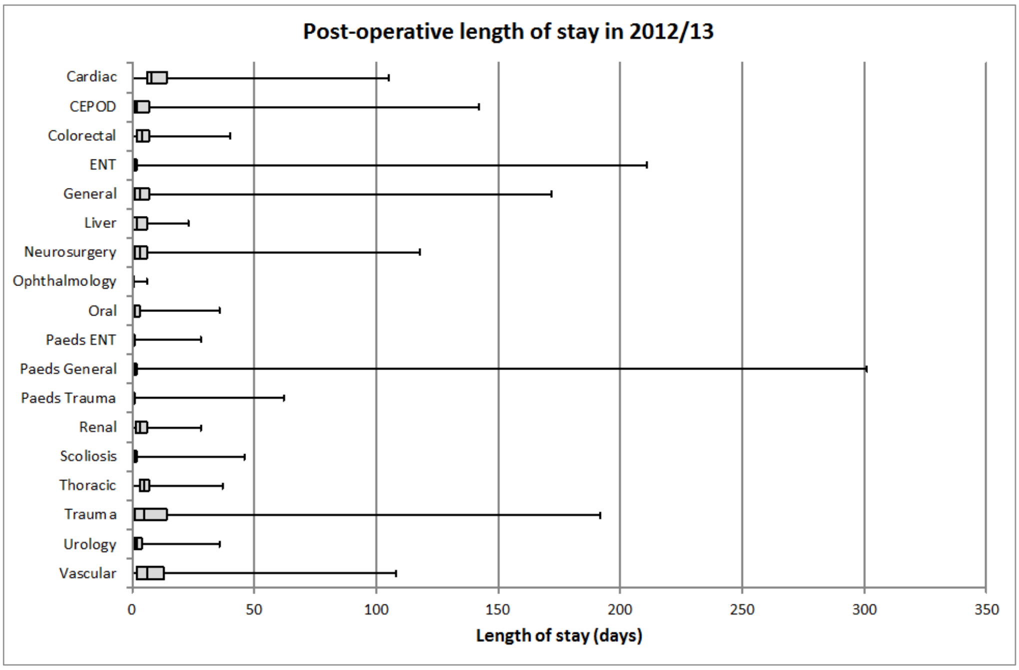


Figure 3.13: Post-operative length of stay for all specialties in 2012/13

As can be seen from Figure 3.13, the post-operative LoS differs greatly between specialties. Extreme outliers have been removed from the data as described in Section 3.2.1, however, it is still evident that large values of LoS that are a long way away from the majority of the data exists in the dataset. The post-operative LoS distributions for all specialties show the typical characteristic of LoS distributions of being skewed to the right, as indicated by the large 'whiskers' to the right of the interquartile range boxes.

Cardiac, Trauma, and Vascular specialties all have the highest average post-operative LoS, whereas Ophthalmology, Paediatric ENT and Paediatric Trauma all have the lowest post-operative LoS. It is interesting to note that in the majority of cases, the Paediatric specialties have a shorter post-operative LoS than their adult counterparts.

In order to correctly model the post-operative LoS of patients, it may be appropriate to fit statistical distributions to the data. LoS distributions are typically skewed to the right with a long tail towards high values of LoS, and there are often outliers in the data that are vastly higher than the majority of LoS values [104]. Lognormal, Weibull and Gamma distributions are commonly fitted to LoS data. The software package Stat::Fit [85] was used to find any statistical distribution that would be suitable to represent the post-operative LoS data. For each surgical specialty, the Anderson-Darling and Kolmogorov-Smirnov goodness of fit tests were performed to test whether any of the Lognormal, Weibull or Gamma distributions could be fitted to the empirical data. For all surgical specialties, the null hypothesis of each goodness of fit test was rejected at the 5% significance level. Thus it can be concluded that the post-operative LoS data cannot be modelled by using one of these distributions typically used for LoS.

As discovered in Section 3.4.1, not all patients are in a bed on the assigned ward for their specialty. It was reported in [12] that being on a ward that is not the intended specialty ward may ‘adversely affect’ LoS and quality of care. It is of interest to see whether being on a ward that is not their assigned ward in UHW affected the post-operative LoS. For the sake of this investigation, the intended wards for the specialty will be called the ‘correct’ wards, and the other wards that patients are on will be called the ‘wrong’ wards. Both the Kolmogorov-Smirnov and Sharipo-Wilk tests concluded that the LoS data on either the correct or wrong wards were not Normally distributed at the 5% significance level. A series of Mann-Whitney tests were therefore used to compare the post-operative LoS on the correct and wrong wards for all specialties. Conclusions of these tests at the overall 5% significance level are given in Table 3.6.

			Median post-op LoS (days)	
Specialty	**p-value**	**Conclusion**	**Correct ward**	**Wrong ward**
Cardiac	0.253	No difference in LoS	-	-
Colorectal	0.640	No difference in LoS	-	-
ENT	0.058	No difference in LoS	-	-
General	0.004	No difference in LoS	-	-
Liver	0.097	No difference in LoS	-	-
Neurosurgery	0.054	No difference in LoS	-	-
Ophthalmology	0.745	No difference in LoS	-	-
Oral	0.088	No difference in LoS	-	-
Paeds ENT	0.398	No difference in LoS	-	-
Paeds General	0.111	No difference in LoS	-	-
Paeds Trauma	<0.0005	Different LoS	0	2
Renal	<0.0005	Different LoS	5	2
Scoliosis	0.227	No difference in LoS	-	-
Thoracic	0.122	No difference in LoS	-	-
Trauma	<0.0005	Different LoS	7	1
Urology	0.019	No difference in LoS	-	-
Vascular	0.345	No difference in LoS	-	-

Table 3.6: Tests for differences in post-operative length of stay in different wards for each specialty

The results of the Mann-Whitney tests indicate that, for most specialties, the median post-operative LoS is the same whether the patients are on a correct or wrong ward. The specialties for which the tests were not able to conclude that the median LoS is the same on both types of ward are Paediatric Trauma, Renal and Trauma. It is not consistent among these specialties that the LoS is longer on the wrong ward.

3.6 Problems Experienced in UHW

3.6.1 Cancellations

Patients are entered into the TheatreMan database when their operation has been booked and assigned to an operating theatre session. According to administrators who book patients for operations, patients are added to theatre lists typically one day before surgery. This booking of an operation is different to when a patient receives a letter from the hospital advising them of their surgery date. Therefore the data available in TheatreMan is only truly representative of the patients who have already entered hospital prior to their operation. If an operation is cancelled a month or a week prior to the date of surgery, then that information is not captured in the TheatreMan database.

In the year 2012/13, just over 18% of operations were cancelled after being assigned to an operating theatre session; this corresponds to over 2500 operations. Figure 3.14 shows the number of operations that were performed and cancelled for each specialty in 2012/13. It can be seen in Figure 3.14 that the Trauma specialty has the highest number of cancellations, however, this is not surprising due to the urgent nature of the procedures for this specialty. It is quite common for Trauma operations to be cancelled when a more urgent patient arrives in hospital who requires surgery sooner than the scheduled patients. It was not possible to find from the data whether cancelled operations were rescheduled and performed at a later date.

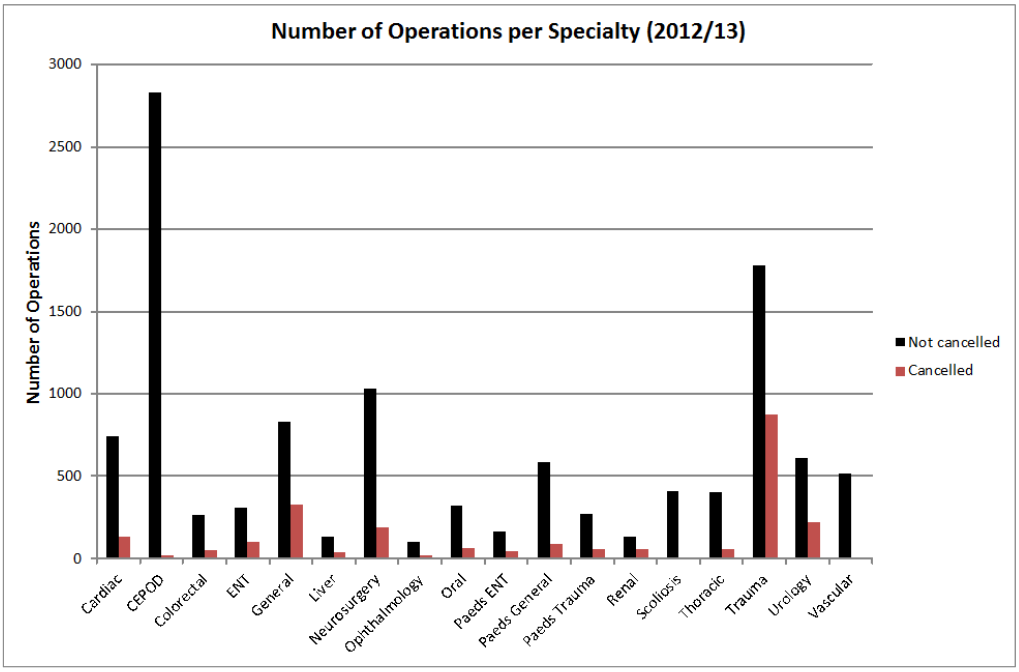


Figure 3.14: Number of performed and cancelled operations in UHW in 2012/13

In the year 2012/13, only 2.9% of the operations that were cancelled after being scheduled onto a theatre list were cancelled before the day of surgery, whereas 93.4% were cancelled on the day of surgery. The remaining 3.7% of cancelled surgeries were recorded as being cancelled after the day of surgery. It is assumed that these operations were indeed cancelled and did not take place when scheduled, but were recorded as cancelled at a later date after a delay by the admin staff. It is therefore not known exactly when it was decided to cancel these operations.

There are three broad categories of reasons why operations may be cancelled, as identified by the NHS: hospital non-clinical, hospital clinical and patient reasons [128]. Cancellation data available for the year 2012/13 contains 22 distinct reasons as to why operations were cancelled and has been re-classified into the three categories as defined by the NHS. In 2012/13, 54.2% of all cancellations were attributable to hospital non-clinical reasons, such as equipment and staff availability, list overrun, and unavailable beds, 26.3% of cancellations were due to hospital clinical reasons, such as the operation became unnecessary or the patient was deemed unfit for surgery, and 19.5% of cancellations were due to patient reasons, such as the patient did not arrive on time for surgery.

A detailed summary of the reasons operations were cancelled in 2012/13 for each specialty is given in Table 3.7. The proportion of non-clinical cancellations that were attributable to a lack of bed availability on the wards is also reported, and the specialties that have a majority of the non-hospital cancellations caused by a lack of beds are highlighted in red. Unfortunately, there were no data available for the cancellations in the Scoliosis and Vascular specialties.

Specialty	Percentage of cancellation types			Percentage of non-clinical cancelled due to no beds
	Hospital non-clinical	Hospital clinical	Patient reasons	
Cardiac	56.0	26.1	17.9	26.7
CEPOD	17.4	65.2	17.4	0.0
Colorectal	72.5	23.5	3.9	73.0
ENT	49.5	27.6	22.9	71.2
General	63.5	25.2	11.2	64.1
Liver	71.4	19.0	9.5	23.3
Neurosurgery	63.2	18.1	18.7	70.5
Ophthalmology	31.8	40.9	27.3	100.0
Oral	60.6	19.7	19.7	77.5
Paeds ENT	53.5	23.3	23.3	87.0
Paeds General	22.2	53.3	24.4	40.0
Paeds Trauma	66.1	22.0	11.9	79.5
Renal	61.4	26.3	12.3	88.6
Scoliosis	-	-	-	-
Thoracic	57.9	28.1	14.0	18.2
Trauma	41.6	29.0	29.5	14.6
Urology	70.4	21.1	8.5	73.2
Vascular	-	-	-	-

Table 3.7: Reasons for cancelled operations in UHW in 2012/13

The high percentage of non-clinical cancellations that were due to a lack of beds can perhaps be explained by the fact that some specialties share wards with other specialties. The Colorectal, General and Urology specialties are all allowed to admit patients onto each others wards, ENT, Oral and Ophthalmology specialties share the same ward, and all of the paediatric specialties send their patients to the specific Paediatric ward.

3.6.2 Outliers on Wards

Figure 3.15 shows the percentages of patients that were in a bed on the assigned ward for their specialty, or related or unrelated wards in 2012/13. For example, patients in the Cardiac specialty should be on the Cardiac ward, however, if there are no beds available on this ward they may be put on related wards such as the Cardiology or Thoracic surgery wards, since these wards will have the correct equipment and nursing staff with the required skills for Cardiac surgery patients.

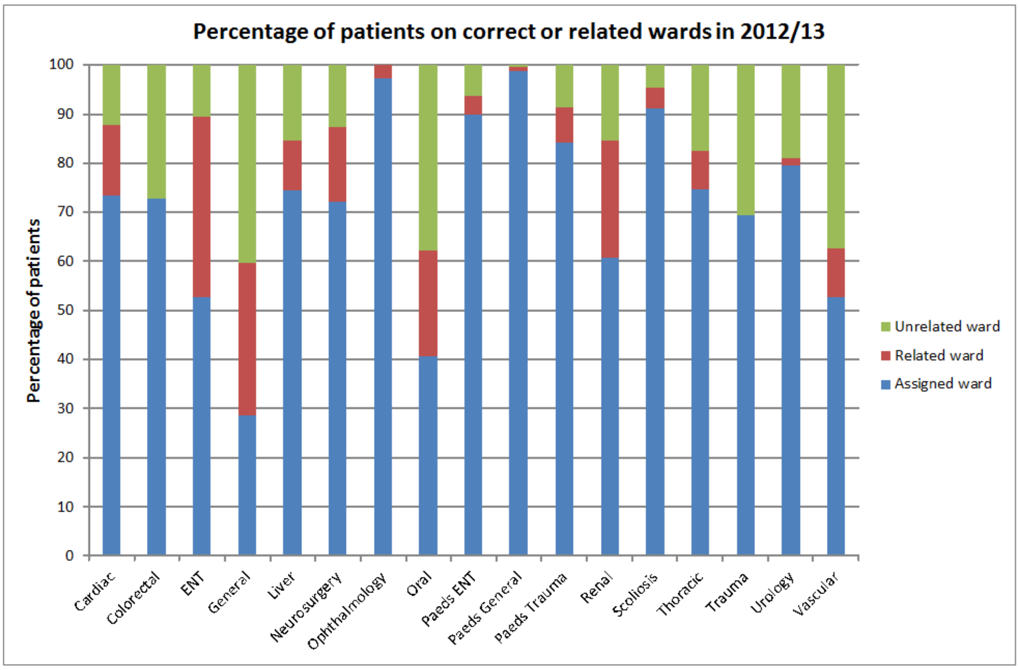


Figure 3.15: Percentage of surgical patients on the assigned or related wards in UHW in 2012/13

As can be seen in Figure 3.15, patients are on the assigned specialty ward(s) for the majority of the surgical specialties. General and Oral surgery are the two exceptions. If patients are not on the assigned ward for their specialty, then it would appear from the data that the patients are on an unrelated ward to their specialty for the majority of specialties. This is concerning due to the issues raised in an audit report [12] that may result from patients not being treated on their specialty's ward. Specialties for which the majority of patients are on a related ward include Cardiac and Neurosurgery. This is as expected due to the highly specialised equipment and nursing skills required for these two specialties.

3.7 Summary

This chapter has been used to introduce the case study hospital and to investigate data provided by CaV UHB relating to the operating theatres and surgical inpatient wards in UHW, Cardiff. The analysis was performed in order to gain an understanding of how the operating theatres are currently utilised by the surgical specialties, and the post-operative demand for beds on the wards.

The data was provided as two separate datasets, so initial processing of the data, as described in Section 3.2, was required in order for the whole patient pathway from admission, through the operating theatre, to discharge could be captured and analysed. Outliers were removed where appropriate, however, it is clear from the above analysis that great variation in the data exists. In particular, the duration of operations and post-operative LoS exhibit great variation with respect to specialty.

The demand for operations in UHW was analysed in Section 3.3, with particular emphasis on the analysis of the occurrence of emergency surgical patients, how the operating theatres are utilised by the specialties, and how many patients are operated on in a typical operating theatre session. The latter two aspects can be used to construct scheduling rules and the demand for operations as inputs to any subsequently developed models.

The subsequent demand for beds on the surgical wards was then analysed in Section 3.4, with additional analysis on pre- and post-operative LoS for each specialty provided in Section 3.5. The post-operative LoS was not found to differ significantly for most specialties depending on whether patients were on their specialty's assigned ward, or a similar specialty's ward. The CCU was found to experience high demand from surgical patients, especially from the General surgery specialty. However, the majority of patients stay in the CCU for a maximum of two days, possibly alleviating an accumulative demand for beds.

Finally, following discussions with hospital managers, problems that are currently experienced in the hospital that are associated with the operating theatres are investigated in Section 3.6. A high proportion (18%) of scheduled operations were found to have been cancelled within two days of the operation date. This is clearly undesirable for both hospital planning purposes and for the quality of care of patients. Particularly relevant to this research, is the fact that over half of these cancelled operations were cancelled due to a lack of beds available on the wards for post-operative recovery. A contributing factor to a lack of beds available on wards, was found to be the existence of outlying patients – patients who are not on the correct ward for their specialty.

Insights gained from the data analysis presented in this chapter will be used to inform the modelling approaches developed in subsequent chapters. The aim of these subsequent chapters is to address the problems of cancelled operations and outlying patients, together with the research aims outlined in Chapter 1, through the construction of operating theatre schedules using operational research techniques.

Chapter 4

Deterministic Optimisation of the MSS

This chapter discusses the development of a set partitioning based optimisation model for the construction of the MSS. A brief review of the literature on the set partitioning problem is given. The proposed model, which includes constraints for both the operating theatres and bed demand, is then developed and validated.

4.1 The Set Partitioning Problem: an Overview

The set partitioning problem (SPP) can be formulated as a binary integer programming optimisation model that determines how items in a set can be partitioned into smaller subsets such that all items in the larger set are contained in exactly one subset. This model has been used successfully for the modelling of scheduling and rostering problems [138], and also vehicle routing problems [15]. In general, the SPP is NP-hard, however, in some cases exact approaches can be used to determine globally optimal solutions [116]. Indeed, with the increasing speed of computer hardware, larger problem instances containing hundreds of millions of variables and hundreds of constraints can sometimes be solved using exact methods [135].

4.1.1 Set Partitioning Problem Formulation

The SPP model stems from a set theoretical approach, so the initial set theory problem is introduced. Let $M = \{1, ..., m\}$ contain elements that need to be partitioned, let $\mathcal{S} = \{S_1, S_2, ..., S_n\}$ for $S_j \subseteq M$ contain n subsets of M, and let $P \subseteq \{1, ..., n\}$.

The set P defines a partition of M if and only if

i)

$$\bigcup_{j \in P} S_j = M$$

i.e. the union of all subsets in partition P form the original set M.

ii)

$$S_j \cap S_k = \emptyset \quad \forall \; j, k \in P, \quad j \neq k$$

i.e. each element of M occurs in exactly one subset S_j of M.

Let c_j be the cost associated with subset S_j, and let $\sum_{j \in P} c_j$ be the cost of the partition P. The SPP can be defined as finding the minimum cost partition, P^*, of M, given $\mathcal{S}$. Balas and Padberg [13] give a general definition of the SPP as follows:

"Given a finite set M, a constraint set defining a family F of 'acceptable' subsets of M, and a cost associated with each member of F; find a minimum-cost collection of members of F which is a partition of M."

The mathematical formulation of the SPP model will now be introduced. In an SPP model, there are m elements that need to be partitioned, and n possible subsets of $M = \{1, ..., m\}$. Let the matrix $A = (a_{ij})$ be defined as

$$a_{ij} = \begin{cases} 1 & \text{if element } i \text{ is included in subset } S_j \quad \forall \; i = 1, \ldots, m, \;\; j = 1, \ldots, n \\ 0 & \text{otherwise.} \end{cases}$$

The decision variables, $x_j, j = 1, \ldots, n$, for the SPP are defined as

$$x_j = \begin{cases} 1 & \text{if subset } S_j \text{ is selected} \quad \forall \; j = 1, \ldots, n \\ 0 & \text{otherwise} \end{cases}$$

with each decision variable having an associated cost c_j.

The SPP is thus the problem of choosing subsets at minimal cost, such that all elements of the original set are partitioned into exactly one subset. The mathematical formulation is as follows:

$$\min \quad \sum_{j=1}^{n} c_j x_j \tag{4.1}$$

$$\text{s.t.} \quad \sum_{j=1}^{n} a_{ij} x_j = 1 \quad \forall \quad i = 1, ..., m \tag{4.2}$$

$$x_j \in \{0, 1\} \quad \forall \quad j = 1, ..., n \tag{4.3}$$

In corresponding matrix notation, the formulation of the SPP is:

$$\begin{aligned} \min \quad & c^T x \\ \text{s.t.} \quad & Ax = e \\ & x \in \{0,1\}^n \end{aligned} \tag{4.4}$$

where A is an $m \times n$ matrix of zeros and ones, c is an arbitrary vector of costs, and $e = (1, 1, 1, ..., 1)^T$ is an m-vector. The rows of the matrix A are associated with elements of the set $M = \{1, ..., m\}$ to be partitioned. Each column of A therefore represents each subset, S_j, of M, for $j = 1, ..., n$. The j^{th} column of A, a_j, has elements

$$a_{ij} = \begin{cases} 1 & \text{if column } j \text{ covers row } i \quad \text{i.e. if subset } S_j \text{ contains element } i \in M \\ 0 & \text{otherwise.} \end{cases}$$

The binary decision variables, x_j, $j = 1, \ldots, n$, can also be thought of as the probability that the j^{th} column is included in a solution [80]. This can be particularly relevant when interpreting the linear programming relaxation solution to this integer programming problem.

4.1.2 Problems Related to the SPP

There are two other optimisation problems that are closely related to the SPP and are also NP-hard [83]: the set covering and set packing problems. The set covering problem is defined as choosing subsets at minimal cost, such that all elements of the original set appear in *at least* one subset. The mathematical formulation of the set covering problem is:

$$\min \quad \sum_{j=1}^{n} c_j x_j \tag{4.5}$$

$$\text{s.t.} \quad \sum_{j=1}^{n} a_{ij} x_j \geq 1 \quad \forall \quad i = 1, \ldots, m \tag{4.6}$$

$$x_j \in \{0,1\} \quad \forall \quad j = 1, \ldots, n \tag{4.7}$$

The set packing problem is defined as choosing the *maximum number of subsets*, such that each subset is disjoint (i.e. no two subsets share an element). The mathematical formulation of the set packing problem is:

$$\max \sum_{j=1}^{n} c_j x_j \tag{4.8}$$

$$\text{s.t.} \quad \sum_{j=1}^{n} a_{ij} x_j \leq 1 \quad \forall \quad i = 1, \ldots, m \tag{4.9}$$

$$x_j \in \{0, 1\} \quad \forall \quad j = 1, \ldots, n \tag{4.10}$$

In both models, a_{ij}, c_j and x_j, for $i = 1, \ldots, m$ and $j = 1, \ldots, n$, have the same definition as in the SPP formulation in Model 4.4.

It can be seen that the three models differ in the constraints 4.2, 4.6 and 4.9, and that set packing is a maximisation problem whereas the other two models are minimisation problems.

Set packing is a special case of set partitioning since it is more tightly constrained than the set partitioning model. This corresponds to the first set partitioning condition being relaxed, i.e. not all elements in M have to be contained in the subsets S_j in the partition P. The set covering model corresponds to the second set partitioning condition being relaxed, i.e. elements of M can be in more than one subsets S_j in the partition P.

There are some models in the literature that deviate from the pure SPP formulation given in Model 4.4. These generalised SPP models are often found in crew rostering applications [135] where the right-hand-side vector, e, need not be a unit vector, and some constraints need not be equalities. It has been shown that there are benefits to constraints of this type in relation to finding integer basic feasible solutions in the linear programming relaxation of the SPP model [137].

4.1.3 Variable Generation

As we have seen, the decision variables in the SPP model are binary and indicate whether or not the corresponding subsets are included in the solution. There can often be hundreds of thousands of variables in an SPP model due to the combinatorial nature of the subsets. If all possible combinations of subsets are considered, the number of variables in the model could reach into the billions [94]. The number of decision variables also corresponds to the number of columns in the A matrix, and is hence referred to as column, variable or matrix generation in the literature. One great advantage of formulating scheduling problems as an SPP is that complex scheduling rules can be implicitly built into the process of generating the various subsets, thereby reducing the number that need to be considered. For example, over 30 rules and conditions regarding desirable flight crew rosters were followed when generating feasible crew schedules [135].

There exist many techniques in the literature for generation of the subsets. Traditionally, there are two general approaches [94]. These are enumeration, in which *all* subsets are considered, and column generation, which is an iterative process of generating *some* feasible subsets and solving the associated SPP model.

Enumeration

Enumeration is the technique of systematically generating all possible combinations of subsets of a larger set. An enumeration process can be used to generate all feasible subsets for the SPP, whilst taking into account certain rules that prevent some (undesirable) subsets from being generated.

The rows of the A matrix need to be partitioned into subsets in the SPP, and hence a large number of subsets can be generated. Marsten [115] showed that there are advantages to be gained by using enumeration in an algorithm for solving SPPs, namely the realisation of optimal solutions. Once an optimal solution has been found, selective exploration of the enumeration tree may then be used to obtain a collection of near optimal integer solutions. This could be desirable if computation times are limited or if the decision maker requires a selection of 'good' solutions to choose from. Geoffrion [86] discusses the advantages of excluding certain solutions in the enumeration tree from further consideration. Implicit enumeration techniques are also discussed by Michaud [119].

Application specific algorithms have been used to enumerate all feasible subsets. An enumeration algorithm is demonstrated with a simple numerical example by Garfinkel and Nemhauser [84]. Ryan [135] describes an enumeration process used to generate all possible lines of work for an airline crew schedule. A skeleton line of work is first constructed for each crew member in turn, then all feasible legal and desirable sequences of trips are added to the skeleton, obeying implicit rules and conditions, until no further additions are possible. Enumeration is also used to generate all feasible train routes by Lusby et al. [110].

Column Generation

Column generation is an iterative process of choosing a subset of rows, the generation of feasible columns for these rows, and the corresponding optimal solution of these set partitioning subproblems. The process repeats until the solution to recent subproblems has not improved the cost function value. In column generation, the original SPP to be optimised is known as the master problem, but not all columns in the A matrix may be known explicitly, so is restricted to form the restricted master problem to contain fewer columns.

Wilhelm [160] gives a review of column generation techniques used in integer programming, and Hoffman and Padbeg [94] discuss column generation for feasible aircrew schedules. An alternative technique used in column generation for the A matrix uses graph theory to generate columns based on shortest path calculations [103].

Preprocessing

Preprocessing of the data for an SPP is often used to make the optimisation easier and quicker to perform and is used particularly for large problem instances, as discussed by Chu and Beasley [53]. Often called 'reduction', rows and/or columns of the A matrix are deleted in order to reduce its dimensions. This corresponds to deleting constraints and variables respectively in the model. Ryan [135] shows how the number of variables can be reduced using filtering techniques, and an approach that reduces the number of variables whilst minimising the potential for fractional solutions is given by Ryan and Falkner [137].

4.1.4 Cost of Variables

The cost vector c that is used as the coefficient in the objective function in the SPP model must be chosen to reflect the relative 'cost' of each subset S_j. There is no common measure which is used to determine the costs of the subsets in a set partitioning optimisation model since it is application dependent. For example, the costs used for a vehicle routing application could be the distance of each route [138], while the costs used in airline crew rostering might reflect the interests of both management in terms of minimising the number of crews needed, and crew members in terms of their time off between long trips [135].

4.1.5 Solution Methods

Depending on the application, exact methods or heuristics can be used to solve SPPs. Branch-and-bound and branch-and-cut algorithms can be used to find optimal solutions, and a variety of heuristics, including tabu search and genetic algorithms, are used to find sufficiently good solutions. It has also been reported that integer programming is a method that is likely to provide integer solutions quickly, as long as the SPPs to be solved are small [116]. Gershkoff [88] shows that solving many small subproblems to optimality can be more successful in finding integer solutions than solving a single large SPP.

The branch-and-bound algorithm uses systematic enumeration and a structured search of the space of all feasible solutions. The problem is first solved without the integer constraint on the decision variables, then large subsets of the solution space are discarded by using upper or lower bounds of the objective function. The technique is used in discrete and combinatorial optimisation problems [102], and was first used in the solution of SPPs by Marsten [115] whose algorithm used linear programming to calculate the lower bounds. Albers [7] and Ryan [135] have also used the branch-and-bound approach, and have demonstrated the technique on large problem instances of thousands of variables and hundreds of constraints. Lagrangian relaxation is used to provide the bounds in their case [79].

The branch-and-cut algorithm is a hybrid of the branch-and-bound algorithm and cutting plane methods. Once a non-integer optimal solution has been found, a cutting plane algorithm is used to find additional constraints for the linear program; the branch-and-bound algorithm is then employed. Balas and Padberg [13] discuss a variety of branch-and-cut based algorithms for the SPP. Branch-and-cut algorithms have been used as solution techniques when set partitioning has been applied to airline crew scheduling [94] and the vehicle routing problem [14].

As discussed, there are many examples in the literature where algorithms are used to solve the SPP to optimality, despite the problem being NP-hard in general. Conversely, in many real-life situations there is no need to achieve the optimal solution, for example in air crew scheduling, where just a ‘good’ solution is required. For these situations, heuristics that find sufficiently good solutions in a short amount of time have been developed. Heuristics are also used to obtain approximate solutions when set partitioning models are considered too large to solve exactly. Ryan and Falkner [137] have imposed additional structure on the set partitioning model in order to find good solutions quickly. Linderoth et al.

[107] also developed a heuristic for solving large SPPs applied to crew scheduling and vehicle routing problems, and exploit the power of parallelism to obtain good solutions. Lee et al. [105] used a heuristic approach with tabu search when applying the SPP to the vehicle fleet mix problem.

In recent literature, evolutionary algorithms have been used for the solution of the SPP. Levine's algorithm [106], based on parallel subproblems of the main SPP, was able to regularly find the optimal integer solution to problems with a few thousand variables. Chu and Beasley [53] present a genetic algorithm that takes a large number of set partitioning constraints into consideration. Their heuristic includes separate fitness and 'unfitness' scores and is able to be generalised and applied to any highly constrained problem. Optimal or near-optimal solutions are reported to be found very quickly using this heuristic.

4.1.6 Applications of the SPP

As mentioned, a major area where set partitioning is used is airline crew scheduling, where exactly one flight crew must be assigned to each flight [147]. Personnel costs are the largest cost faced by airlines, so it is important to schedule the flight crews appropriately. This application is popular in the literature as the problem instances are very large, often with thousands of variables and hundreds of constraints. There are two stages to solving the crew scheduling optimisation problem: the generation of feasible tours of duties (that form the subsets), and the optimisation of the SPP [94]. A column of the A matrix is created for every feasible tour of duty for each crew and an associated cost is then assigned. An optimal schedule is then selected so that every flight is assigned a crew, whilst the cost is minimised.

The vehicle routing problem has also been a popular area of SPPs in the literature. Balinski and Quant [15] first proposed a set partitioning formulation of the vehicle routing problem as an alternative to previous methods which used heuristics to find an approximate solution. In the SPP model, each column of the A matrix represents a feasible route for each vehicle with an associated cost. Again, this application has very large problem instances depending on the number of vehicles and number of nodes to visit. Foster and Ryan [80] have developed an SPP model for an extended vehicle routing problem that incorporates restrictions on work load and coverage, reflecting real world situations. A different application of an SPP model for the vehicle routing problem by Lee [105] incorporates tabu search to find the capacity mix and routes for a fleet of vehicles.

A more recent use of a set partitioning model is concerned with the routing of trains through a railway junction. Lusby et al. [110] apply the method to obtain train schedules that are needed quickly at various times of the day due to the dynamic nature of the problem. A branch-and-price solution approach is then used to obtain good solutions which have been tested with data from a major German railway company. In another railway application, Rezanova and Ryan [133] use an SPP model to re-schedule train driver duties if a disruption to the timetable occurs. By using a branch-and-bound approach, it is reported that integer solutions are found within seconds.

Another application of the SPP model is the division of students in a class into several smaller groups that provides a good representation of the overall classroom population. Desrosiers et al. [59] use an enumeration of all possible groups and an SPP model in order to balance the attributes among the groups.

Applications of the SPP in Healthcare

Set partitioning methods have also been applied to a wide range of healthcare related problems. Fei et al. [78] have used a SPP model to assign elective surgical patients to operating theatre slots, taking into consideration constraints relating to operating theatre and surgeon availability. However, emergency surgical cases are not taken into account in this study. In their set partitioning model, a subset of feasible plans is selected in order to minimise the cost of scheduling the individual patients. Here, a plan represents an assignment of surgical cases to an operating theatre, and the cost reflects the number of unused or overtime hours of the corresponding operating theatres. A column generation based heuristic is then employed to solve the problem.

Set partitioning has also been used to schedule anesthesiologists for surgery, based on the matching of skills with specific tasks [73]. Here, special consideration is given to the generation of the set of possible tasks for each anesthesiologist, and the corresponding cost which reflects the relative desirability of the tasks. One practical benefit of scheduling in this way, as identified in the paper, is the reduction of time spent by physicians in the hospital on scheduling the anesthesiologists.

Milburn and Hall [120] have used the SPP as an alternative approach to location analysis for home-health district nursing. This problem concerns the allocation of district nurse subunits to district centers such that each subunit must be allocated to exactly one district. According to Milburn and Hall, an advantage of using the SPP framework over location analysis is that the SPP model does not require a fixed set of district centers. By considering every possible combination of subunits that could form a district center, variable sizes of district centers are allowed.

Finally, a decision support system has been developed by researchers in Sweden to aid the allocation of patient visits to care providers [75]. The system uses an SPP model to evaluate all feasible schedules of timetables of visits for the carers. The system has been implemented in several home care organisations in Sweden and it has been reported that considerable amounts of time have been saved on the daily planning time, and that the quality of the timetables produced has been improved.

As illustrated with the above examples, SPP models have been used in a wide variety of healthcare research; however, it is not apparent that a set partitioning approach has been used to construct an MSS with relation to the bed constraints prior to this research.

4.2 Set Partitioning Based Optimisation Model for the Construction of the MSS

A set partitioning optimisation model is adopted here for the construction of the MSS because of its scope to include constraints on the operating theatres and post-operative bed requirements. The ability to generate and limit the number of candidate schedules as inputs into the optimisation model can also help to reduce the size of the problem.

4.2.1 Development of the Proposed Model

The proposed model for the construction of the MSS is based on an SPP model and is outlined below. The aim of the model is to select a subset of possible 'plans' for each surgical specialty subject to a number of constraints. A plan for a specialty defines which operating theatre the specialty has use of on which day of the week and during which session, a.m. or p.m. Plans also reflect a specialty's preferences of theatres and days through the use of scheduling rules. The solution to the model will provide one plan for each specialty which, when put together, will form the optimal MSS.

Recall from Section 4.1.1 that the basic SPP model is:

$$\begin{aligned} \min \quad z &= c^T x \\ \text{s.t.} \quad Ax &= e \\ x &\in \{0,1\}^n \end{aligned} \tag{4.11}$$

The constraints specified in the A matrix above can be interpreted as choosing the optimal partition of subsets of operating theatre sessions. The basic proposed set partitioning based model for the construction of the MSS, that also takes into account post-operative bed constraints, takes the following form:

$$\min \quad z = c^T x \tag{4.12}$$

$$\text{s.t.} \quad Ax = e \tag{4.13}$$

$$Bx \leq d \tag{4.14}$$

$$x \in \{0,1\}^n \tag{4.15}$$

There are now two sets of constraints: one set for the operating theatre sessions (4.13) and the other for the post-operative bed constraints (4.14). The additional bed constraints ensure that there are not more beds required than available on each ward on each day. The addition of these bed constraints results in a deviation from the pure SPP formulation as shown in Model 4.11. As discussed in Section 4.1.2, this often occurs in practice when applying this optimisation method to real world applications, however, the characteristics of the SPP can still be exploited.

In the model, x_j, for $j = 1, \ldots, n$, are binary decision variables that indicate whether or not plan j is selected in the final solution, and c is a vector giving the cost of each plan. Together they form the total 'cost' of the chosen plans, which we seek to minimise. The objective function and cost associated with each plan will discussed in Section 4.2.2.

A is an $m \times n$ binary matrix where the columns represent possible plans for each surgical specialty. The generation of the A matrix will be discussed in Section 4.2.3. The first s rows of A represent generalised upper bound (GUB) constraints that relate each plan to a specific specialty. These GUB constraints specify that only one plan can be chosen for each specialty. The remaining rows of A represent constraints for each operating theatre session and consist of elements to be partitioned. An operating theatre session is characterised by the theatre, day and whether it takes place in the morning or afternoon. The constraints ensure that only one specialty is allocated to an operating theatre session in the optimal MSS. The A matrix has elements:

$$a_{ij} = \begin{cases} 1 & \text{if operating theatre session } i \text{ is used in plan } j \\ 0 & \text{otherwise} \end{cases}$$

for $i = 1, \ldots m$ and $j = 1, \ldots, n$.

The right-hand side values of the constraints associated with the A matrix are given in the m-vector $e = (1, 1, 1, \ldots, 1)^T$. This indicates that only one plan must be selected in the solution for each specialty (the GUB constraints), and that only one specialty can occupy an operating theatre session at any one time (the operating theatre constraints).

The entries of the B matrix are determined from the plans in the A matrix and represent the number of surgical inpatients who require beds for each plan on each ward on each day. The bed requirements consists of patients in beds for pre- and post-operative care; the generation of the B matrix will be discussed in Section 4.2.3. The elements of B are:

$$b_{kj}^{(l)} = \text{ number of beds required on ward } k \text{ on day } l \text{ for plan } j$$

for $k = 1, \ldots, p$, $j = 1, \ldots, n$ and $l = 1, \ldots, q$.

Bed constraints are constructed so that the number of beds required on each ward on each day must be less than or equal to the number of beds available. The right-hand side values of these constraints are in the vector d and represent:

$$d_k^{(l)} = \text{ number of beds available on ward } k \text{ on day } l$$

for $k = 1, \ldots, p$ and $l = 1, \ldots, q$.

Any optimisation software using a Simplex-based algorithm will intrinsically convert inequality constraints into equality constraints via the use of slack variables [54]. Using this idea, the bed constraints in the MSS SPP model will be converted into equality constraints through the use of slack and surplus decision variables [136]. The bed constraints will be treated as elastic in order to keep track of the difference between the available and required beds on each ward. The inclusion of slack and surplus variables will also allow for the 'sharing' of beds between different wards within the model.

In this context, slack variables in a bed constraint can be thought of as the number of unused or *empty* beds on a certain ward on a certain day. If there are fewer beds required than available on a ward, i.e. $b_{kj}^{(l)} < d_k^{(l)}$, then there is some slack in the system, comprising of $d_k^{(l)} - b_{kj}^{(l)}$ empty beds. Accordingly, the surplus variables can be thought of as the number of *additional* beds required on a ward in order to meet the patient bed requirements. This would occur when there are more patients who require a bed on a ward than there are physical beds on the ward, i.e. $b_{kj}^{(l)} > d_k^{(l)}$. The surplus for this ward is therefore $b_{kj}^{(l)} - d_k^{(l)}$ additionally required beds.

It is known from discussions with managers in UHW that not all wards can share beds with other wards. In order to control the transference of beds between wards, or equivalently which slack and/or surplus variables can be used in the bed constraints, a matrix is used to define allowable transitions of patients between wards. This transition matrix, W, is a square $p \times p$ matrix of zeros and ones. W is not necessarily symmetric, since a ward does not have to reciprocate the sharing of beds with another ward. W is informed from knowledge obtained from the hospital on which wards each specialty can use, and so W is assumed to be constant for each day l. The elements of W are:

$$w_{kv} = \begin{cases} 1 & \text{if patients meant to be on ward } k \text{ are able to use beds on ward } v \\ 0 & \text{otherwise} \end{cases}$$

for $k = 1, \ldots, p$ and $v = 1, \ldots, p$.

In order to be able to determine how many beds are transferred between each ward, consider a $(p \times p) \times q$ matrix $Z^{(l)}$ whose elements, $z_{kv}^{(l)}$, are slack and surplus decision variables that specify how many beds are moved from ward k to ward v on day l. It is important to note that the W matrix concerns the allowable transitions of patients between wards (which is what would happen in reality: patients would be moved to an empty bed on a different ward), but the $Z^{(l)}$ matrix concerns the number of beds that are 'moved' between wards in the model. Of course, in reality the patients would be moved and not the beds, but the notion of beds being moved must be used in the model because the bed constraints concern beds, not patients. This correspondence between the W and $Z^{(l)}$ matrices is shown in Figure 4.1.

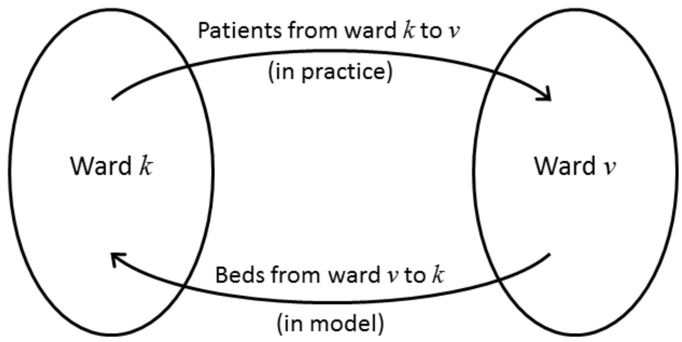


Figure 4.1: Correspondence between the W and $Z^{(l)}$ matrices

Each element of row k of $Z^{(l)}$, $z_{kv}^{(l)}$ $\forall$ v, represents the number of empty beds on ward k on day l that are 'given' to ward v. Following from the definition of the slack variables for this application, the sum of the elements of row k of $Z^{(l)}$ represents the number of empty beds on ward k on day l.

Each element of column k of $Z^{(l)}$, $z_{vk}^{(l)}$ $\forall$ v, represents the number of additional beds used by ward k on ward v on day l. Similarly, following from the definition of the surplus variables for this application, the sum of the elements of column k of $Z^{(l)}$ represents the number of extra beds required by ward k on day l.

Combining elements in W and $Z^{(l)}$ gives the total allowable slack and surplus for each ward as follows:

i) The total number of empty beds (slack) on ward k on day l is:

$$\sum_{v=1}^{p} w_{vk} z_{kv}^{(l)} \quad \forall \quad k = 1, \ldots, p, \;\; l = 1, \ldots, q$$

i.e. the sum product of column k of W and row k of $Z^{(l)}$.

ii) The total number of additionally required beds (surplus) on ward k on day l is:

$$\sum_{v=1}^{p} w_{kv} z_{vk}^{(l)} \quad \forall \quad k = 1, \ldots, p, \;\; l = 1, \ldots, q$$

i.e. the sum product of row k of W and column k of $Z^{(l)}$.

Hence, the bed constraints for each ward on each day (constraint 4.14) can be formulated as:

$$\sum_{j=1}^{n} b_{kj}^{(l)} x_j - \sum_{v=1}^{p} w_{kv} z_{vk}^{(l)} + \sum_{v=1}^{p} w_{vk} z_{kv}^{(l)} = d_k^{(l)} \quad \forall \quad k = 1, \ldots, p, \;\; l = 1, \ldots, q \qquad (4.16)$$

Now that slack and surplus decision variables are included in the model, an additional constraint is needed to ensure that the number of extra beds should not exceed the number of empty beds across all wards on each day. This is achieved by enforcing that the sum of the surplus variables across all wards on each day does not exceed the sum of the slack variables across all wards on each day. This also prevents the total number of beds in the hospital from being exceeded. The daily constraint for the slacks and surpluses for each ward is:

$$\sum_{k=1}^{p} \sum_{v=1}^{p} w_{kv} z_{vk}^{(l)} \leq \sum_{k=1}^{p} \sum_{v=1}^{p} w_{vk} z_{kv}^{(l)} \quad \forall \quad l = 1, \ldots, q \qquad (4.17)$$

From discussions with hospital managers in UHW, it is apparent that some specialties do not want their beds to be used by any other specialty. An occurrence of a patient from a different specialty using a bed on a ward is known as having 'outliers' on the ward, and can be seen as exacerbating the problem of a shortage of beds when demand is high. Specialties may not want outliers on their ward(s) due to clinical reasons (it may be medically unsafe to have patients of a specialty on a different ward), or it might be the case that they want to 'reserve' their own beds in case of an influx of emergency patients. Both are valid reasons, so the bed constraints will be modified to incorporate the ability to 'reserve' beds on wards if required.

Allowable movements of patients between wards are declared in the transition matrix, W; however, it is not possible in the present formulation to state the fact that empty beds on a ward should not be shared with another ward. To illustrate this, let there be f empty beds on ward k, i.e. the sum of slack variables for ward k is f:

$$\sum_{v=1}^{p} w_{vk} z_{kv}^{(l)} = f \quad \text{for ward } k.$$

However, if ward k does not allow any other ward to use its empty beds, i.e. does not allow any surpluses from other wards to fill its slacks, then in the current W matrix, $w_{vk} = 0 \;\forall\; v = 1, ..., p$ for ward k would have to be specified. This would then imply $\sum_{v=1}^{p} w_{vk} z_{kv}^{(l)} = 0$ for ward k. Evidently this causes a conflict between the desire to reserve beds, and the current formulation of the bed constraints.

The suggested solution is to introduce the notion of a conceptual 'dummy ward' that would act as a holding place for reserved beds in the model. Empty beds need to be shared with a ward in the model, hence wards that do not want to give their empty beds to other real wards should send the beds to the dummy ward. In the model, this is achieved by giving the slacks from the real ward to the surplus of the dummy ward.

The dummy ward needs to be included in both the W and $Z^{(l)}$ matrices. As before, there are p real wards, for which the W matrix states between which wards patients may be moved, and the $Z^{(l)}$ matrix states how many beds are transferred between those wards. The dummy ward acts by taking the slack from real wards that do not want patients from other specialties in their beds. This corresponds to having an extra row in the W matrix, and an extra column in the $Z^{(l)}$ matrix.

Let the wards be denoted as $k = 1, \ldots, p+1$ and $v = 1, \ldots, p+1$, and let wards $1, \ldots, p$ be the real wards and let ward $p+1$ be the dummy ward. If ward k does not want any other ward to use its empty beds, then ward k should give its empty beds to the dummy ward. This corresponds to allowing only patients to move from the dummy ward to the real ward k (and no patients from any other real ward), i.e.

$$w_{kv} = \begin{cases} 0 & \forall\ v = 1, ..., p \text{ and for any ward } k \in 1, \ldots, p, \text{that does not want} \\ & \text{to share its beds with any other ward,} \\ 1 & \text{for ward } v = p+1 \text{ and for any ward } k \in 1, \ldots, p. \end{cases}$$

For the $Z^{(l)}$ matrix, this corresponds to 'giving' empty beds on ward k only to ward $p+1$ (and not to any other real ward). i.e.

$$z_{kv}^{(l)} = \begin{cases} 0 & \forall\ v = 1, ..., p \text{ and for any ward } k \in 1, \ldots, p, \text{that does not want} \\ & \text{to share its beds with any other ward,} \\ > 0 & \text{for ward } v = p+1 \text{ and for any ward } k \in 1, \ldots, p. \end{cases}$$

The bed constraint (4.16) needs to be re-written in order to account for the dummy ward, as the total slack and surplus for each ward on each day now need to include the dummy ward. The surplus (the number of additional beds required on ward k

on day l), can now be expressed as follows:

$$\text{Surplus} = \sum_{v=1}^{p} w_{kv} z_{vk}^{(l)} \quad \forall \quad k = 1, \ldots, p, \;\; l = 1, \ldots, q$$

For the surpluses, v corresponds to the column number of W and the row number of $Z^{(l)}$. When a dummy ward is used, there remain p rows in the W matrix and p columns in the $Z^{(l)}$ matrix. Hence the values of v go from 1 to p. Also for the surpluses, k corresponds to the row number of W and the column number of $Z^{(l)}$. In the case that a dummy ward is used, there are now $p+1$ columns in W and $p+1$ rows in $Z^{(l)}$; however, because the dummy ward never receives real patients from real wards, no surpluses are ever given to the dummy ward from real wards. Therefore the values of k go from 1 to p (not $p+1$).

The slack (number of empty beds on ward k on day l), can now be expressed as follows:

$$\text{Slack} = \sum_{v=1}^{p+1} w_{vk} z_{kv}^{(l)} \quad \forall \quad k = 1, \ldots, p, \;\; l = 1, \ldots, q$$

For the slacks, v corresponds to the row number of W and the column number of $Z^{(l)}$. When a dummy ward in used, there are $p+1$ rows in the W matrix and $p+1$ columns in the $Z^{(l)}$ matrix. Hence the values of v go from 1 to $p+1$. Also for the slacks, k corresponds to the column number of W and the row number of $Z^{(l)}$. In the case that a dummy ward is used, there remain p columns in W and rows in $Z^{(l)}$.

Constraint 4.17 is also altered with the new expressions of the slacks and surpluses. The formulation of the set partitioning based model for the construction of the MSS is therefore given in Model 4.18:

$$\min \quad \sum_{j=1}^{n} c_j x_j \tag{4.18a}$$

$$\text{s.t.} \quad \sum_{j=1}^{n} a_{ij} x_j = 1 \quad \forall \quad i = 1, ..., s \tag{4.18b}$$

$$\sum_{j=1}^{n} a_{ij} x_j \leq 1 \quad \forall \quad i = s+1, ..., m \tag{4.18c}$$

$$\sum_{j=1}^{n} b_{kj}^{(l)} x_j - \sum_{v=1}^{p} w_{kv} z_{vk}^{(l)} + \sum_{v=1}^{p+1} w_{vk} z_{kv}^{(l)} = d_k^{(l)} \quad \forall \quad k = 1, ..., p, \;\; l = 1, ..., q \tag{4.18d}$$

$$\sum_{k=1}^{p} \sum_{v=1}^{p} w_{kv} z_{vk}^{(l)} \leq \sum_{k=1}^{p} \sum_{v=1}^{p+1} w_{vk} z_{kv}^{(l)} \quad \forall \quad l = 1, ..., q \tag{4.18e}$$

$$x_j \in \{0, 1\} \quad \forall \quad j = 1, ..., n$$

$$z_{kv}^{(l)} \geq 0 \text{ and integer} \quad \forall \quad k = 1, ..., p, \;\; v = 1, ..., p+1, \;\; l = 1, ..., q$$

The objective (4.18a) is to minimise the ‘cost’ of using each plan. Equality constraints in (4.18b) represent the GUB constraints: only one plan must be chosen for each specialty. Constraints for the use of each operating theatre session (4.18c) specify that only one specialty can use each operating theatre during each session. Bed constraints (4.18d) for each ward on each day ensure that if more beds are required than physically available on a ward, then sharing of beds is allowed between wards through the use of slack and surplus variables. Constraint 4.18e is required in order to ensure that the overall total number of beds is not exceeded on each day of the planning horizon. The decision variables x_j are binary, and the slack and surplus decision variables, $z_{kv}^{(l)}$, must be non-negative integers.

A summary of the notation used in Model 4.18 is given in Table 4.1:

Notation	Indicies	Definition
c_j	j = 1, . . . , n	Cost of plan j.
a_{ij}	$i = 1, \ldots, s$	Indicates if plan j refers to specialty i.
	$j = 1, \ldots, n$	
a_{ij}	$i = s+1, \ldots, m$	Indicates if specialty is scheduled in
	$j = 1, \ldots, n$	operating theatre session i in plan j.
$b_{kj}^{(l)}$	$k = 1, \ldots, p$	Bed requirement in ward k on day l for plan j.
	$l = 1, \ldots, q$	
	$j = 1, \ldots, n$	
w_{kv}	$k = 1, \ldots, p$	Indicates if bed sharing is allowed between
	$v = 1, \ldots, p+1$	wards k and v.
$d_k^{(l)}$	$k = 1, \ldots, p$	Number of beds available on ward k on day l.
	$l = 1, \ldots, q$	
x_j	$j = 1, \ldots, n$	Decision variable: plan j is chosen or not.
$z_{kv}^{(l)}$	$k = 1, \ldots, p$	Decision variable: number of beds transferred
	$v = 1, \ldots, p+1$	between wards k and v on day l.
	$l = 1, \ldots, q$	

Table 4.1: Notation used in MSS optimisation model

4.2.2 Objective Function

In this section, a number of candidate objective functions will be discussed in order to choose the most suitable objective function for Model 4.18. The objective function needs to align with the aims of the hospital for constructing an MSS and easily distinguish between good and bad plans.

The first, and possibly simplest, candidate objective function is a monetary cost applied to each of the possible plans. This cost could reflect the cost per session of assigning each specialty to each operating theatre, and the cost of having a patient in a bed per day. The minimisation of the cost of running the operating theatres and patient recovery on the wards is certainly desirable for the hospital management, however, there are many other factors that affect the monetary cost of scheduling surgeries. The main variation in costs would come from the operating theatre aspect of this objective function, since the cost per bed per day of patients in the wards will only be affected by the duration of their LoS. The hospital managers in UHW have expressed the opinion that the operating theatres are not the restrictive resource in the problem they are experiencing. Therefore it does not seem appropriate to construct the MSS based on the cost of assigning specialties to the operating theatre sessions. It was not deemed suitable to continue with a monetary objective function due to the many other factors that would need to be accounted for and the difficulty in obtaining the required data from the hospital.

It is desirable to construct an MSS that satisfies the rules and preferences of everyone involved with conducting surgery. Preferences are often thought of as soft constraints that are desirable to satisfy, but not absolutely necessary to satisfy for a feasible solution. Hard scheduling rules, such as which operating theatre each specialty can use and which day of the week any specialty should or should not operate on, are specified by directorate and hospital managers and are built into the plans that make the A matrix (discussed in Section 4.2). The ability to build these rules implicitly into the generation of the subsets is an advantage of formulating this scheduling problem as an SPP [135]. However, there may also be other soft scheduling rules, or preferences, that could be reflected in the objective function by assigning lower values of 'cost' to a preferred plan. Therefore, when the objective function is minimised, the most preferable plans with lower 'cost' values will be chosen with respect to the operating theatre and bed constraints. However, there would be difficulty in discerning this preference information from stakeholders in the hospital. Experience gained through liaison with staff in the hospital suggests that different stakeholders have different preferences, e.g. the directorate managers might give their preferences based on data, which might be different to the preferences given by surgeons who might base theirs on their experience of working in the operating theatres. This approach was also deemed inappropriate because it was predicted that conflicting preferences would be collected from various hospital staff, and it was unclear how to resolve these conflicts.

The first two candidate objective functions discussed above are based on information that would need to be acquired from the hospital. If either of these were chosen, the data for the objective functions would have to be altered if the model was applied to a different hospital. This is perfectly valid, since the resulting MSS will be very specific to the particular hospital; however, it is not considered the best choice here where we are seeking to illustrate the flexibility of a generic scheduling model. Hence, the subsequent candidate objective functions use information from within the optimisation model to determine the 'costs' of each plan. The optimisation model allows for the sharing of beds between wards which could be particularly useful in reality when demand for beds on a ward becomes acute. However, it could also be argued that sharing beds, or in reality moving patients onto different wards, is not a desirable practice to encourage in the hospital. This could be because there may not be the correct equipment or skilled nurses on the different ward. Therefore, the next candidate objective function penalises any sharing of beds between wards. This approach bears similarities with that of Bard and Purnomo [18] for meeting the scheduling preference of nurses. The sharing of beds will still be allowed through the use of the slack and surplus variables, though it will be discouraged by the imposed penalties. The total surplus, or additionally required beds, is summed over all wards and all days. It may be the case that it is more undesirable for some wards to require additional beds than other wards. Therefore a penalty, say f_k, will be specified for ward k in the objective function, for all wards $k = 1, \ldots, p$. The objective function would therefore take the form:

$$\min \quad \sum_{l=1}^{q} \sum_{k=1}^{p} f_k \sum_{v=1}^{p} w_{kv} z_{vk}^{(l)}$$

Although this objective function minimises the amount of patients being moved to different wards in the model, after discussions with hospital management it was decided that this objective function did not quite align with the hospitals primary aim of reducing the number of cancelled patients. If it is medically safe for a patient to be put onto a different ward, then surgeons do this in order to free up bed space for more patients and thereby reduce the waiting list. There also still remains the capability within the model to eradicate any bed sharing, by setting the corresponding elements of the W matrix to zero.

Another of the hospital's primary objectives is to increase the number of patients moving through the system in order to reduce the waiting lists for elective surgery. This can be thought of as equivalently reducing the amount of unmet demand as a proportion of the total demand over the planning horizon. The amount of unmet demand can be considered as the difference between the demand and activity performed over a specified planning horizon for all surgical specialties. Unfortunately, this cannot be calculated because the developed model optimises the MSS based on the *desired* activity (the B matrix consists of required beds for all planned patients). Therefore, a fixed amount of patient demand will be seen in the model, so the difference in demand and activity will always be constant, and so cannot be used as an objective function.

Single objective functions have been discussed thus far, however a multiobjective function could be used in this setting. Zhang et al. [162], for example, use an objective function that consists of five different cost and penalty terms. They consider the delay in meeting surgery demand and unmet demand for inpatient and outpatient surgery, and penalise any undersupply of operating theatre hours to each specialty. A multiobjective function that uses any of these aspects could be appropriate, however, it will not be used here.

The objective function that is believed to be the most appropriate and consistent with one of the hospital's primary objectives is based on the idea of reducing the amount of unused bed days in order to make best use of the existing capacity on the wards. According to the OECD [129], a bed day is defined to be 'a day during which a person is confined to a bed and in which the patient stays overnight in a hospital'. This is a commonly used measure that is reported and used by hospital managers for the management of patients and hospital wards.

The number of bed days used on a ward on one day is the sum of the required beds over all the chosen plans for the optimal schedule, i.e. $\sum_{j=1}^{n} b_{kj}^{(l)} x_j$. Perhaps most important to the hospital management, is the utilisation of the beds on the wards. There is a fixed cost associated with running a ward, which includes such costs as equipment, nurses, building overheads, cleaners and even catering for patient's meals. Therefore, because these costs are fixed, hospital management are keen for these wards to be utilised as fully as possible. This can be thought of as maximising the number of patients on the wards at any one time. Equivalently, it can be thought of as minimising the number of empty beds on the wards at any one time. If the number of empty beds are kept to a minimum, the overhead costs for running the wards will be reduced when considering the cost per bed per patient. The number of empty beds on a ward can be expressed as the difference between the number of beds on the ward, and the beds required by inpatients, i.e. $d_k^{(l)} - \sum_{j=1}^{n} b_{kj}^{(l)} x_j$.

The measure of unused bed days will be used in the objective function since the objective will need to be minimised, which is consistent with the traditional SPP formulation. If the number of unused bed days is minimised, the notion of maximising the throughput of patients can then also be investigated – if there are more unused bed days on a ward, then this would imply that the throughput of patients could be increased in order to use these empty beds. This could be achieved by increasing the number of patients operated on, assuming that their LoSs follow similar distributions as at present.

Since the whole system of hospital beds is being considered in the model, it is of interest to include the unused bed days used on all wards on all days of the planning horizon. Hence the number of unused bed days is summed over all wards and days, i.e. $\sum_{k=1}^{p}\sum_{l=1}^{q}(d_k^{(l)} - \sum_{j=1}^{n} b_{kj}^{(l)} x_j)$. Therefore, the chosen objective function for the optimisation model to construct an MSS is:

$$\min \quad \sum_{k=1}^{p}\sum_{l=1}^{q}(d_k^{(l)} - \sum_{j=1}^{n} b_{kj}^{(l)} x_j)$$

On occasions when the chosen plans in the optimal solution result in more beds being required than available for a particular ward on a particular day, i.e. $\sum_{j=1}^{n} b_{kj}^{(l)} x_j > d_k^{(l)}$, this will result in a negative term in the overall summation for the objective function. This will help to reduce the objective function value, which is in line with the minimisation objective. However, this could be seen to be artificially reducing the objective function value, because it seems counterintuitive to get benefit (a lower objective value) from requiring more beds than available. This is allowable, however, since if a feasible solution is able to be found, then it must be the case that beds are being shared between wards via the slacks and surpluses. If this is undesirable, then this can easily be prevented by disallowing the sharing of beds by altering the W matrix. Hence a negative term in the objective function summation is not invalid.

From discussions with staff at UHW, the minimisation of the unused bed days seems to be consistent with the hospital's own objectives of utilising expensive resources: the beds on wards. This objective can also be used as part of the hospital's capacity planning strategy, since plans that bring the greatest throughput of patients will be selected during the optimisation. If the optimal MSS has spare capacity, i.e. empty beds on the wards, then there could be scope for increasing the number of patients brought in for surgery, if other constraints on the operating theatres would allow (e.g. enough operating theatre time for more operations).

4.2.3 Operating Theatre Constraints

As described previously in Section 4.2.1, the A matrix is an $m \times n$ binary matrix whose columns represent different possible plans for each surgical specialty. The first s rows of A represent GUB constraints enforcing that exactly one plan must be selected for each specialty in the optimisation. The remaining rows then represent constraints for each operating theatre session; only one specialty is allowed to be allocated to an operating theatre session in the optimal MSS. The structure of the A matrix is illustrated in Figure 4.2.

***n* possible plans**

***s* specialties**

***m* OR sessions**

			Specialty 1			Specialty 2				Specialty 3			...				
			1	1	1	0	0	0	0	0	0	0				=	1
			0	0	0	1	1	1	1	0	0	0				=	1
			0	0	0	0	0	0	0	1	1	1				=	1
			0	0	0	0	0	0	0	0	0	0				=	1
OR 1	Mon	AM	1	1	0	1	0	1	0							≤	1
		PM	1	0	0	0	0	1	0							≤	1
OR 1	Tues	AM	0	1	0	1	1	1	0							≤	1
		PM	0	0	0	0	1	1	1							≤	1
OR 1	Weds	AM	0	0	0	0	0	0	1							≤	1
...	...	...															
OR 14	Fri	AM	0	0	1	0	0	0	1							≤	1
		PM	0	0	1	0	0	0	1							≤	1

Figure 4.2: Illustrative diagram of the A matrix

The rows that correspond to the GUB constraints (indicated in yellow in Figure 4.2) are automatically filled in when each plan is generated. The remainder of this section describes how the rows that correspond to the operating theatre constraints are generated.

Enumeration in used to generate all possible plans for each specialty since this technique has been shown in the literature to be very effective. An approach similar to Ryan [135] is adopted here to enumerate the plans for each specialty. A skeleton plan is first generated for each specialty that fixes in the plan any zeros or ones to indicate that a specialty must be or must not be scheduled in certain operating theatre sessions respectively. Then using the skeleton plan, all combinations of plans are added to the A matrix until no further additions are possible. This process is repeated for each specialty, resulting in the A matrix being constructed in the following way as described in Algorithm 1.

Algorithm 1 Generation of the A Matrix

```
for each specialty i do
    Make the skeleton plan for specialty i (Algorithm 2)
    Generate all plans for specialty i via enumeration (Algorithm 3)
    Append block of plans for specialty i onto A matrix
end for
```

A skeleton plan is constructed for each specialty using rules and preferences for defining which operating theatre or day of the week is allowed or preferred. The skeleton plan is an initial column of the A matrix for a particular specialty that consists of three different numbers: '0' denotes that the specialty must not be allocated to that operating theatre session, '1' denotes that the specialty must be allocated to that operating theatre session, and '-1' denotes that fact that the specialty could be allocated to that operating theatre session, i.e. it is a temporary value that will be replaced by the value '0' or '1' in the final A matrix. The skeleton plan is constructed according to Algorithm 2.

Algorithm 2 Construction of the skeleton plan for specialty i

```
Input scheduling rules
for each session j do
    if specialty i is not to be scheduled in session j then
        skeletonPlan[j] = 0
    else if specialty i is to be scheduled in session j then
        skeletonPlan[j] = 1
    else
        skeletonPlan[j] = −1
    end if
end for
Append skeletonPlan to the A matrix
```

The enumeration of every possible plan for one specialty starts with the skeleton plan and is achieved by passing through each element in turn, and deciding whether or not to change a '-1' entry into a '0' or '1'. If there are already the required number of operating theatre sessions, then all subsequent '-1' entries are changed to '0's. If there are still operating theatre sessions to allocate, then the entry in the current column will be changed to '0' and a copy of the column will be taken and appended to the A matrix for later use. At each decision point, the number of operation room sessions that have already been allocated in the plan is checked against the required number of operating theatre sessions for that specialty. As the plans are generated, the algorithm pairs the allocated operating theatre sessions so that, if a specialty has been allocated to a morning session, then it will also be allocated to an afternoon session if more sessions are required. This reflects the hospital's preference that specialties have whole day sessions rather than half day sessions. All '-1' entries will be changed in one plan before the algorithm moves onto the next plan. The enumeration algorithm performed is described in Algorithm 3.

Algorithm 3 Enumeration of all plans for a specialty

for each column j **do**
  for each session i **do**
    if $a_{ij} = -1$ **then**
      if Number of 1's in column j of the A Matrix = the required number of sessions **then**
        $a_{ij} = 0$
      else if (Number of 1's in column j of the A Matrix < the required number of sessions) $\cap$ (Number of remaining 1's < Number of -1's in column j) **then**
        $a_{ij} = 0$
        copyCol = a copy of column j of the A Matrix
        copyCol(i) = 1
        Append copyCol onto the RHS of the A Matrix
        numCols = numCols + 1
      else if Number of 1's in column j of the A Matrix < the required number of sessions $\cap$ (Number of remaining 1's = Number of -1's in column j) **then**
        $a_{ij} = 1$
      end if
    end if
  end for
end for

Since all desirable plans for each specialty are being enumerated, it is of interest to estimate the size of the A matrix, or equivalently the number of decision variables that will be required. SPPs can be very large as reported in [135], so it is of interest to see how the size of typical instances of this application problem compare. The number of plans for each specialty is the same as the number of distinct combinations of zeros and ones in each column.

For each specialty i, let

$$x_i = \text{number of operating theatre sessions available to specialty } i$$
$$y_i = \text{number of operating theatre sessions required by specialty } i$$

The enumeration algorithm groups the sessions in which a specialty is scheduled into whole day sessions where possible. Moving from using half day sessions to whole day sessions, the number of operating theatre sessions available to specialties is halved. The number of operating theatre sessions available to specialty i is thus $x_i/2$.

If y_i is even, then the number of plans for specialty i is ${}^{x_i/2}C_{y_i/2}$, since the number of operating theatre sessions required by the specialty also needs to be halved when whole day sessions are considered.

If y_i is odd, then $y_i - 1$ required operating theatre sessions are grouped into whole day sessions, as in the case when y_i is even, and the 'leftover' odd session is then allocated to an a.m. or p.m. session. The number of possible allocations of this remaining odd session is equal to the number of zero elements in the A matrix that was constructed with the $y_i - 1$ sessions, since each zero can be changed to a one. The number of plans generated with the $y_i - 1$ sessions is ${}^{x_i/2}C_{(y_i\,-\,1)/2}$. The number of zeros in these plans is equal to the difference between the total number of elements in these plans and the number of ones in these plans, i.e.

$$\text{Number of zeros in the plans} = \left(x_i \cdot {}^{x_i/2}\, C_{(y_i\,-\,1)/2}\right) - \left((y_i - 1) \cdot {}^{x_i/2}\, C_{(y_i\,-\,1)/2}\right)$$

In summary, the number of plans generated through enumeration for each specialty is

$$\text{The number of plans for specialty } i = \begin{cases} {}^{x_i/2}C_{y_i/2} & \text{if } y_i \text{ even} \\ (x_i - y_i + 1) \cdot {}^{x_i/2}\, C_{(y_i\,-\,1)/2} & \text{if } y_i \text{ odd} \end{cases}$$

The number of plans in the A matrix is then the total number of plans that have an even number of required operating theatre sessions and those that have an odd number of required operating theatre sessions. i.e.

$$\text{Number of plans in } A \text{ matrix} = \sum_{i \in \mathbb{E}} {}^{x_i/2}C_{y_i/2} + \sum_{i \in \mathbb{O}} (x_i - y_i + 1) \cdot {}^{x_i/2}\, C_{(y_i\,-\,1)/2}$$

where $\mathbb{E}$ is the set of specialties that require an even number of operating theatre sessions, and $\mathbb{O}$ is the set of specialties that require an odd number of operating theatre sessions.

Using the above formula, the number of plans in the A matrix resulting from using the current scheduling rules for the MSS used in UHW are calculated and given in Table 4.2.

	Number of sessions		
Specialty (i)	**Available** (x_i)	**Required** (y_i)	**No. plans in A matrix**
Cardiac	20	20	1
CEPOD	10	10	1
Colorectal	10	8	5
ENT	10	5	60
General	20	8	210
Liver	10	2	5
Neuro	20	20	1
Ophthalmology	10	2	5
Oral	30	6	455
Paeds ENT	10	1	10
Paeds General	10	8	5
Paeds Trauma	10	2	5
Renal	30	6	455
Scoliosis	10	4	10
Thoracic	10	8	5
Trauma	10	2	5
Urology	10	10	1
Vascular	20	8	210
		Total no. plans in A matrix =	1449

Table 4.2: Size of the A matrix for current UHW MSS scheduling rules

As a validation check, 1449 plans are generated in the A matrix when the current scheduling rules that are used in UHW are used to construct an MSS. This is a large problem, though not as large as the problem instances reported by [147] which involved thousands of decision variables.

4.2.4 Bed Constraints

As described in Section 4.2.3, the A matrix defines in which theatre and at what time each surgical specialty will operate. Using the A matrix, the B matrix is then generated by filling in the predicted number of beds required on each ward on each day for each plan. This section describes how the B matrix is generated and the different methods used to calculate the bed requirements. The structure of the B matrix is illustrated in Figure 4.3.

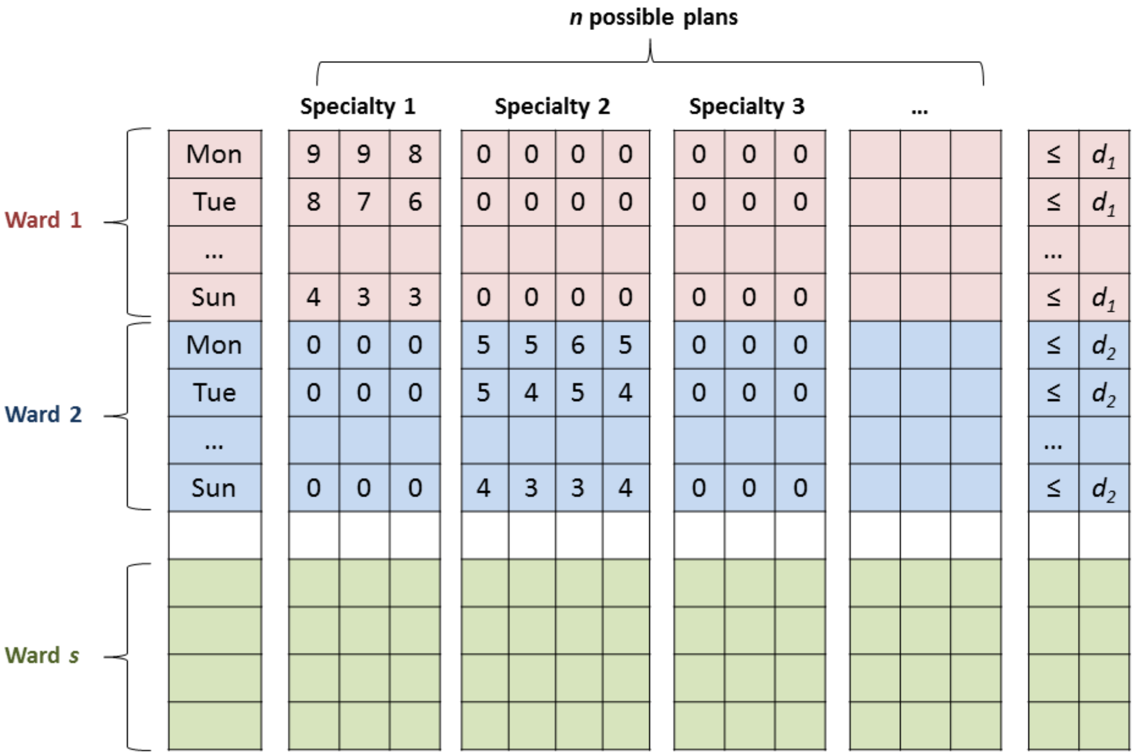


Figure 4.3: Illustrative diagram of the B matrix

This B matrix illustrates the relationship between the A and B matrices. For example, it can be seen in Figure 4.2 that the first plan in the A matrix has scheduled Specialty 1 in operating theatre 1 in both the morning and afternoon sessions on Monday. The first plan in the B matrix therefore gives the predicted bed requirements of Specialty 1 for this plan on Ward 1 for each day of the week. Nine beds are required on Monday, then eight on Tuesday. This is assumed to continue to reduce to four beds on Sunday, reflecting the fact that patients are discharged from hospital according to the specialty's LoS distribution.

As discussed in Section 3.5, the LoS of an episode in hospital for a surgical patient can be split into two separate LoSs: pre-operative (before surgery) and post-operative (after surgery). The LoS is rounded up to the nearest whole day since, according to the OECD [129], a bed day is defined as 'a day during which a person is confined to a bed and in which the patient stays overnight in a hospital'. It is deemed appropriate to use whole day pre-operative and post-operative LoS in order to capture the fact that a bed day involves an overnight stay.

Whenever a specialty is scheduled to operate in the A matrix, a bed is required in the model (in the B matrix) for the total duration of the patient's pre-operative and post-operative LoS. The number of patients that are operated on per-session is known, the pre-operative LoS is a user-specified duration, and several methods of generating the bed requirements for the post-operative LoS are investigated. The optimisation model concerns the bed requirements for mainly elective surgical inpatients, with the exception of surgical patients being admitted to hospital via the emergency CEPOD theatre. This theatre generates a high demand for beds, so is deemed necessary to include it in the model.

The basic algorithm for generating a B matrix based on the A matrix is given in Algorithm 4.

Algorithm 4 Basic generation of the B matrix

```
for each column j do
  Look-up which specialty plan j refers to from the GUB constraints.
  for each session i do
    if a_ij = 1 then
      Find corresponding day of surgery in B matrix.
      Put the new arrivals in the B matrix before the day of surgery for their
      pre-operative LoS.
      Put the new arrivals in the B matrix on the day of surgery.
      Update B matrix with the post-operative bed requirement.
    end if
  end for
end for
```

A look-up table of which ward(s) each specialty sends their patients to is used in the generation of the B matrix in order to get the bed requirements for the appropriate bed constraints. It is assumed that as soon as a patient's post-operative LoS is complete, they leave the hospital and are not modelled as going to another ward in the hospital. Although in reality a surgical patient may move to a different ward for a different medical need, the current surgical episode is only being considered in the model and so if the patient moves to a different ward, they are starting a new episode.

There is, however, one ward that is an exception to this rule: the CCU. The CCU is a special ward in which the patients who require the most intense medical attention are treated, typically for a short period of time. In the model, patients are either deemed well enough to leave the CCU and move onto another surgical ward for further post-operative recovery, or unfortunately they die while in the CCU, ending their episode in hospital. The CCU is the only ward in the model from which patients can move into another ward. A separate look-up table is used to send patients from the CCU to appropriate surgical wards to continue their post-operative recovery. It is assumed that no patients return to the CCU once they have left.

Data concerning patients' post-operative LoSs in each ward and the CCU is used to determine how long patients will remain in beds in the model. Several approaches to modelling the post-operative LoS are discussed below. The method used to determine which wards patients go to and how the post-operative LoS is used to generate the B matrix is given in Algorithm 5.

Algorithm 5 Assignment of patients to wards for the generation of the B Matrix

```
for each column j do
  Look up the specialty plan j refers to and the number of new arrivals that will
  require a bed.
  for each session i do
    if a_ij = 1 then
      Calculate the number of new arrivals that go to each of the wards (including
      the CCU) using a look-up table.
      Calculate, using i, the day of surgery.
      Put a specified proportion of patients into beds on wards one and/or two
      days before day of surgery for pre-operative LoS.
      if Some patients are to be sent to the CCU then
        Send those patients to the CCU and update the B matrix with their
        CCU LoS.
        Calculate how many patients are discharged from the CCU on each day
        of the week.
        Send the other patients from surgery straight to other wards (not the
        CCU) and update the B matrix with their post-operative LoS.
      else
        Send patients from surgery straight to wards (not the CCU) and update
        the B matrix with their post-operative LoS.
      end if
      Re-adjust number of CCU discharges according to the mortality rate.
      for each day, d, in the planning horizon do
        if The number of discharges from the CCU on day d is > 0 then
          Distribute these discharges between all other wards using a look-up
          table.
          for Each ward l do
            if Any CCU discharges are sent to ward l then
              Send these patients to ward l and update the B matrix with their
              post-operative LoS.
            end if
          end for
        end if
      end for
    end if
  end for
end for
```

In this research, several different methods of using the data on post-operative LoS are used to generate the B matrix. A patient's LoS is categorised based on their surgical specialty, except when the patients are on the CCU, in which case the LoS is based on the overall CCU LoS. Patients who were operated on in the CEPOD theatre have a LoS based on their actual surgical specialty. As we saw, the LoS data was analysed and discussed in Section 3.5. Here the B matrix is generated using three methods: (1) the average LoS for each ward, (2) the expected bed count on each ward on each day after surgery, and (3) the conditional probability of leaving the hospital on each day after surgery. Each method is discussed in the following sections.

Generating the B Matrix: Method 1 – Mean LoS

The first method of using the LoS data to generate the B matrix is based on using the mean LoS for each specialty. It is assumed that each patient in the model has a LoS equal to the mean LoS of their specialty, and it does not vary. In this method, all patients who have surgery during the same operating theatre session will therefore leave the ward together. This method does not take the uncertainty of patient LoS into account.

Generating the B Matrix: Method 2 – Conditional Probability of Failure

In this method of filling the B matrix, an example, or 'scenario', of the number of beds required is generated based on each plan. Using information from the LoS data, on each day after surgery, the probability of each patient leaving hospital will be evaluated to determine whether or not they will stay in the bed until the next day.

Let the post-operative LoS for a patient be denoted by the random variable T. More specifically, T denotes the duration of time after surgery until the patient either leaves hospital (end of the spell in hospital) or moves to the care of a different specialty (end of episode). In either case, T is effectively the time taken for the patient to 'recover' from surgery.

For each specialty, we need to calculate the probability that a patient leaves on day d, given that the post-operative LoS has already reached d days. In survival analysis, this is known as the conditional probability of failure. To enable this, the post-operative LoS will be used to find the survival distribution: the probability of a patient staying in a bed past day d. The Kaplan-Meier estimate of the survivor function is used because no parametric distribution has been found fit to our LoS data [97].

From the theory of survival analysis, the survivor function, $S(t)$, is the probability that the random variable T takes a value greater than a specified time, t.

$$S(t) = P(T > t)$$

In this application, $S(t)$ is the probability that the post-operative LoS is longer than a specified time t. Theoretically, the function $S(t)$ decreases smoothly from $S(t) = 1$ at $t = 0$, towards zero as t increases towards infinity. If a parametric distribution cannot be fitted to $S(t)$, then $S(t)$ can be estimated by step functions that provide the survival probability for discrete time points; days in this case. An estimate of $S(t)$ can be calculated using the Kaplan-Meier method, sometimes known as the product-limit estimate. This method is non-parametric, so no assumptions about the data are made.

For the Kaplan-Meier estimate of the survivor function, let $t_{(j)}$, $j = 0, 1, 2, \ldots, k$, be the ordered LoSs. The estimate of the survivor function, $S(t)$, at time $t_{(j)}$, is given by the general Kaplan-Meier formula:

$$\hat{S}(t_{(j)}) = \hat{S}(t_{(j-1)})P(T > t_{(j)} \mid T \geq t_{(j)})$$

An alternative expression for the estimated survivor function is found if we substitute for the survival probability $\hat{S}(t_{(j)})$, and is given in the Kaplan-Meier product limit formula:

$$\hat{S}(t_{(j)}) = \prod_{i=1}^{j-1} P(T > t_{(i)} \mid T \geq t_{(i)})$$

Using information from the LoS data such as the number of patients that have each distinct LoS, $m_{(j)}$, and the number of patients that could have left at each LoS time, $n_{(j)}$, the Kaplan-Meier estimate is calculated as:

$$\hat{S}(t_{(j)}) = \frac{n_{(j)} - m_{(j)}}{n_{(j)}}$$

In order to fill the B matrix, the conditional probability of each patient leaving the hospital is evaluated on each day after surgery, and the number of patients remaining in hospital is updated. The conditional probability of failure, $L(t_{(j)})$, is the probability that the event (patient leaves hospital) occurs in a small time interval of length h after time t, and is defined as:

$$L(t_{(j)}) = P(t < T < t + h \mid T > t)$$

$L(t_{(j)})$ can be estimated when finding the Kaplan-Meier estimate of $S(t)$ as follows:

$$\hat{L}(t_{(j)}) = \frac{m_{(j)}}{n_{(j)}}$$

i.e.

$$\text{Cond. prob. of leaving hospital on day } d = \frac{\text{No. of patients leaving on day } d}{\text{No. of patients in hospital at start of day } d}$$

The B matrix is generated using the conditional probability of failure estimate, $\hat{L}(t_{(j)})$, according to Algorithm 6.

Algorithm 6 Generation of the B Matrix using the conditional probability of failure

for each column j **do**
  Look up the specialty plan j refers to and the number of new arrivals that will require a bed (newArrivals).
  for each session i **do**
    if $a_{ij} = 1$ **then**
      Enter the number of new arrivals in the row in the B matrix that corresponds to the weekday of the operating theatre session.
      Let remainingArrivals = newArrivals.
      for Day $d = 1$ **to** maximum LoS for this specialty **do**
        for $k = 1$ **to** remainingArrivals **do**
          Generate a random number, $r \in [0, 1]$.
          if $r >$ conditional probability of leaving hospital on day d **then**
            Decrease remainingArrivals by 1.
          end if
        end for
        Update B matrix with number of remaining arrivals on this day
      end for
    end if
  end for
end for

Generating the B Matrix: Method 3 – Expected Bed Count

By sampling from the LoS distributions to generate the B matrix as in Method 2, the observed LoSs are being anticipated. Here B is generated using a non-anticipatory method so that nothing about the patients' LoS (distribution) is assumed.

The principle behind this expected bed count approach is to determine how many patients are expected to require a bed on day d after surgery, given a particular schedule. Using the theory of survival analysis on empirical distributions of LoS data, the expected number of beds required on day d after surgery is calculated as follows:

Let e_d be the expected bed requirement on day d after surgery. Let p_d be the probability of a patient staying in a bed from day d to the next day, day $d+1$.

Then,

$$\text{Expected bed requirement on day } d \text{ after surgery, } e_d = e_{d-1}p_{d-1}$$

where $e_0 = N =$ number of patients having surgery during an operating theatre session, and who will later require a bed on a ward.

The probability of a patient staying in a bed from day d to the next day is found using the conditional probability of failure as described in Method 2, i.e. the probability that a patient leaves the hospital on day d, in the following way:

$$p_d = 1 - \text{conditional probability of failure on day } d$$

Two counters of expected bed count are used for the generation of the B matrix: the true, fractional form of e_d, and the corresponding rounded value. It is important to use the fractional form in the calculations of expected bed count to avoid rounding error; however, the rounded value is required as it represents a whole number of beds for use in the model.

4.3 Simulation of an Optimal MSS

The second stage of the modelling process involves the simulation of an optimal schedule obtained from the previous optimisation stage. The simulation provides a number of output measures that can be used to evaluate how well an optimal schedule might perform if implemented.

The simulation is performed by producing a snapshot of future bed requirements for each ward. Future bed requirements are generated using the same method as in Method 2 in Section 4.2.4; the conditional probability of leaving hospital on each consecutive day after surgery. This method of simulation is equivalent to generating a B matrix, using the plans of the optimal schedule to form an A matrix as the input required for generating a B matrix. It is important to note that the B matrices generated as part of the simulations are different to those generated as inputs to the optimisation problem. Typically, 1000 simulations of an optimal MSS are performed, unless specified differently later in the thesis.

Performance measures concerning an optimal MSS that are obtained from the simulation include whether more beds are required than available on at least one day on one ward, the average number of these violations (over all wards on all days) and the expected bed shortage. The expected bed shortage is the total number of beds required in addition to those available on all wards on all days. Further details and discussion of these results are given in Chapter 5.

4.4 Model Validation

The remainder of this chapter concerns the validation of the developed model. An important stage of the modelling process is to verify that the model is valid and represents the real-life system accurately. Robinson [134] discusses how validation is a process of increasing confidence in a model to ensure that the model is sufficiently accurate for purpose for which the model is to be used. This is particularly relevant to models applied to heathcare problems as there is a high degree of uncertainty associated with many hospital systems, and it is important to develop models that the end-users feel are sufficiently accurate for their purposes. Model credibility is also concerned with developing in the end-users the confidence in the information derived from the model [141].

The aim of validation of the developed optimisation and simulation model is to establish whether what is currently happening in UHW can be modelled. That is, the current MSS being used can be modelled correctly with relation to the resulting bed requirements. This will instill confidence in potential end-users of the model so that it can be used to determine whether improvements can be made to the MSS that will reduce the number of cancellations and level demand for beds on wards.

There are a number of aspects to the optimisation and simulation model that can be investigated for validation. It would be expected that the throughput of surgical inpatients in the model, i.e. the number having operations, would be similar to that observed in the data from UHW. This ensures that we are not planning the MSS based on too few or too many patients over the planning horizon. This affects the bed requirements modelled for each ward, and so needs to be compared to the observed bed count from the UHW data. Checking the bed count ensures that the number of patients and their LoSs are similar to that observed in UHW. All validation is performed based on the current MSS used in UHW, and with the data from UHW for the year 2012/13.

4.4.1 Baseline Scenario

A baseline scenario for the problem to be modelled is required in order to assess the model and to compare with the results obtained from parameter variation. The baseline scenario is defined by the values of parameters that are used within the optimisation and simulation model and are chosen to reflect the current set-up at UHW, such as the number of operating theatres and number of days in the MSS.

Parameter values are either based on knowledge of the current practices at UHW, or from the extensive data analysis as discussed in Chapter 3. A summary of the parameters and their values used in the baseline scenario is given in Table 4.3.

Parameter	Value in baseline scenario
No. of operating theatres	14 operating theatres
No. of days in MSS	5 days a week
No. of sessions per day	2 sessions per day
No. of specialties	18 specialties
No. of sessions required	Same number as current UHW MSS for each specialty
No. patients per session	From 2012/13 UHW data (see Section 4.4.2)
Pre-operative LoS	Every patient has pre-operative LoS = 1 day
Post-operative LoS	From 2012/13 UHW data (see Section 3.5.2), based on specialty
No. of wards	11 wards
Wards used by specialties	From 2012/13 UHW data (see Section 3.4.1)
No. of beds	As in UHW (see Section 3.4.1), constant over all days.
Allowable ward transfers	Only allowed from the Colorectal to the General/Liver ward
CCU mortality	Current mortality rate (10%) from 2012/13 data

Table 4.3: Parameter Values for the Baseline Scenario

4.4.2 Number of Operations per Session

The number of operations per session, as calculated by dividing the duration of surgical procedures into the total session duration, was found in Section 3.3.3. Since the number of operations per session has a big influence on the demand for beds on the wards, the calculated number of operations is validated with respect to total patient throughput and simulated bed requirements in the model. A more detailed discussion of the validation of these two measures are given in Sections 4.4.3 and 4.4.4 respectively.

The number of operations per session for the baseline scenario is found by rounding the calculated number of operations as in Section 3.3.3. Adjustments are made as necessary in order to improve the total patient throughput and simulated bed count in the model to better match the observed numbers in UHW. The number of operations per session used in the baseline scenario is given in Table 4.4.

Specialty	**Calculated operations per session**	**Operations per session (rounded)**	**Baseline scenario operations per session**
Cardiac	1.16	1	1
CEPOD	2.77	3	4
Colorectal	0.89	1	1
ENT	1.35	1	1
General	2.10	2	3
Liver	1.47	1	2
Neurosurgery	1.94	2	2
Ophthalmology	2.05	2	1
Oral	0.92	1	1
Paeds ENT	2.10	2	3
Paeds General	1.39	1	1
Paeds Trauma	1.72	2	2
Renal	2.06	2	2
Scoliosis	3.42	3	3
Thoracic	0.98	1	1
Trauma	2.98	3	5
Urology	2.19	2	2
Vascular	1.20	1	2

Table 4.4: Number of operations per session used in the baseline scenario

The rounded calculated number of operations per session is used for all specialties apart for CEPOD, Ophthalmology and Paediatric ENT for which the number of operations have been adjusted in order for the total number of patients in the model to be better aligned with observed data. It has also been adjusted for the General, Liver, Trauma and Vascular specialties so that the simulated bed count better reflected that of the observed bed count in 2012/13.

4.4.3 Patient Throughput

In order to validate the throughput of patients in the operating theatres, the number of patients operated on in the model is compared to the observed number of operations performed in UHW in 2012/13 for each surgical specialty. In the model, the number of operations per session controls how many patients enter the system. This data was extracted from the UHW 2012/13 data by calculating how many surgeries could be included in an operating theatre session, based on the average length of operations for each specialty (discussed in Section 3.3.3). It reflects the number of operations that *could* be performed given the operating theatre time available; however, it does not take into account the cancellation rates observed over the same period which will be considered subsequently.

The total number of surgical patients expected to be seen over a year in the model, assuming a 50 working-week year, is calculated for each specialty and is shown in Table 4.5. The number of sessions per week for each specialty is found from inspection of the current UHW MSS.

Specialty	**Operations per session**	**Sessions per week**	**Sessions per year**	**Operations per year**
Cardiac surgery	1	20	1000	1000
CEPOD	4	10	500	2000
Colorectal	1	8	400	400
ENT	1	5	250	250
General	3	8	400	1200
Liver	2	2	100	200
Neurosurgery	2	20	1000	2000
Ophthalmology	1	2	100	100
Oral surgery	1	6	300	300
Paeds ENT	3	1	50	150
Paeds General	1	8	400	400
Paeds Trauma	2	2	100	200
Renal	2	6	300	600
Scoliosis	3	4	200	600
Thoracic	1	8	400	400
Trauma	5	12	600	3000
Urology	2	10	500	1000
Vascular	2	8	400	800

Table 4.5: Number of planned operations in the model

Since the number of operations per session are calculated based on the length of time for surgery within each session, cancellation rates need to be considered in order to better reflect reality. The cancellation rates for 2012/13 and the corresponding adjusted number of planned operations in the model are shown in Table 4.6.

Specialty	**Planned operations per year**	**Cancellation rate (%)**	**Adjusted planned operations per year**
Cardiac	1000	15.3	847
CEPOD	2000	0.8	1984
Colorectal	400	15.9	336
ENT	250	25.4	187
General	1200	28.4	860
Liver	200	24.1	152
Neurosurgery	2000	15.7	1685
Ophthalmology	100	17.5	83
Oral surgery	300	16.9	249
Paeds ENT	150	20.6	119
Paeds General	400	13.2	347
Paeds Trauma	200	17.8	164
Renal	600	29.5	423
Scoliosis	600	0.0	600
Thoracic	400	12.4	351
Trauma	3000	33.0	2011
Urology	1000	26.6	734
Vascular	800	0.0	800

Table 4.6: Number of planned operations in the model adjusted for cancellation rates

The absolute percentage error can be used to compare the adjusted number of planned operations with the observed number of operations for the year 2012/13. The absolute percentage error for each specialty is given in Table 4.7, and is calculated as follows:

$$\text{Absolute percentage error} = \frac{|\text{No. planned from model - No. observed}|}{\text{No. observed}} \times 100\%$$

Specialty	**Adjusted number of planned operations from the model**	**Observed number of operations (2012/13)**	**Absolute percentage error**
Cardiac	847	731	15.9
CEPOD	1984	2391	17.0
Colorectal	336	270	24.5
ENT	187	309	39.6
General	860	806	6.7
Liver	152	132	14.9
Neurosurgery	1685	1019	65.4
Ophthalmology	83	104	20.6
Oral surgery	249	310	19.6
Paeds ENT	119	165	27.8
Paeds General	347	566	38.7
Paeds Trauma	164	273	39.8
Renal	423	136	210.9
Scoliosis	600	411	46.0
Thoracic	351	395	11.3
Trauma	2011	1673	20.2
Urology	734	592	23.9
Vascular	800	403	98.5

Table 4.7: Percentage error between the adjusted planned number of operations from the model and the observed number of operations

As can be expected, there is some variation between the number of adjusted planned operations predicted by the model and the observed operations. This can be attributed to the various uncertain aspects of planning operations. The number of operations per session may not always be constant for all sessions throughout the year, hence some variation in the total number observed might occur. Another reason that could cause this difference is that the number of sessions available per week in the MSS to each specialty may have changed during the year 2012/13, again causing the number of actual operations to be different than expected. However, having discussed this with hospital management this is not believed to be the case.

Most specialties have an absolute percentage error of less than 25%. Specialties that have particularly high percentage errors are Renal (210.9%), Vascular (98.5%) and Neurosurgery (65.4%). The Renal specialty has the second highest cancellation rate (29.5%) observed in 2012/13. This is explained by the nature of the Renal specialty: the uncertain nature of transplant surgery means that scheduled surgery time is often not utilised because of a lack of transplant organs. The throughput in the model for Vascular is almost twice as many patients as was observed in 2012/13. However, when bed count is considered in Section 4.4.4, the bed requirements produced in the model by having two operations per session is almost exactly what was observed in 2012/13. Two operations per session is deemed appropriate to use in the model, since it is important to match the modelled bed requirements to that observed. There are almost two-thirds too many operations in the model than observed in 2012/13 for the Neurosurgery specialty. As found in Section 3.3.3, the average total time for a neurosurgery operation is 216 minutes which implies that, on average, two operations can fit into a 420 minute session. Therefore, two operations per session will still be used for Neurosurgery.

4.4.4 Comparison of Predicted and Observed Bed Count

The combination of the input data concerning the number of operations per session and the LoS data needs to be validated to ensure it is being used correctly in the model to produce the bed requirements for the B matrix. In order to validate the bed requirements, the simulated bed requirements for the current UHW MSS are compared to the observed bed count on the wards from the 2012/13 UHW data.

The simulated bed requirements in the model will be considered similar to the observed bed count when the simulated bed requirement lies within two standard deviations of the observed mean bed count. By applying the Central Limit Theorem to the observed bed count data, 95% of the data falls within two standard deviations either side of the mean. Of course, there will be fluctuations in the bed count throughout the week in the simulation, so the mean simulated bed requirement throughout the week will be compared to the range around the observed mean found from the data. Results on the simulated bed requirements for the current UHW MSS and the corresponding observed bed count for each ward is given in Figure 4.4.

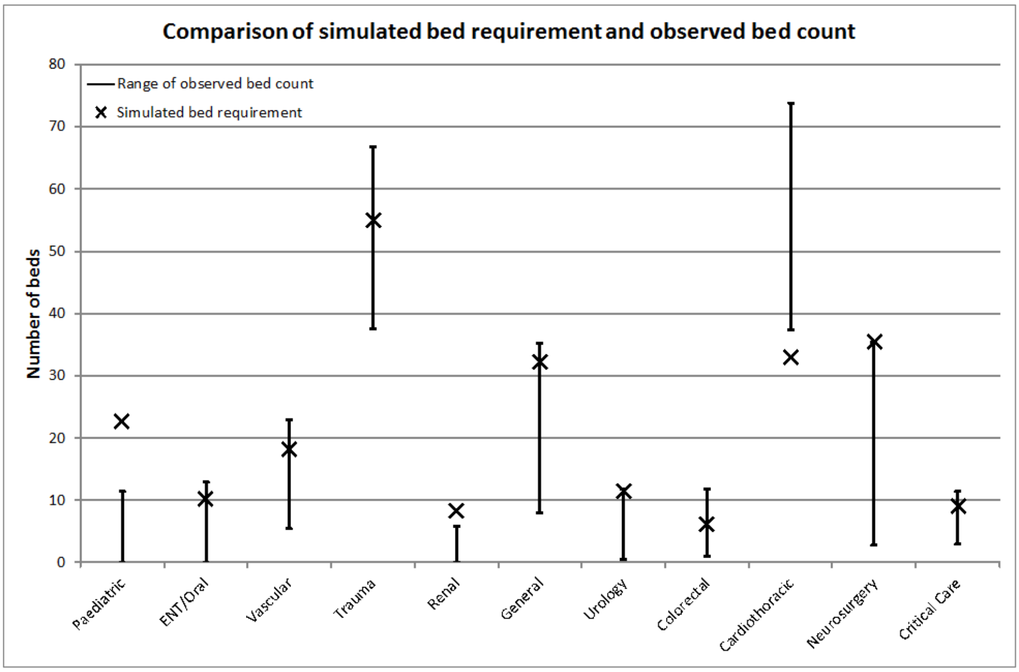


Figure 4.4: Comparison of simulated bed requirement and observed bed count in UHW in 2012/13

It can be seen in Figure 4.4 that the mean simulated bed requirement lies within the upper and lower bounds of observed bed count for most wards. The bounds represent two standard deviations above and below the mean observed bed count. This is not the case for the Paediatric, Renal and Cardiothoracic wards. The mean simulated bed requirement for the Renal ward is slightly above the upper bound of the observed data, although throughout the week the fluctuations in the simulated values do lie within the observed range. Therefore, the simulated bed requirements for the Renal ward are considered to be acceptably close to the observed data.

The simulated bed requirements for the Paediatric and Cardiothoracic wards, however, never lie within the observed range. The simulated bed requirement for the Cardiothoracic ward is always lower than the observed bed count range. 95% of the observed bed count for the Cardiothoracic ward lies between 37.3 and 73.8, whereas the simulated bed requirement fluctuates between 4.3 and 30.3 throughout the week. It could be possible that the number of beds on the Cardiothoracic ward was higher at some point during 2012/13 and has since reduced to the current 37 beds being used in the model. There could also have been data input errors where patients were recorded as being on the Cardiothoracic ward, but were actually outlying on other wards. Considering there are 37 beds available on the Cardiothoracic ward, the simulated bed count for inpatients seems safe, as this allows for spare beds to be available for emergency patients on the ward.

The simulated bed requirement for the Paediatric ward is always higher than the observed bed count, which ranges between 0 and 11.5 beds. The simulated bed requirement ranges between 13.1 and 30.5 throughout the week, which is closer to the 28 beds actually available on the ward. Hence is it deemed that the simulation for the Paediatric ward is in reasonable agreement with the observed data.

The differences between the simulated bed requirements and observed bed count might be attributable to the uncertain occurrence and number of unplanned and emergency surgical inpatients. Figure 4.5 shows how the beds were used in the year 2012/13 by planned and unplanned surgical patients in each surgical ward in UHW.

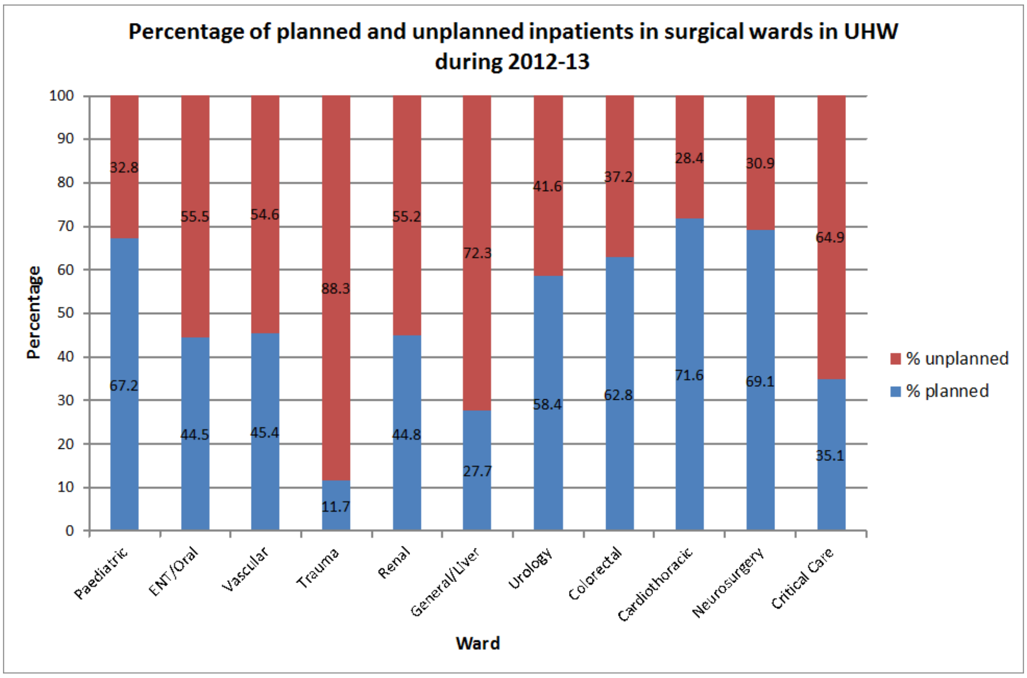


Figure 4.5: Use of beds by planned and unplanned patients in surgical wards in 2012/13

As can be seen in Figure 4.5, the beds on the Paediatric, Urology, Colorectal, Cardiothoracic and Neurosurgery wards were used by a majority of planned inpatients during the year 2012/13. On all other wards, the majority of beds were used by unplanned inpatients. This is understandable for the Trauma ward, as the trauma specialty primarily deals with emergency patients who have broken bones. There is also uncertainty associated with the occurrence of unplanned patients inherent in the nature of the ENT/Oral, Vascular, Renal and General surgical specialties. CCU beds are also dominated by unplanned patients, probably due to the fact that it receives the majority of its patients from the General/Liver specialty and the CEPOD theatre, as discussed in Section 3.4.1.

The maximum number of simulated beds required over a week by planned surgical inpatients on each ward for the current UHW MSS is shown in Figure 4.6. The maximum number of simulated beds is shown by the black sections of the bar chart, and the number of unused beds by the grey sections. The total number of physical beds available on each ward can therefore be interpreted as the total height of the bars. The bar for the Peadiatric ward is slightly different due to there being more beds required than available.

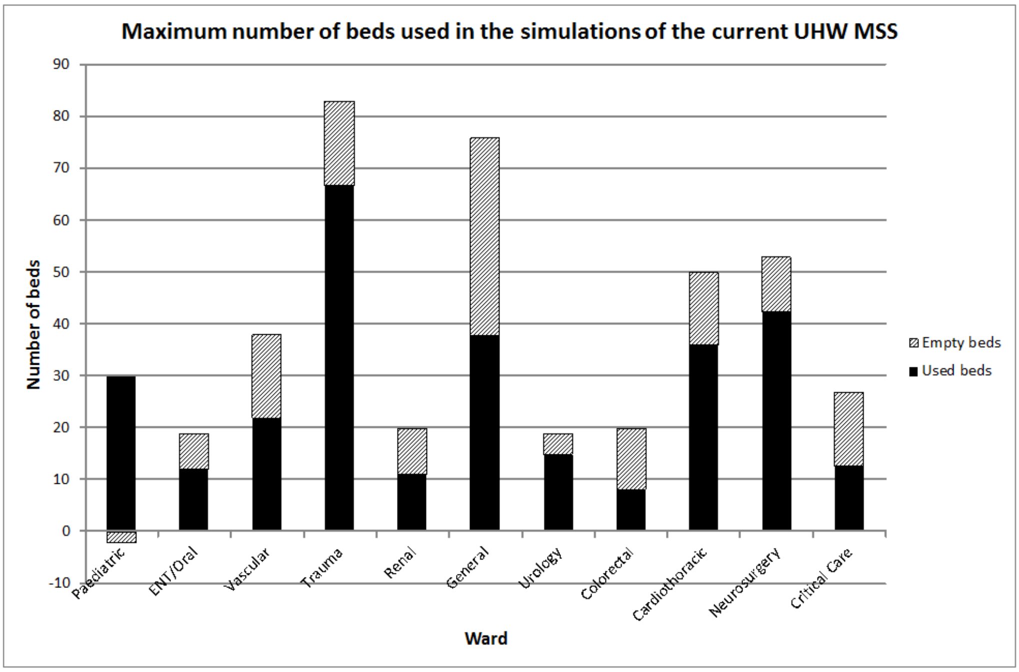


Figure 4.6: Maximum number of beds used in simulations of the current UHW MSS

The percentage of beds used by planned surgical inpatients in the data and simulation is given in Table 4.8. The simulation agrees with the data for all wards apart from the ENT/Oral, Vascular, Trauma, Renal and Colorectal wards. A higher proportion of planned inpatients occur in the simulation than the data for all of these wards apart from the Colorectal ward. Hence, a greater number of inpatients are being handled in the model than were observed in the data, so any optimal schedule produced from the model would be able to cope with more planned patients than observed in past data.

	Percentage of beds used by planned inpatients	
Specialty	**In 2012/13 data**	**In simulation of current UHW MSS**
Paediatric	67.2	107.4
ENT/Oral	44.5	64.3
Vascular	45.4	57.8
Trauma	11.7	80.4
Renal	44.8	55.7
General	27.7	49.5
Urology	58.4	78.5
Colorectal	62.8	40.3
Cardiothoracic	71.6	71.9
Neurosurgery	69.1	79.8
Critical Care	35.1	46.9

Table 4.8: Comparison of beds used by planned inpatients as observed in the 2012/13 data and in the simulation of the current UHW MSS

4.5 Summary

This chapter has introduced the set partitioning based optimisation model that has been developed for the construction of an MSS. An overview of the SPP and solution methods was provided in Section 4.1. As we have discussed, a set partitioning based model seems appropriate for this scheduling problem due to the combinatorial nature of the problem, together with the benefit of being able to choose an optimal schedule from a selection of possible schedules that are defined within the optimisation model.

The proposed model has been developed and explained in Section 4.2. The model aims to find an optimal MSS that minimises the number of unused bed days over the planning horizon, subject to constraints on both the operating theatres and demand for beds on the wards. The developed model is a deterministic model, since a 'snapshot' of bed demand is used to form the bed constraints. LoS data can be used to generate this bed demand input in three different ways: assuming each patient requires a bed for the mean LoS of their specialty, using techniques from survival analysis to calculate the conditional probability of each patient leaving hospital on each day after surgery, and by calculating the expected bed count on each consecutive day after surgery. The possibility of bed transference between wards is also present in the bed constraints so that the model better reflects reality.

An essential part of any model development is the validation and verification of the model to ensure that the model is accurate and represents reality as best it can. Validation of the model was performed in Section 4.4 by comparing the patient throughput and bed count that was observed in the data and predicted by the model. A baseline scenario is also defined for later use of comparison with experimental results. These experiments and results are presented in Chapter 5.

Chapter 5

Results of the Deterministic Model for the MSS

This chapter discusses the results produced by applying the deterministic model developed in Chapter 4 to the current situation at UHW. It then investigates the effects on the MSS by varying a subset of the model's parameters. A number of 'what-if' scenarios of potential interest to the managers of UHW are also investigated.

5.1 Optimisation of the Baseline Scenario

The chosen optimisation software for this research is Xpress-MP. Inputs for the set partitioning based optimisation model are generated in Java and passed to Xpress-MP for automatic optimisation. Optimal values of the decision variables are then passed back to Java so that simulations of the optimal MSS can be performed.

Performance measures that will be used to assess the optimal schedules across the different parameter experiments include the optimal objective function value, the percentage of simulations that have more beds required than are available, and the average number of these violations per simulation. The expected bed shortage for an MSS is also investigated. This is determined from the simulations of the MSS and is the total number of beds required in addition to those available on all wards on all days in the model. This performance measure has been used previously by Beliën and Demeulemeester [19] and is of interest here because it corresponds to the number of cancelled operations that would be expected based on the MSS under investigation. Hence a lower expected bed shortage is sought.

Finally, a desirable feature of an MSS is that the demand for beds on wards is levelled throughout the week. The number of simultaneous sessions, i.e. surgical sessions occurring at the same time but in different theatres, that each specialty has in the MSS is therefore inspected in order to investigate if the demand has been levelled.

5.1.1 Feasibility of the Current UHW MSS

It is important to investigate whether the current MSS used in UHW is a feasible solution of the developed model. That is, is the current UHW MSS feasible with respect to the bed requirements for each ward in the hospital? It is also of interest to determine whether or not a different MSS can be found that would result in fewer cancelled elective operations given the shortage of beds currently experienced in UHW.

When only the operating theatre constraints are considered in the optimisation, i.e. only the A matrix is in the optimisation model, the current UHW MSS was found to be a feasible solution. This is not surprising as there are far fewer constraints in the optimisation model. Managers at the hospital have also commented that the operating theatres are not the main hospital resource causing cancelled operations, but rather bed availability is the cause.

It is also of interest to investigate whether the current UHW MSS remains a feasible solution when the B matrix is taken into account. This is achieved in our optimisation model by forcing the variables that correspond to each specialties' current schedule to be chosen. For this experiment, 1000 B matrices were constructed using the conditional probability of failure approach discussed in Section 4.2.4. The optimisation failed to find the current UHW MSS as a feasible solution in any of the 1000 instances. Hence, it can be inferred that the current UHW MSS used in UHW is not a feasible solution to the scheduling problem when bed constraints are taken into account. This supports the opinion of the hospital managers and findings from the hospital data that the availability of beds on wards can greatly influence the scheduling of operations.

Despite the current UHW MSS not being a feasible solution, the performance of the current UHW MSS can still be investigated to determine whether there are particular problem areas that cause the current UHW MSS to become infeasible when taking into account bed constraints. In order to do this, the current UHW MSS was simulated 1000 times and the simulated bed count inspected. The simulated bed requirement on each ward on each day of the week is shown in Figure

5.1.

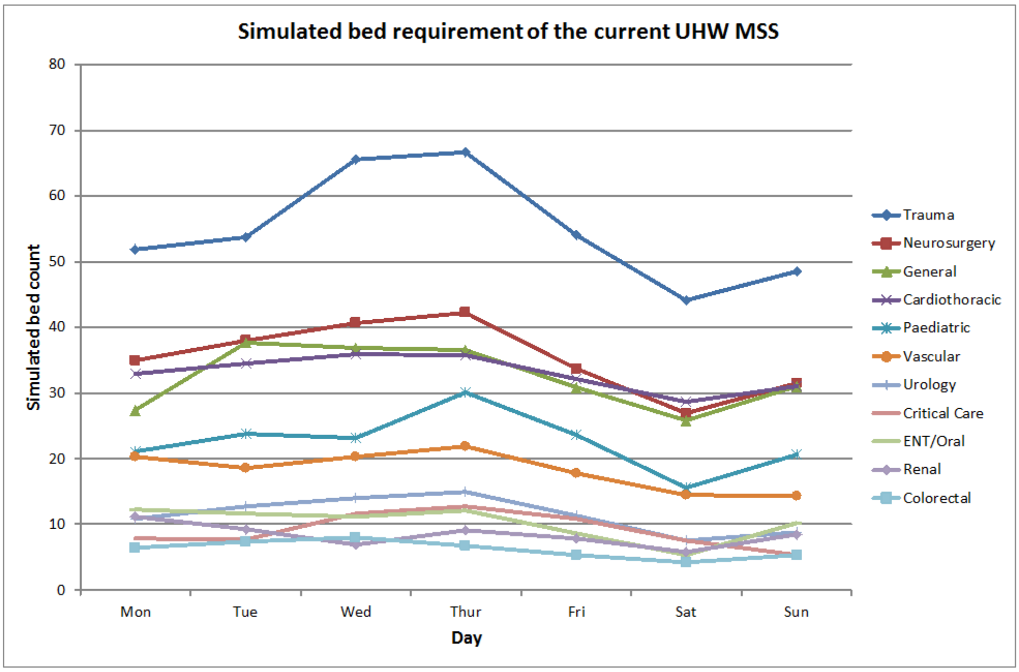


Figure 5.1: Simulated bed requirement for the current UHW MSS

As can be seen in Figure 5.1, the simulated bed requirements follows a similar pattern on all wards: higher bed requirements in the middle of the week, and lower bed requirements during the weekend. Some wards are busier than others, but this is due to the nature of the specialties using those wards, and the number of operations they perform each week.

It was found that 66.8% of the simulations of the current UHW MSS had at least one violated bed constraint. Figure 5.2 shows how many violated bed constraints were in each simulation of the current UHW MSS, providing an indication of how frequent the violated bed constraints are in the simulations. It can be seen that most frequently only one bed constraint is violated in the simulations, and very rarely are four or more constraints violated.

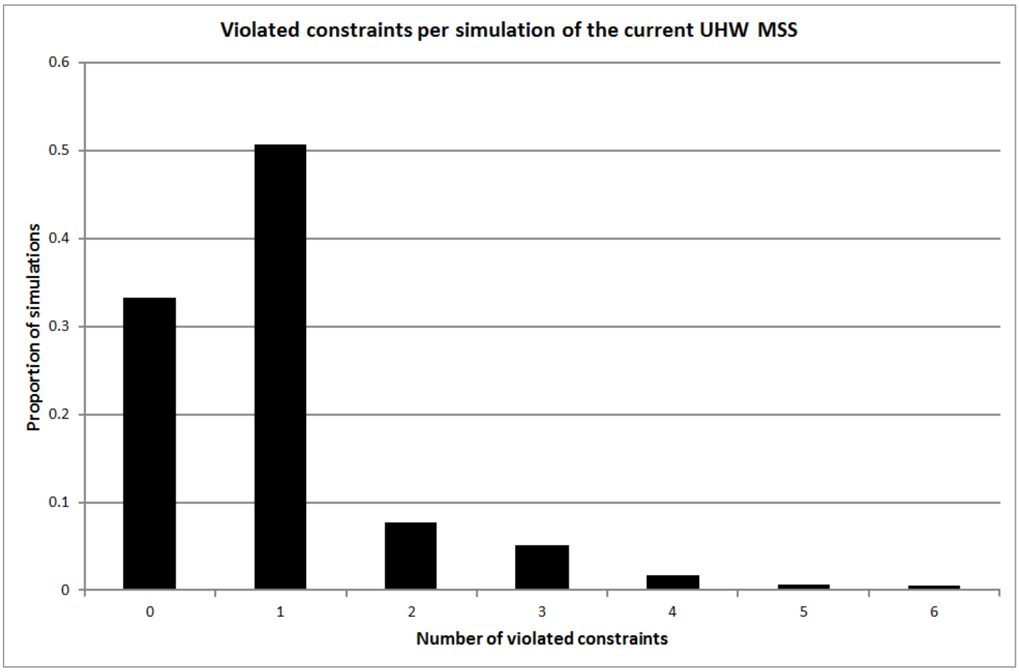


Figure 5.2: Violated bed constraints in the simulations of the current UHW MSS

The total expected bed shortage is on average 3.3 beds (standard deviation = 0.3). To gain an insight into which wards in particular experience an expected bed shortage, the simulated bed requirements are compared to the number of beds available on each ward, i.e. the RHS of the bed constraints. Across the simulations, the average number of additional beds required on each ward is given in Table 5.1. It can be seen that the problem of requiring more beds than available primarily occurs on the Paediatric ward, but also very rarely on the ENT/Oral, Urology and Neurosurgery wards.

	Average expected bed shortage						
Specialty	**Mon**	**Tue**	**Wed**	**Thur**	**Fri**	**Sat**	**Sun**
Paediatric	0.03	0.21	0.17	2.62	0.17	0	0.03
ENT/Oral	0.02	0.01	0.01	0.01	0	0	0
Vascular	0	0	0	0	0	0	0
Trauma	0	0	0	0	0	0	0
Renal	0	0	0	0	0	0	0
General	0	0	0	0	0	0	0
Urology	0	0	0	0.01	0	0	0
Colorectal	0	0	0	0	0	0	0
Cardiothoracic	0	0	0	0	0	0	0
Neurosurgery	0	0	0.01	0.01	0	0	0
Critical Care	0	0	0	0	0	0	0

Table 5.1: Average expected bed shortage on each ward for the current MSS

As can be seen in Table 5.1, the Paediatric ward has, on average, the highest expected bed shortage throughout a simulated week. The Paediatric ward is now examined in more detail in order to determine the causes for the high level of expected bed shortage. It can be seen in Table 5.1 that the simulated bed requirement is higher than the number of beds available on all days apart from Saturday. The average simulated number of patients requiring a bed in the Paediatric ward is shown in Figure 5.3.

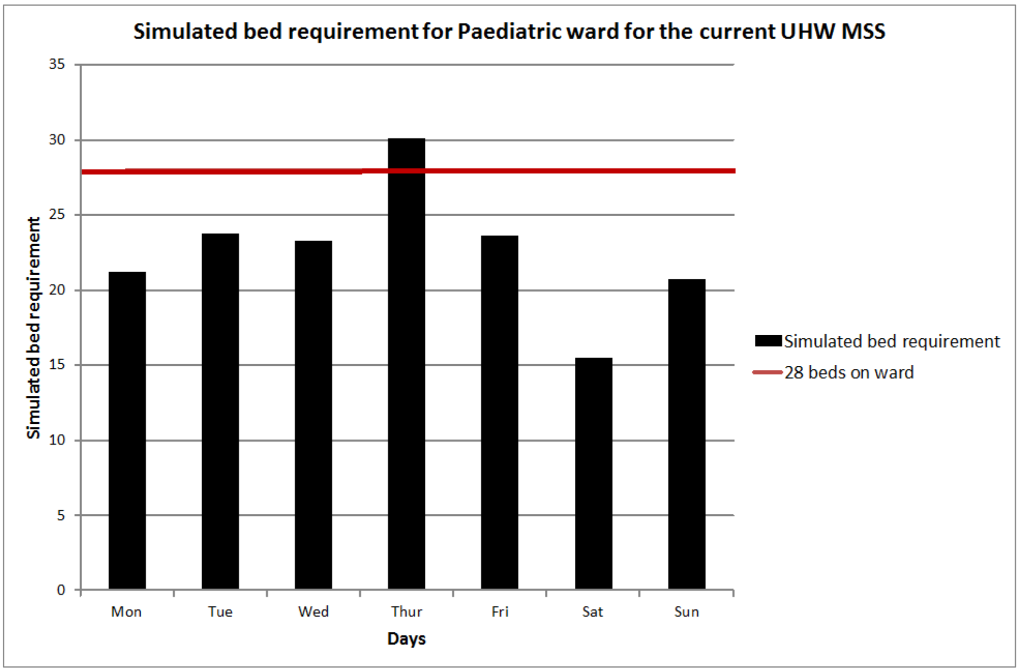


Figure 5.3: Simulated bed requirement for the Paediatric ward for the current UHW MSS

The simulated bed requirements on the Paediatric ward is below the number of physical beds available (28) on all days apart from Thursday. On Thursday, approximately 2 more patients require a bed than are available on the ward. The Paediatric ward receives patients from the four paediatric surgical specialties and the CEPOD theatre. Figure 5.4 shows when the demand for beds on the Paediatric ward is produced from each paediatric surgical specialty. This demand includes pre- and post-operative stays in the ward. As can be seen in Figure 5.4, Thursday is the day on which the demand for beds on the Paediatric ward is the greatest, i.e. the five sessions in the MSS are contributing to the high demand on the Paediatric ward on Thursday.

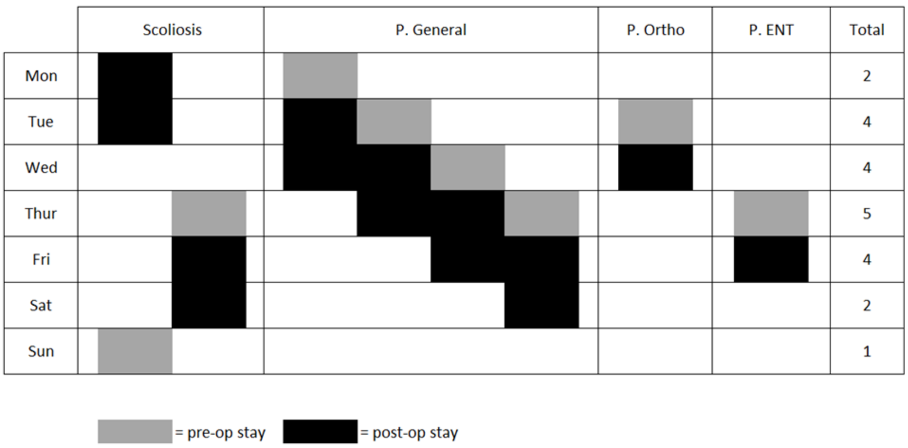


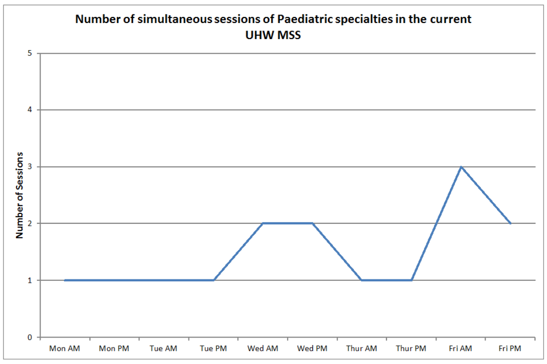
Figure 5.4: Demand for beds on the Paediatric ward in the current UHW MSS

The demand for pre- and post-operative stays on beds in the Paediatric ward from sessions in the current UHW MSS is consistent with the simulated bed requirement, in that there is excessive demand for beds on the ward, particularly on Thursday. This indicates that the current UHW MSS results in a demand for beds on the Paediatric ward that is higher than the number of beds available on the ward due to the combined influx of patients from five specialties.

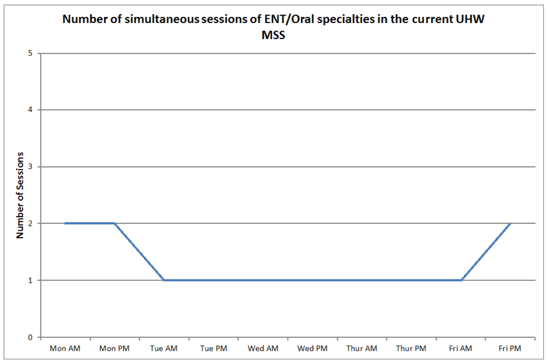
It is of interest to investigate when peaks in the number of simultaneous sessions occur in the MSS because this corresponds to an influx of patients that will be sent to the ward for post-operative recovery. Experiencing a high demand from patients could result in not being able to find a feasible solution to the optimisation problem since the bed constraints would not be satisfied. In reality, this would result in cancelled operations which is undesirable.

The Pediatric ward is one of five wards in which beds on a ward are shared between specialties and in which bed contention has been identified as a particular problem. These wards often experience high demand for beds because there are multiple surgical specialties in the MSS that send their patients to these wards on the same or near similar days. Since these wards have been identified as 'pinch-points' in the system, schedules will be investigated in terms of the number of sessions that are scheduled simultaneously that result in patients going to these shared wards.

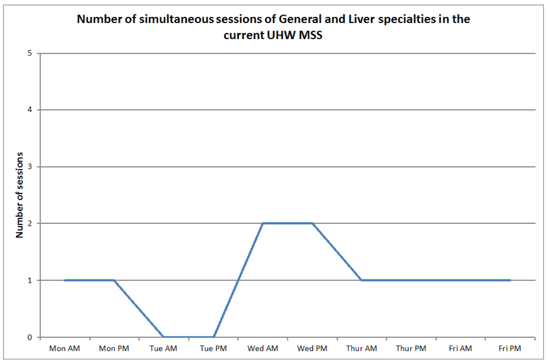
Figure 5.5 shows the number of sessions in the current UHW MSS that are scheduled simultaneously throughout the week for the specialties that send their patients to shared wards.



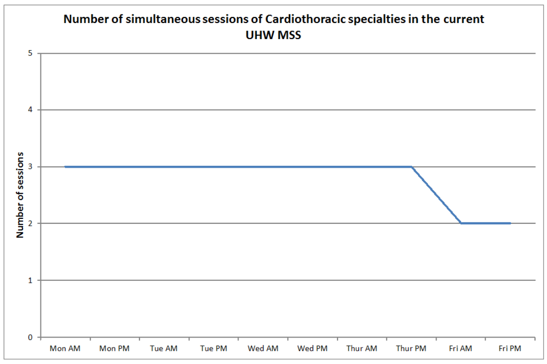
(a) Paediatric ward



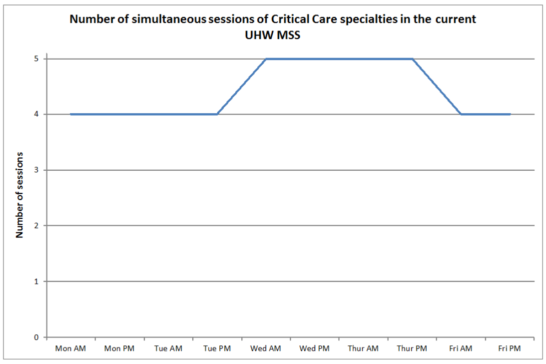
(b) ENT/Oral ward



(c) General/Liver ward



(d) Cardiothoracic ward



(e) CCU

Figure 5.5: Number of sessions that are scheduled simultaneously in the current UHW MSS

In Figure 5.5(a), we see that there is always one scheduled session that results in patients being sent to the Paediatric ward, apart from peaks on Wednesday where there are two simultaneous sessions, and on Friday when there are either two or three simultaneous sessions. There does not appear to be a cyclic pattern across the week, suggesting that the MSS has not been constructed to take into account the typical LoS of patients on the Paediatric ward. If this had been considered, a cyclic pattern would be expected with a cycle peroid similar to the average LoS of the Paediatric ward. Figure 5.6 illustrates the expected cyclic pattern and the relationship between when sessions are scheduled in the MSS and the demand it generates for beds on the wards.

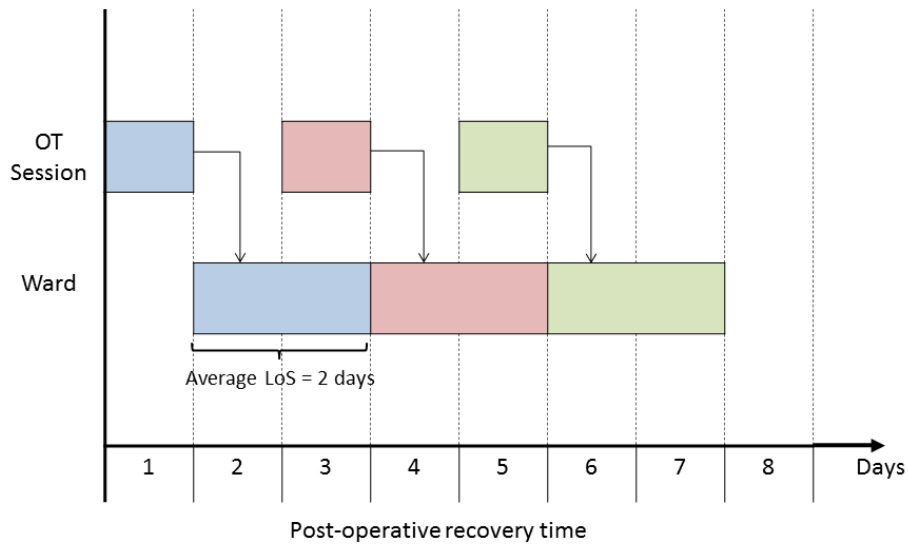


Figure 5.6: Expected cyclic pattern between operating theatre and ward

The number of simultaneous sessions that result in patients going to the ENT/Oral ward is shown in Figure 5.5(b). There is no real trend evident across the week, though there are more simultaneous sessions on Monday and Friday than during midweek. This does not correspond to the average LoS of 2.4 days for the ENT/Oral ward. This also suggests that the current UHW MSS was not constructed with respect to the LoS of patients on these wards.

The number of simultaneous sessions for the General/Liver ward (Figure 5.5(c)) appears to be the most erratic out of all of the shared wards. There is no trend or cyclic pattern throughout the week. For the majority of the week there is only one session; however, there are no sessions on Tuesday, followed by a peak of two sessions on a Wednesday.

The Cardiothoracic ward receives patients from the Cardiac and Thoracic specialties. The Cardiac specialty has exclusive use of two operating theatres in the MSS, so there are always two simultaneous sessions of cardiac surgery throughout the week in the MSS. The only variation in Figure 5.5(d) is attributable to the Thoracic specialty because it is more flexible as to when it can be scheduled, since it is constrained to one theatre and only requires eight sessions per week. There is no real trend or cycle in the graph of simultaneous sessions for the Cardiothoracic ward.

Similarly, the CCU also receives patients from specialties that have a fixed number sessions per week: Trauma and Neurology always have one and two simultaneous sessions throughout the week respectively. Hence the variation in this graph is attributable to the General and Vascular surgical specialties. There is no trend or cycle evident in Figure 5.5(e), however there is a peak in the number of simultaneous sessions on Wednesday and Thursday.

In summary, it appears that sessions for specialties using these shared wards have not been scheduled in the current UHW MSS in such a way that the peaks in the number of simultaneous sessions are spread evenly throughout the week. In particular, they have not been scheduled with respect to the average LoS of each ward. It has been shown that the current UHW MSS is not a feasible solution when bed constraints are considered due to an excessive demand for beds on (some) wards. This influx in patient demand appears to be particularly relevant to shared wards, and is attributable to the number of simultaneous sessions of specialties that send their patients to these shared wards.

5.1.2 Optimal Baseline MSS

In this section, the baseline scenario will be optimised with respect to the operating theatre and bed constraints. Results from the three methods of generating the B matrix from the LoS data, as discussed in Section 4.2.4, are compared.

B Matrix Generated Using the Mean LoS

The mean LoS, rounded to the nearest whole day, used in the model is shown in Table 5.2.

Specialty	**Mean LoS (days)**	**Rounded mean LoS used in model (days)**
Cardiac	7.7	8
CEPOD	3.2	3
Colorectal	3.9	4
ENT	1.2	1
General	3.8	4
Liver	3.1	3
Neurosurgery	3.8	4
Ophthalmology	0.4	1
Oral	1.7	2
Paeds ENT	1.0	1
Paeds General	1.4	1
Paeds Trauma	1.0	1
Renal	3.5	4
Scoliosis	1.5	2
Thoracic	4.9	5
Trauma	4.2	4
Urology	2.5	3
Vascular	4.6	5
Critical Care	3.0	3

Table 5.2: Mean length of stay used to generate the B matrix

In these experiments, no feasible MSS solutions were found using this method of generating the B matrix. From inspection of the B matrix generated for each instance, it was established that this is because there are simply too many beds that are required on each ward. By assuming every patient requires a bed for the same pre-defined time (the mean LoS), there is no staggered departure from the wards allowing new patients to replace them. Instead, there is an influx of patients that all require a bed all at the same time. In reality some patients have a shorter than average LoS, thus freeing up beds for new patients to use. Hence, it can be concluded that this approach for generating the B matrix is not useful for optimising the MSS, since it does not take into account the variance of the patients' LoSs.

B Matrix Generated Using the Expected Bed Count

The number of operations per session for the expected bed count uses data on the number of planned operations per session in UHW, i.e. the number of operations that were intended to happen during each session, regardless of whether or not they actually took place or were cancelled. This is seen as a non-anticipatory approach, since it cannot be known before a session which, if any, operations will be cancelled. By using data from non-cancelled operations per session, it is thought that an MSS can be generated that allows for the desired number of operations to take place and is constructed in such a way that all of the patients' demand for beds can be satisfied.

An optimal solution is found when the optimisation is not restricted to the current UHW MSS. A comparison of the optimal solution and the current MSS used in UHW is provided in Table 5.3. Figure 5.7 shows how prevalent the violated constraints are in the simulations of the optimal MSS. It is much worse than for the current UHW MSS (shown in Figure 5.2) since it is more likely to have a larger number of violated constraints.

Result	**Current MSS**	**Optimal MSS**
Optimal value	984	1285
Simulations with violated bed constraints	66.8%	99.7%
Expected bed shortage	3.3	23.5

Table 5.3: Comparison of the current and optimal MSS using expected bed count

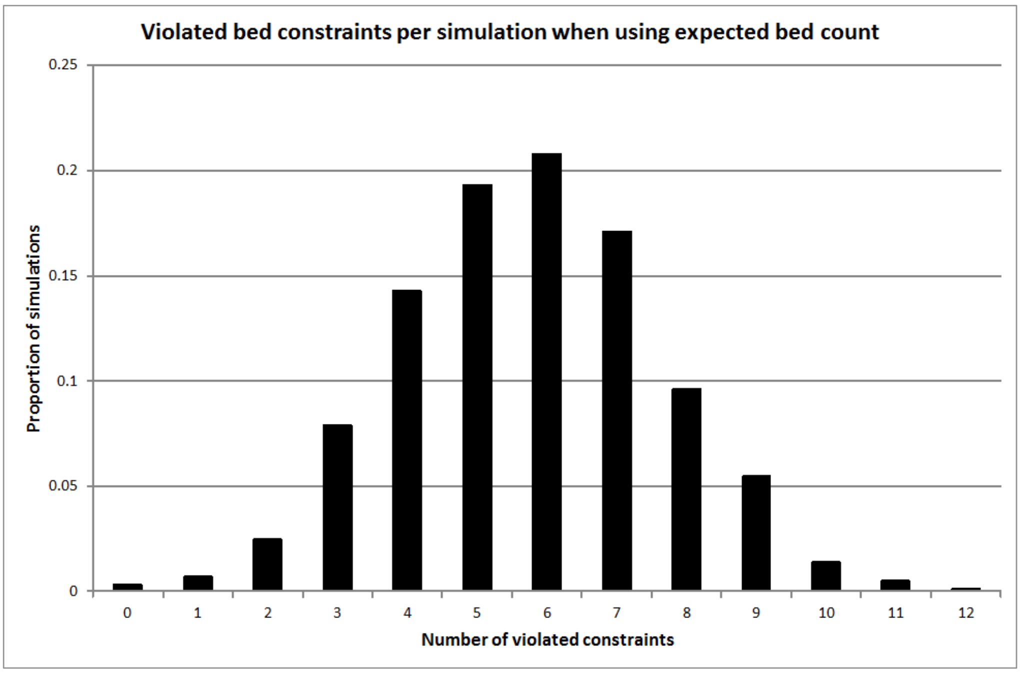


Figure 5.7: Violated bed constraints when using expected bed count in baseline scenario

It is clear from these output measures that the optimal MSS found using the expected bed count to generate the B matrix does not perform as well as the current UHW MSS, despite it being a feasible solution.

B Matrix Generated Using the Conditional Probability of Failure

In this approach, 1000 instances of the problem were used, and each resulting optimal schedule was simulated 1000 times. The average optimal objective function value was 1175.6 unused bed days, with a standard deviation of 46.0 days. On average, 55.6% of the simulations had at least one violated bed constraint.

It is of interest to look in more detail into the prevalence of violated bed constraints. Figure 5.8 shows the proportion of violated bed constraints in each simulation of each run for the baseline scenario. It can be seen that, most frequently, no constraints are violated, and very rarely are six or more constraints violated. This is an improvement over the current UHW MSS for which one constraint was most frequently violated in the simulations.

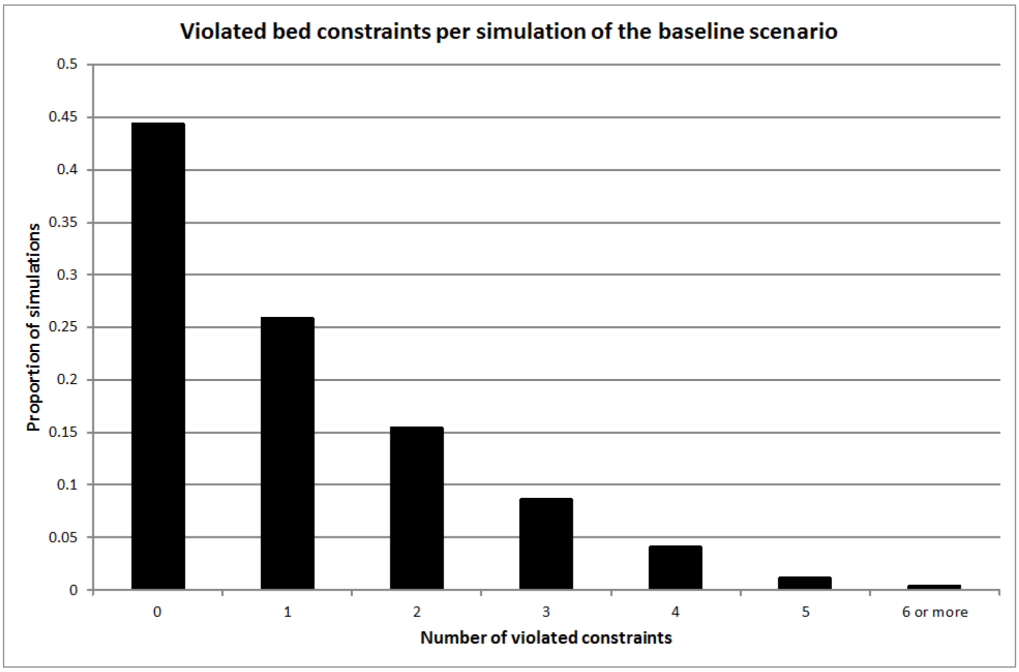


Figure 5.8: Violated bed constraints in the baseline scenario

The average expected bed shortage across the simulations for all instances is 3.0 beds, with a standard deviation of 1.2. On average over all of the instances considered, this is a slight improvement from the results of the simulations of the current UHW MSS, for which the expected bed shortage is 3.3 beds. It is clear that some MSSs exist that will decrease the expected bed count. This indicates that by optimising the schedule with respect to the bed constraints, the expected bed shortage can be reduced.

Table 5.4 shows the simulated number of beds required in the baseline scenario, averaged over all instances. This takes into account different optimal schedules that are generated for different instances of the baseline scenario.

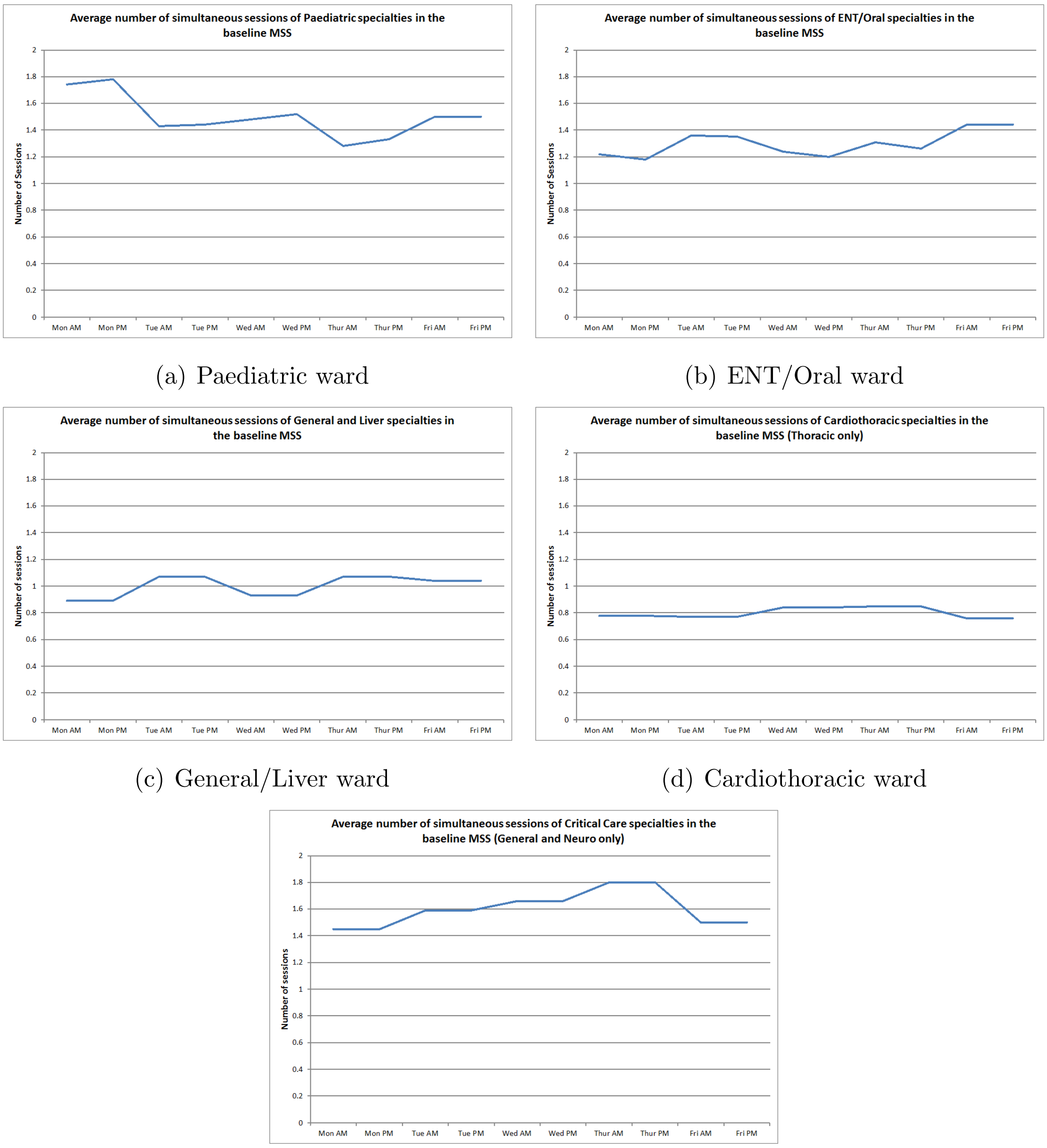
	Average bed requirements							
Specialty	**Mon**	**Tue**	**Wed**	**Thur**	**Fri**	**Sat**	**Sun**	**Beds available**
Paediatric	24.5	25.4	25.6	26.9	20.2	14.1	20.6	28
ENT/Oral	11.2	12.3	13	13.2	8.4	5.2	8.5	19
Vascular	18.1	19.5	20.9	21.7	17.3	14.4	16.5	38
Trauma	54.7	58.4	61.9	64.5	52.9	43	48.9	83
Renal	8.3	9.2	9.8	10.1	8	6	6.8	20
General	32.5	34.5	35.9	37.4	29.8	25.2	30.4	76
Urology	10.9	12.7	14.1	14.9	11.3	7.5	8.8	19
Colorectal	6	6.5	7	7.4	6.1	4.8	5.4	20
Cardiothoracic	32.6	33.7	34.8	36.3	32.7	29.3	31.2	50
Neurosurgery	35.1	38.1	40.5	42.4	33.8	26.8	31.5	53
Critical Care	7.9	9.3	10.4	11.2	11.7	8.1	5.7	27

Table 5.4: Average simulated bed requirement for the baseline scenario

It can be seen in Table 5.4 that for all wards, for every day of the week, the number of beds required does not exceed the number available. This is a definite improvement over the current UHW MSS.

Figure 5.9 shows the average number of sessions in the baseline scenario, over all instances considered, that are scheduled simultaneously throughout the week for the specialties that send their patients to shared wards.

Some theatres are given the same assignment of specialties in all instances due to the scheduling rules: namely the Trauma, CEPOD, Urology, Cardiac and Neurosurgery specialties are always scheduled in the same theatres in the same sessions. These specialties are therefore ignored in the comparison of optimal schedules.



(a) Paediatric ward

(b) ENT/Oral ward

(c) General/Liver ward

(d) Cardiothoracic ward

(e) CCU

Figure 5.9: Average number of sessions that are scheduled simultaneously in the baseline scenario

The number of simultaneous sessions that result in patients going to the Paediatric ward, as shown in Figure 5.9(a), decreases throughout the week. In addition to this downward trend, there are slight peaks in the number of simultaneous sessions on Monday, Wednesday and Friday. The average LoS for the Paediatric ward is 1.95 days, which is approximately equal to the period of this cyclic pattern. Hence there is evidence to suggest that the MSS for the baseline scenario has been constructed with respect to the LoS distributions of patients on the Paediatric ward. The Paediatric sessions are spaced throughout the week so that there is enough time to discharge patients between high influxes of patients requiring beds on the Paediatric ward. In comparison with the number of simultaneous paediatric sessions for the current UHW MSS (Figure 5.5(a)), by including bed constraints in the optimisation model, there is an improvement in the spacing of these sessions throughout the week. This helps to avoid peaks in demand for beds on the Paediatric ward, and when peaks do occur, the elapsed time between peaks is roughly the same as the average LoS on the ward.

Figure 5.9(b) shows the number of simultaneous sessions that result in patients going to the ENT/Oral ward. There does not appear to be any strong trend in the graph, though there is a slight increase in the number of simultaneous sessions towards the end of the week on Thursday and Friday. There is also no obvious cyclic pattern throughout the week, though there are slight peaks on Tuesday and Thursday/Friday. The average LoS on the ENT/Oral ward is 2.4 days, so this could be an explanation for these slight peaks because they are separated by approximately 2 – 2.5 days. Similar to the results obtained for the Paediatric ward, this could provide evidence that the model is constructing the schedule for the ENT/Oral specialties with respect to the LoS on the ward. This graph for the baseline scenario is markedly flatter than the graph for the current UHW MSS (Figure 5.5(b)). The ENT/Oral sessions in the baseline MSS are more equally spaced throughout the week, and the differences in peaks and troughs in demand have been reduced.

From visual inspection, there could be a slight increasing trend in the number of simultaneous sessions throughout the week for the General/Liver ward, as shown in Figure 5.9(c). There is one peak on Tuesday and possibly another on Thursday; however, the graph is much flatter than the graph for the current UHW MSS (Figure 5.5(c)), indicating that the demand for the General/Liver ward has been leveled over the week. Since there is only one certain peak on Tuesday, this would indicate a cycle length of 5 days, which is similar to the average LoS on this ward which is 5.4 days. Hence this also provides evidence that the optimisation model is spreading the simultaneous sessions throughout the week based on the average LoS for this ward.

The graphs for the Cardiothoracic and the CCU wards are shown in Figures 5.9(d) and 5.9(e) respectively, and do not differ markedly from the graphs for the current UHW MSS. The specialties that use these wards and that have a fixed schedule during the week have been removed from these graphs in order to show the variation of the specialties that can be scheduled at different times of the week; hence the Cardiothoracic graph only shows the simultaneous sessions for the Thoracic specialty, and the CCU graph only shows the simultaneous sessions for the General and Vascular surgical specialties. It can be seen that there is little variation in the number of simultaneous sessions throughout the week in both graphs. There is a peak in the Cardiothoracic graph that is spread over Wednesday and Thursday, and the graph for the CCU slowly increases throughout the week to a peak on Thursday, after which it decreases rapidly on Friday. Both the Cardiothoracic wards and CCU have an average LoS that is longer than 5 days, making it hard to determine whether or not a cycle exists in the graphs. The Thoracic specialty has an average LoS of 5.7 days, which could correspond to the wide peak evident in Figure 5.9(d) over Wednesday and Thursday. The combined average LoS of the General and Vascular specialties is 8.8 days, which is much longer than the duration of the MSS. It could be argued that the slowly increasing peak in Figure 5.9(e) corresponds to a long average LoS, however, it is not possible to confirm this from our results.

In summary, it appears that there may well be a relationship between the cycle length in the graphs of simultaneous sessions and the average LoS for each ward. This is especially evident from the graphs for the Paediatric and ENT/Oral wards, though it becomes much more speculative for the other shared wards for which the average LoS is longer than five days, making it difficult to determine from inspection of the graphs. The average LoS and corresponding cycle length determined from the above graphs for each of the communal wards in the optimal MSS's obtained in the baseline scenario is shown in Figure 5.10.

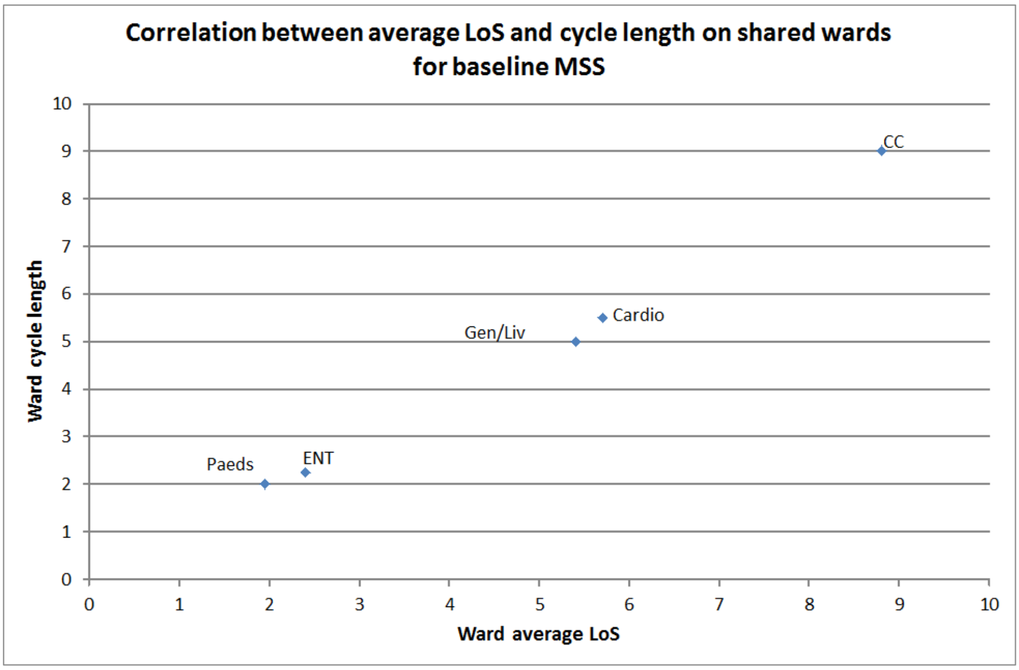


Figure 5.10: Correlation between the average length of stay and cycle length for shared wards in the baseline scenario

In Figure 5.10, there appears to be a positive, linear relationship. As there are only five shared wards to analyse, there are not many data points available to give a very reliable analysis of correlation between the average LoS and the cycle length. Despite this, a Kolmogorov-Smirnov test confirms that both variables are Normally distributed with a high probability, and Figure 5.10 indicates that there is a linear relationship between the two variables and that there are no significant outliers. Based on these assumptions, a Pearson product-moment correlation analysis was run at the 5% significance level. The results indicate that there is a strong, statistically significant, positive correlation between the average LoS and cycle length on the communal wards ($r = 0.997, p < 0.0005$).

Based on the inspection of the graphs in Figure 5.9 and the confirmation of a strong, positive correlation between the average LoS on a ward and the cycle length between peaks for simultaneous sessions in the MSS, there is evidence to suggest that the inclusion of bed constraints serves to level the bed requirements in shared wards. Furthermore, the sessions that send their patients to shared wards are distributed throughout the week with respect to the average LoS for each ward. This leveling of the demand for beds on the shared wards could be a contributing factor to the reduction of the expected bed shortage from the current UHW MSS to the average optimal baseline scenario MSS.

5.2 'What-if' Scenarios

As we have seen, there are many inputs and parameters that are used within the optimisation and simulation model. By having so many parameters, the model is very flexible and can be customised to many different situations or hospitals. Parameters of interest are listed in Table 5.5, together with some examples of experiments that can be performed. A subset of parameters will be chosen to demonstrate the flexibility of the model and to investigate interesting scenarios.

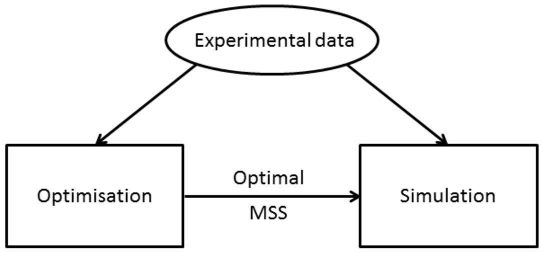
Parameter	**Experiments**
No. days theatres in use	Increased number of days in the MSS
No. sessions per day	Whole day sessions compared to half day
No. patients per session	More/less operations performed during a session
LoS data	Increase/decrease in post-operative LoS by $x\%$
Pre-operative arrivals	Increase/decrease in pre-operative LoS by $x\%$
No. wards	More/less wards open to surgical specialties
Transition matrix (W)	More/less allowable transitions of patients between wards
No. beds available	More/less beds available on the wards

Table 5.5: Experiments associated with model parameters

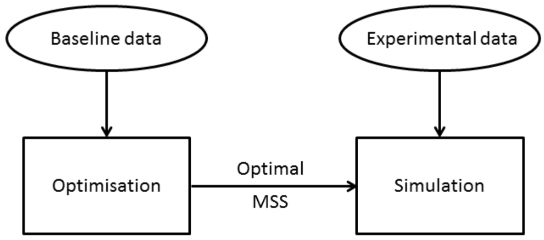
5.2.1 Post-Operative Length of Stay

It is of interest to the hospital to see how changes to the post-operative LoS affect the characteristics of the optimal MSSs. In particular, how will the demand for beds on wards change with varying LoSs. Both an increase and decrease in LoS will be investigated, however, the hospital is expected to be more interested in the results from a decrease in LoS as they are implementing initiatives on wards to reduce LoS and therefore free up beds for more patients.

The model will be used in two ways to investigate how LoS affects demand for beds on the wards. It will first be used to determine whether schedules can be found for different changes in LoS by inputting the experimental data into the optimisation and simulation (Figure 5.11(a)). The second use is to investigate how schedules that were generated with current levels of LoS in the optimisation cope with different LoS using the experimental data in the simulations only (Figure 5.11(b)). The first use is linked to hospital planning; the second use is linked to robustness to see how well an MSS copes with different realisations of bed demand.



(a) Experimental LoS data used in both the optimisation and simulation



(b) Experimental LoS data used in simulation only

Figure 5.11: How different length of stay data will be used in the model

In this case, the number of new arrivals per session is kept the same as in the baseline scenario. The LoS is increased or decreased by the same amount for all wards, except for the CCU. It is not considered appropriate to change the LoS for the CCU because it experiences a very high demand from the most seriously ill patients who are already moved out of the CCU as soon as possible to allow other patients to be treated. Results of 1000 instances using increased or decreased LoS in the optimisation and simulation are shown in Table 5.6.

% change in post-op LoS	**Optimal value (Unused bed days)**	**Average percentage of simulations with bed violations**	**Expected bed shortage (beds)**
-30%	1537.9	19.7	0.5
-20%	1418.6	29.8	1.0
-10%	1293.4	42.6	1.8
Baseline	1175.6	55.6	3.2
+10%	1070.9	69.1	5.3
+20%	954.4	82.3	8.5
+30%	860.5	93.0	14.0

Table 5.6: Results of changing the length of stay in the optimisation and simulation

The results of this experiment indicate that it is still possible to find feasible and optimal schedules when the LoS departs from the baseline scenario. As the LoS decreases, the average optimal objective function value increases, indicating that the number of unused bed days in the system increases. This is as expected, since more bed days will be used if patients stay in hospital for longer. As the LoS decreases, the average percentage of simulations with bed violations also decreases. This is as expected, since the earlier patients leave hospital, the more beds are available for new patients to take and it is less likely that bed capacity will be exceeded. As the LoS decreases, the expected bed shortage also decreases. This is consistent with fewer beds being used if the LoS is shorter.

The above results show that is is possible to construct optimal MSSs if a change in the post-operative LoS is predicted, i.e. the changed LoS data is used in the optimisation and simulation. The ability of an MSS constructed using current levels of LoS as in the baseline scenario to cope with changes in the LoSs is now investigated. The results of using different levels of LoS in the simulation only are shown in Table 5.7.

% Change in post-op LoS	**Average % simulations with bed violations**	**Expected bed shortage (mean)**	**Expected bed shortage (StdDev)**
-30%	14.9	0.4	1.3
-20%	25.8	0.8	2.1
-10%	39.6	1.7	3.3
Baseline	55.6	3.2	5.1
+10%	69.8	5.3	7.1
+20%	82.7	8.6	9.6
+30%	93.5	14.3	12.9

Table 5.7: Results of changing the length of stay in the simulation only

As can be seen in Table 5.7, the average percentage of simulations with bed violations decreases as the LoS decreases. This is as expected, since the shorter a patient's LoS, the sooner they will leave hospital, freeing up a bed for any incoming patients.

The expected bed shortage is also reported in Table 5.7. As the LoS decreases, the average expected bed shortage of all wards decreases, as does the standard deviation of the expected bed shortage. It is not surprising that the mean expected bed shortage increases as the LoS increases, because more patients will require beds for longer, therefore increasing the demand for beds for longer. The increase in the standard deviation of the expected bed shortage as the LoS increases can be explained due to the fact that more patients have longer LoSs causing higher expected bed shortages, but there are also some simulations that do not result in any bed shortages, ie. the expected bed shortage is being stretched from 0 to an increasing maximum, as the LoS increases.

By using the same baseline LoS data as input to the optimisation in all of these LoS experiments, the same optimal schedules will be found for each run of the trials. It is therefore possible to compare the results of the LoS experiments on a run-wise basis as the same schedules are used in the simulations. Matched pairs tests can be used to determine at what level of change of the LoS causes the change in expected bed shortage to become significant.

The expected bed shortage results from all instances of different levels of LoS in the simulations were not found to be Normally distributed at the 5% significance level. A related-samples Wilcoxon signed rank test, the non-parametric equivalent to a matched pairs t-test, is used to compare the difference in expected bed shortage for different levels of LoS. Six pairwise tests were carried out, in each case testing the null hypothesis that there was no difference between the medians of the data on expected bed shortage in the baseline scenario and in the changed LoS scenarios. The Wilcoxon signed rank tests, carried out at an overall significance level of 5%, found that there is a significant difference in the expected bed shortage when the LoS is increased by 10–30% above the baseline scenario and decreased by 10–30% below the baseline scenario. A recommendation can therefore be made to the hospital that if the LoS can be reduced by at least 10% across all wards, then the expected bed shortage will be significantly lower (at least 47% lower) than if the LoS remained at current levels.

5.2.2 Number of Beds Available on Wards

It is of interest to investigate the effects of changing the number of beds available on the hospital wards on the MSS. Here, all parameters are set to the values as in the baseline scenario, but in each experiment the number of beds available on all wards will be altered by the same percentage. It is important to note that for these experiments, the same level of sharing of beds between wards, as specified in the W matrix, is as used in the baseline scenario. The effects of the sharing of beds between wards will be investigated in Section 5.2.3.

The number of beds that have been used in the baseline scenario is the number of actual physical beds on each ward in UHW. This does not take into account the very real possibility that some of those beds might be occupied by patients from other surgical or medical specialties. As learnt from discussions with hospital managers, it is often the case in UHW that beds on surgical wards, that should really be used for inpatients of the assigned surgical specialty, are used by other surgical specialties or outlying medical patients. This occurs when the demand for beds on other wards outweighs the supply on those wards and the patients must be placed on another ward within the hospital. Another factor that can affect how many beds are available to planned inpatients on surgical wards is the uncertain occurrence and number of unplanned and emergency surgical inpatients, as discussed in Section 4.4.4.

The number of physical beds available on each ward was the same as used in the baseline scenario (Section 4.4.1). Optimal schedules were then simulated using a number of different levels of bed availability on the wards. Results for these experiments, when the bed constraints were changed by a certain percentage for all wards, are given in Table 5.8.

% Change in number of beds	**Average % simulations with bed violations**	**Expected bed shortage (mean)**
-20%	99.8	26.6
-10%	89.0	9.3
Baseline	55.6	3.0
+10%	23.0	0.9
+20%	6.3	0.2

Table 5.8: Results of changing the number of beds available in the simulation

As expected, when using the same input data as the baseline scenario for the optimisation, as the number of beds available on the wards decreases, a higher percentage of simulations contain violated bed constraints and the expected bed shortage increases. Table 5.9 shows at what level of bed change each ward experiences a shortage of beds in the simulations. An 'X' indicates that a ward features a shortage of at least one bed in the simulations.

	Amount of beds available on each ward				
Specialty	**-20%**	**-10%**	**Baseline**	**+10%**	**+20%**
Paediatric	X	X	X	X	X
ENT/Oral	X	X	X	X	
Vascular					
Trauma	X	X			
Renal					
General					
Urology	X	X	X		
Colorectal					
Cardiothoracic	X				
Neurosurgery	X	X	X		
Critical Care					

Table 5.9: When wards experience bed shortages in the simulations

It can be seen in Table 5.9 that there are four wards in the baseline scenario that experience a bed shortage: Paediatric, ENT/Oral, Urology and Neurosurgery. When the number of beds are increased on all wards by 10%, only the Paediatric and ENT/Oral wards continue to have a shortage of beds, and only the Paediatric ward continues to have a bed shortage when the number of beds are further increased by 20%. When the number of beds is reduced by 10%, the Trauma ward also then has a bed shortage, and when the number of beds is reduced further by 20%, the Cardiothoracic ward experiences bed shortages in the simulation. This gives an indication of which wards are the most sensitive to a change in the number of beds on the hospital wards. The Paediatric ward always experiences a bed shortage, whereas the Vascular, Renal, General and Colorectal wards and the CCU can all withstand a 20% reduction in the number of beds on their wards.

Figure 5.12 gives more detail on the scale of the expected bed shortages over a simulated week when the number of beds available on the wards is changed. Only wards that experience an expected bed shortage are shown. Again, the Paediatric ward has the greatest expected bed shortage, and the ENT/Oral ward has some bed shortage in the baseline scenario and below, but not on the same scale as the Paediatric ward. Bed shortages only become a sizable problem on the Trauma, Urology and Neurosurgery wards once the number of beds have been reduced by 20% on each ward.

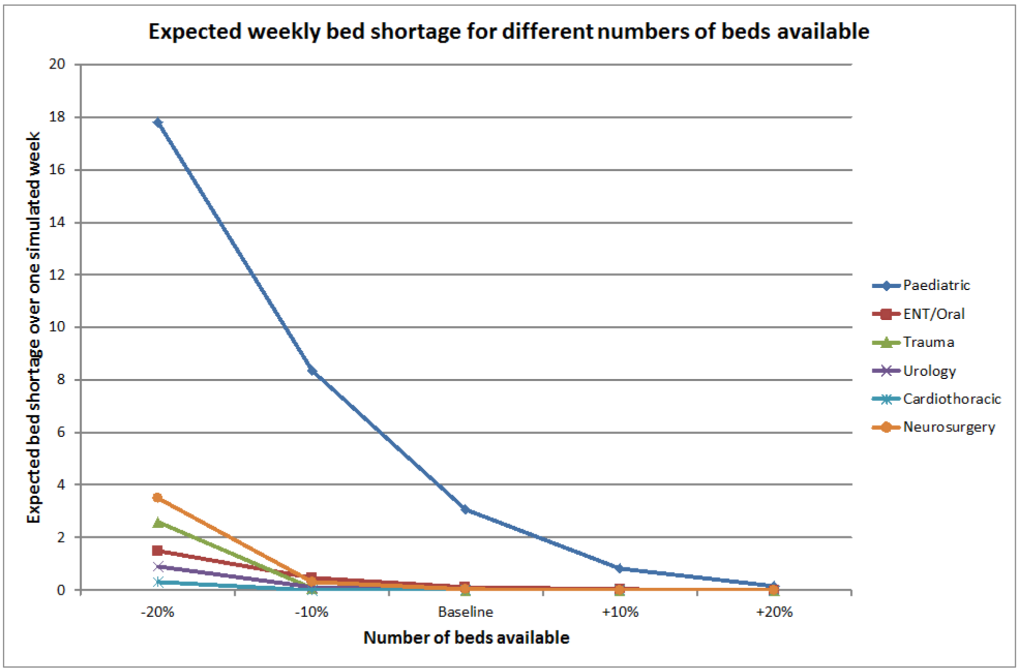


Figure 5.12: Effect of different amounts of beds available on each ward on the expected bed shortage

5.2.3 Sharing of Beds Between Wards

As we saw in Section 4.2, our formulation of the optimisation model incorporates the facility of sharing beds between wards through the use of slack and surplus variables. The sharing is controlled by the W matrix in the formulation by stating which wards each can share with. It is of interest to see how the sharing of beds between wards affects the performance of the resulting MSS.

The baseline scenario involves sharing between a limited number of wards to reflect the reality at UHW. Patients who should be sent to the Vascular ward are allowed to be put on the Cardiothoracic ward, and patients whose home ward is the Colorectal ward can go to the General surgery ward. This sharing is based on current practices in UHW and the closely related nature of the surgical procedures of these specialties.

Two experiments will be performed based on the allowed sharing of beds in the optimisation model. The first will investigate the effect of absolutely no sharing of beds between wards. This scenario is based on the premise that specialties in the hospital could become responsible for all of their resources; from the scheduling of their operations to the management of their beds. For this to happen, no other surgical specialty would be allowed to use the beds of another specialty; beds would be completely ring-fenced and managed within the specialty.

The second experiment will investigate the opposite extreme of bed sharing; allowing the sharing of beds between many different wards. The creation of large pools of beds using wards of specialties of similar surgical nature could be a strategy to even out demand for beds across the hospital. In this scenario, it would make it easier for a specialty to borrow a bed from another specialty if they were experiencing high demand for beds, and vice versa if the demand for beds was reversed. The pools used in the scenario are summarised in Table 5.10. They are based on the similarity of the surgical nature of the specialties, and assuming that the specialties require similar equipment and specially trained nurses for post-operative care. The CCU is not pooled with any other specialty due to the very specialised nature of the care given in that ward. Within each pool, beds are allowed to be shared between every ward within that pool, but not with any other ward outside their pool.

<table>
<tr><td>Cardiothoracic
Vascular</td><td>General/Liver
Colorectal
Renal
Urology</td></tr>
<tr><td>Paediatric
ENT/Oral
Trauma</td><td>Neurosurgery</td></tr>
<tr><td colspan="2">CCU</td></tr>
</table>

Table 5.10: Grouping of the wards to create bed pools

In this case, the Paediatric ward has been pooled with the ENT/Oral and Trauma wards which both cater for adult patients. It is not common practice in UHW for children to be put on adult wards, and this would never happen in reality. The Paediatric ward has been pooled with adult wards here in order to demonstrate that the Paediatric ward requires more beds than it currently has, and that if there is the possibility of the ward sharing beds with *any* ward, then it would benefit the whole system. It is left to the hospital managers to decide which ward would be able to be pooled with the Paediatric ward if this bed pooling strategy was implemented.

In order to investigate the different levels of sharing of beds between wards, the optimisation model will be run using the number of beds on each ward as in the baseline scenario, i.e. the number of physical beds available on each ward. The average slacks and surpluses of the resulting optimal solutions will be inspected to determine how many beds, if any, are shared between wards. This will then be used to inform the simulation model, in which the altered number of beds available on each ward will be used. For example, if it was found from the slacks and surpluses from the optimal solutions that Ward 1 shares four of its empty beds with Ward 2, then Ward 1 will have four more beds available in the simulation of the optimal schedules, and Ward 2 will have four fewer beds available in the simulation.

For the first experiment, when no sharing of beds is allowed, the same optimal solutions as in the baseline scenario were found. This is as expected due to the very restricted W matrix. The values of the slacks and surpluses indicate that the empty beds on all (real) wards were given to the dummy ward in the model as part of the optimal solutions. This is to ensure that the equality bed constraints (4.18d) were satisfied, and has the added property that the beds on each ward are safeguarded from other wards using their beds.

For the second experiment, when extra sharing compared to the baseline scenario is allowed via the use of a number of bed pools, the average optimal objective function value was 1156.3 unused bed days. This is a 1.7% reduction from the baseline scenario in which the average optimal value is 1175.6 unused bed days, suggesting that the additional sharing through the use of bed pools results in more bed days being used, and thus the beds being utilised more. The slacks and surpluses in the optimal solutions also indicate that more beds are being shared between wards than in the baseline scenario. Figure 5.13 shows where beds are shared between wards within bed pools in the optimal solutions. For example, in the Cardiothoracic and Vascular bed pool, beds on the Cardiothoracic ward are used by Vascular patients.

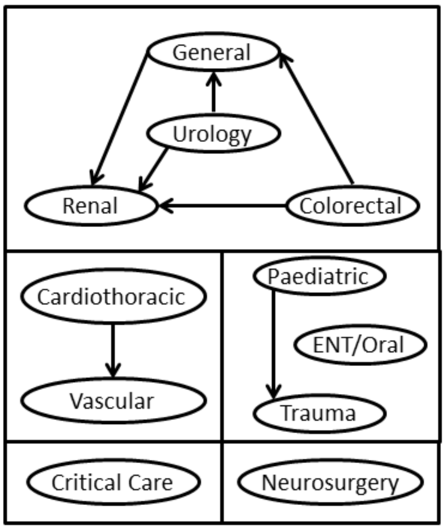


Figure 5.13: Sharing of beds between wards in bed pooling experiments

The average number of empty beds on each ward for this scenario are shown in Table 5.11.

Specialty	**Average number of empty beds**	**Average % of empty beds**
Paediatric	2.9	10.2
ENT/Oral	5.9	30.8
Vascular	17.4	45.9
Trauma	23.6	28.4
Renal	9.9	49.3
General/Liver	38.6	50.8
Urology	7.4	39.1
Colorectal	12.6	62.9
Cardiothoracic	15.6	31.1
Neurosurgery	17.9	33.7
Critical Care	15.4	57.1

Table 5.11: Amount of empty beds when more sharing between wards is allowed

As can be seen in Table 5.11, on average, all wards have empty beds. This is in agreement with the current UHW MSS simulation results in Section 4.4.4. However, the Paediatric ward is the only ward that experiences a shortage of beds and so uses some beds on another ward in the optimisation model. The average number of empty and additionally required beds on the Paediatric ward over a simulated week is shown in Table 5.12. The ward has empty beds on Monday, Friday, Saturday and Sunday, uses all of its beds on Tuesday, but requires additional beds from other wards on Wednesday and Thursday. From inspection of the slack and surplus decision variables, these additional beds were acquired from the Trauma ward. Therefore, additional beds were moved from the Trauma ward to the Paediatric ward in the model in order for all patients in the Paediatric ward to be able to have a bed. However, what would happen in reality is that these additional Paediatric patients would outlie on the Trauma ward. As discussed previously, children would never be put on an adult ward as a matter of policy, but these results are used to illustrate the benefits that could be achieved if the Paediatric ward could be pooled with *any* ward in order to help with its high demand for beds.

	Mon	**Tue**	**Wed**	**Thu**	**Fri**	**Sat**	**Sun**
Beds	1 empty	Full	1 extra	2 extra	5 empty	11 empty	6 empty

Table 5.12: Number of empty and additionally required beds on the Paediatric ward

All empty beds on other wards are given to the dummy ward. This ensures the equality bed constraints are satisfied in the optimisation model, but also has the implication that these empty beds are safeguarded for the sole use of the assigned specialty of each ward.

The results of the experiment using bed pooling is summarised in Table 5.13, along with results from the first experiment and the baseline scenario for comparison. It can be seen that the more sharing that is allowed in the system, the better the system performs. This is shown by the reduction in the percentage of simulations that have bed constraint violations and the average violation per simulation is also reduced. The expected bed shortage is also reduced, indicating that if more beds are allowed to be shared between wards, then fewer cancellations would occur over the surgical specialties as a whole.

Scenario	**% simulations with violated bed constraints**	**Av number of violations per simulation**	**Expected bed shortage (mean, SD)**
No sharing	55.6	1.07	3.19, 5.08
Baseline	55.6	1.07	3.19, 5.08
Bed pools	46.3	0.73	2.14, 3.75

Table 5.13: Summary of results of the bed sharing experiments

5.3 Conclusion

This chapter has concentrated on the analysis of results from the deterministic model for the construction of the MSS which was developed in Chapter 4. Investigation into whether the current MSS used in UHW is feasible with respect to the operating theatre and bed constraints has been performed, followed by and investigation of optimal MSSs under experiments relating to hospital variables.

A baseline scenario was used to reflect the current reality at UHW, and associated investigations were described in Section 5.1. Interestingly, the current MSS used in UHW was not found to be a feasible solution for the optimisation model. Optimal schedules could, however, be found for a less restricted baseline scenario, and were found using the expected bed count and conditional probability of failure methods of generating bed demand for the bed constraints. These optimal schedules were found to perform better than the current MSS used in UHW in terms of a more levelled number of simultaneous sessions across the week, and fewer simulations with violated bed constraints.

A series of 'what-if' scenarios that were chosen to be of interest to hospital managers were investigated in Section 5.2. In particular, it was found that feasible MSSs could be found for different levels of post-operative LoS, different number of beds available on each ward, and different bed pools composed of wards sharing beds.

Note that while the deterministic model presented and investigated in Chapters 4 and 5 provides some important insights into the interplay between the MSS and resulting bed requirements on surgical wards, it does not take into account the stochastic nature of the post-operative bed requirements. Extensions of the model are derived in Chapters 7 and 8 that aim to create more robust MSSs that safe-guard against this uncertainty.

Chapter 6

Optimisation Under Uncertainty: an Overview

As previously discussed, there is uncertainty associated with the resources involved with the scheduling of operating theatres. This chapter includes a literature review on the various techniques that can incorporate uncertainty into the optimisation process.

6.1 Optimisation Under Uncertainty

Traditional optimisation methods of linear programs in the form of Model 6.1, for example, implicitly assume that the parameters for a given problem are known. That is, the coefficients in the objective function, c, and constraints, A and b are known. However, it is not always the case that these inputs to optimisation problems are fixed and known accurately. Uncertainty in the values of the parameters can be due to a variety of reasons, including measurement error or the fact that the parameters represent some information about the future. Examples include the costs of products for the optimal inventory mix and the future demand for a product, which may not be known with certainty.

$$\begin{aligned} \max \quad & \mathbf{c^T x} \\ s.t. \quad & \mathbf{Ax} \leq \mathbf{b} \\ & \mathbf{x} \in X \end{aligned} \tag{6.1}$$

Point estimates or expected values can be used as 'snapshots' of these uncertain parameters in deterministic optimisation methods in order to give an indication of an optimal solution. However, Ben-Tal and Nemirovski [24, 25] have shown through the use of case-studies that small perturbations in these uncertain parameters can result in infeasible solutions. Hence, even if an optimal solution can be found for certain point estimates, they may not remain feasible when the data changes or is not as expected once implemented. This is not appealing to decision makers, since they require solutions in which they can have confidence for highly uncertain problems.

It could even be the case that each instance, or snapshot, of input data results in a different solution being found from the optimisation, as illustrated in Figure 6.1. This situation is also undesirable for decision makers because it is not clear as to which optimal solution is best to choose since it is not known which of the input data will be realised on implementation.

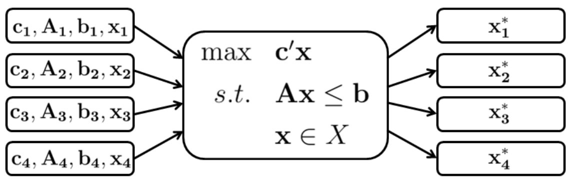


Figure 6.1: Multiple inputs resulting in multiple solutions

Ideally, all possible realisations of uncertainty should be taken into account by the optimisation model to provide a single optimal solution, as in Figure 6.2. The field of optimisation under uncertainty is concerned with this 'black box' method of optimisation. A number of approaches to optimisation under uncertainty have been identified in the literature: stochastic programming that uses the probability distributions of the possible realisations of uncertainty, the use of recourse for making decisions in stages once more data becomes certain, and robust optimisation in which a range of possible values that the uncertain parameters could take is specified.

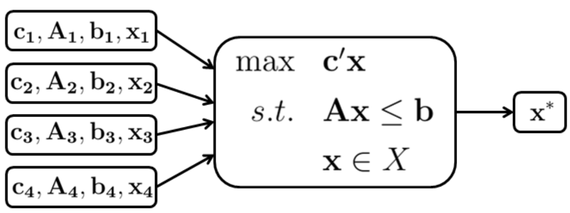


Figure 6.2: Multiple inputs resulting in one solution

By taking into account the possible realisations of uncertainty, either by using their probability distributions in stochastic programming or by using specified ranges of uncertain values, the notion of a 'good' solution may be found. The definition of good, or 'robust', for these solutions is different for each application, and the robust solutions should be more resilient to uncertainty than their non-robust or deterministic counterparts found from traditional optimisation methods.

6.2 Approaches to Optimisation Under Uncertainty

Two broad approaches to optimisation under uncertainty have been identified in the literature; stochastic programming and robust optimisation. A brief overview of both methodologies are given in this section.

6.2.1 Stochastic Programming

Stochastic programming (SP), also referred to as stochastic optimisation, provides a framework for finding solutions to problems that involve uncertainty. Pioneered by Dantzig [54] in 1955, SP uses the fact that the uncertain data in the model can be described by probability distributions in order to find a solution for all (or at least most) of the possible instances of realisations of the data which, in some sense, is optimal. The reader is directed to the textbook by Birge and Louveaux [35] for a comprehensive overview of SP.

A widely used technique in SP is recourse. Multistage stochastic programs with recourse are problems in which some decisions, or recourse actions, are taken once the uncertainty has been disclosed, for example, choosing the product mix when the availability of the resources required to make the products are uncertain [8]. In solving the problem, the mix must be chosen before the uncertain availability is known, then additional decisions (recourse decisions) are made once the uncertainty is realised to adjust for the new conditions. Although two-stage programs are most common, more than two stages can be used depending on the application. Two-stage problems assume that data in either stage can be modelled as a random vector with a known probability distribution. Decisions at either stage are based on the data available at the time, and should not depend on future observations.

If the probability distributions of the random variables are known, then numerical integration is employed over the random continuous probability space. Due to computational difficulties with this approach, it is often assumed that the random data have a finite number of possible realisations, known as scenarios. This discretisation of the probability space helps to solve the two-stage problem numerically since it can be re-formulated as a single linear programming problem. The more scenarios considered in the probability space, the more likely that the actual realisation of uncertainty will be covered in the stochastic program, however, this causes the model size, and hence computational time, to increase rapidly. A number of approaches have been developed in order to smartly choose scenarios, such as using Monte Carlo simulation [143], and to combine a number of scenarios into a reduced number [68].

Chance-constrained programming is also a common technique used in SP. First developed by Charnes and Cooper [50], the approach is based on satisfying the constraints of a linear program up to a pre-specified level of probability. By limiting the probability of constraint violation in this way, solutions are very difficult to find due to the computational intractability of the chance-constrained problem. Another drawback to this approach is that probability distributions of the uncertain parameters are required. Approximations of chance-constrained programs have been developed [126], including the sample average approximation method [99] that replaces the probability distribution of the constraints with an empirical distribution obtained from a random sample of the uncertain parameters.

SP approaches to operating theatre scheduling have been commonly used in the literature. Bruni et al. [40] use a stochastic model with recourse for the scheduling of surgeries when the occurrence of emergency patients and surgery durations are uncertain. Belien and Demeulemeester [19] use SP, via MIP based heuristics, to construct an MSS with resulting levelled bed occupancy when the number of patients and their LoS is uncertain.

The main disadvantages of using the SP approach to optimisation under uncertainty include the requirement of specifying the probability distributions of the possible outcomes at each decision stage. Reliable data is needed to estimate them accurately, which often is not available in practice. If the alternative scenario approach is used for SP, the number of scenarios to generate is unclear. More scenarios will give a more complete picture of the uncertainty, however, this will increase the computational time required in order to solve the problem. Scenario-based approximation methods to SP are discussed in more detail in Section 6.4.

6.2.2 Robust Optimisation

As noted by Gabrel et al. in [81], the term 'robust optimisation' (RO) has assumed different definitions since its conception by Soyster [145] in 1973 and Beyer and Sendhoff [34] note that the term even encompasses several approaches within the application of robust design. The term was first made popular in the 1990s by Mulvey et al. [123], however, their scenario-based robust optimisation model is completely different to that of Soyster, and will therefore not be covered in this review. The original concept of RO proposed by Soyster has continued to be extended since the late 1990s, and will be taken to be the definition of RO herein.

The theory of RO, first proposed by Soyster [145], uses the notion of bounded, convex uncertainty sets to define the nature of the uncertain data. The objective function is then optimised over the uncertainty set, while maintaining feasibility for the worst-case value of the constraints. The use of uncertainty sets is especially useful if a stochastic model of the uncertainty is unknown, rendering SP impossible.

An uncertainty set specifies a set of values that the uncertain data could realise. By optimising over an uncertainty set, the original problem is reformulated and replaced by what is known as its robust counterpart. Probability distributions for the uncertain data are not assumed, though the shape of the uncertainty set must be defined. Uncertainty sets that are defined as having an ellipsoidal form [23, 71], result in the original linear program being transformed into a non-linear robust counterpart. This approach is less conservative than that of Soyster [145], however, the robust counterparts are harder to solve computationally. More recent efforts have focused on the definition of simpler uncertainty sets that preserve the tractability of the original linear program and are thus more computationally efficient [31, 32].

Recent research directions include trying to bridge the gap between RO and SP. Chen et al. [51] provide an RO perspective on SP in which they develop a tractable approximation for multistaged chance-constrained linear programming problems. Bertsimas and Goyal [29] and Bertsimas et al. [28] also apply RO techniques to multistage stochastic problems. Düzgün and Thiele [70] develop an RO approach that describes the uncertainty in objective coefficients using multiple ranges for each coefficient. This approach avoids a very large single range that would be required by the traditional RO model, and which would lead to overly conservative results.

One of the main advantages of RO over SP is that a (full) stochastic model of the uncertain data is not required. When sufficient information on the probability distribution of the uncertain data is not known, RO becomes an attractive alternative due to the relatively simple requirement of an uncertainty set. The selection of an appropriate uncertainty set is an issue which will be discussed further in Section 6.3, however, the benefits that RO provides are generally considered worth it. The fact that RO techniques are available that ensure the tractability of the original problem is maintained in the robust counterpart, and that the degree of conservatism can be controlled are particularly appealing.

6.3 Robust Counterpart Optimisation

6.3.1 Development of Robust Optimisation

Soyster [145] first suggested modelling uncertainty in linear programming problems through the use of bounded, convex sets. Unknown coefficients were assumed to take values from a realistic subset called the uncertainty set, often centered on the nominal values of the unknown coefficients. The model developed by Soyster [145] resulted in each uncertain parameter taking its worst-case value from the uncertainty set. This ensures that the constraints remain feasible for any of the possible realisations of uncertainty (within the uncertainty set), whilst the objective function is optimised with respect to the worst-case realisation. By taking the worst-case value, this approach is at maximum conservatism. In a similar approach developed in 1976, Falk [77] considered uncertainty in the objective function whose coefficients were assumed to lie in a closed, convex set.

Interestingly, this approach to RO was not advanced in the operational research literature until the late 1990s. Soyster's model of using the worst-case scenario was deemed too conservative in practice, since complete protection against the uncertainty often results in severe worsening of the objective function value. As a result of this view, the earlier RO framework of Soyster was extended to consider other forms of uncertainty sets.

Extensions to the original RO model involved the use of ellipsoidal uncertainty sets by Ben-Tal and Nemirovski [23, 24, 25], El-Ghaoui and Lebret [71], and El-Ghaoui et al. [72]. Ellipsoidal uncertainty sets were assumed because the corners (or extremes) of the box representation employed by Soyster [145] were considered unlikely to occur once the uncertainty was realised. Ben-Tal et al. [22] provide an overview of RO using ellipsoidal uncertainty sets and the resulting robust counterparts. This approach reduces the level of conservatism of Soytser's model and tractable reformulations for the robust counterparts can be produced. However, the robust counterparts obtained from this approach are second-order cone problems and are computationally complex.

Due to the drawback of increased computational complexity of the robust counterparts encountered using ellipsoidal uncertainty sets, Bertsimas and Sim [30, 31] developed a methodology that considers the uncertainty set as a polyhedron. Specifically, the uncertainty set is an interval of a range of values that each uncertain parameter can take. An additional parameter is also introduced to each constraint, called the protection level [31] or the budget of uncertainty [32], that the decision maker can use to control the degree of conservatism of the solution by limiting the number of coefficients that can take their worst-case value. This approach preserves the tractability of the nominal problem; the robust counterpart of a linear problem is linear. The reader is referred to Bertsimas et al. [28] for a comprehensive review of RO using different uncertainty sets.

The choice of uncertainty set is not always clear, so Bertsimas and Brown [27] provide a prescriptive methodology for constructing uncertainty sets. The approach of Bertsimas and Sim [31] that uses ranges of realistic values for the uncertainty set is particularly appealing to practitioners due to its simplicity [81]. No special assumptions about the probability distribution of the uncertain data are required, and the intuitive nature of the protection level aids the model's interpretation.

Although in some situations it may be advantageous to be able to control the level of conservatism of the robust solution, as will be discussed in Section 6.3.2, there are many applications in industry, such as robust control theory [67] that deals with bounded system uncertainty, that require a worst-case analysis. A large branch of RO still focuses on this worst case optimisation, as discussed by Ben-Tal and Nemirovski [26].

6.3.2 Recent Developments in Robust Optimisation

A number of interesting advances in RO have occurred in recent years; Gabrel et al. [81] provide a detailed overview of developments since 2007. Key developments include the 'robustification' of stochastic optimisation, using work from risk theory to describe uncertainty sets, and the development of non-linear and multistage RO models.

The 'robustification' of SP tries to bridge the gap between RO and SP by assuming the uncertain parameters belong to unknown probability distributions.

Bandi and Bertsimas [16] propose a new approach to finding tractable methods of analysing stochastic systems using RO. The field of distributionally robust optimisation, in which a robust formulation for SP is constructed using a set of probability distributions that is assumed to include the true distribution of the uncertain parameters [56, 89], is also related to this concept.

The early work in RO focused on static problems in which the values of all decision variables had to be chosen at once. In recent years however, dynamic RO, in which recourse decisions are incorporated in a tractable way into a modelling framework, has seen a rise in popularity. Thiele et al. [149] discuss approaches to RO with recourse, and Assavapokee et al. [11] develop tractable algorithms specific to two-stage robust problems that minimise the worst-case regret. The reader is directed to Düzgün and Thiele [69] for an overview of recent findings in dynamic RO research.

6.3.3 Applications of Robust Optimisation

RO has been shown to be applicable to a wide range of applications due to its flexible framework for dealing with uncertainty in optimisation problems. Applications include inventory management, such as finding robust policies for supply chains that are subject to stochastic demand [33], robust portfolio selection in which stock returns are uncertain [76], and robust unit commitment schedules in the energy sector [144]. The reader is directed to an in-depth review of the different application areas given in Gabrel et al. [81].

RO has been used extensively in different scheduling applications. Lu et al. [108] study the single machine scheduling problem with uncertainty associated with job processing times. The total flow time of jobs is minimised by measuring the schedule robustness as the maximum absolute deviation from the optimal solution in the worst-case scenario. The robust project scheduling problem, in which there is uncertainty associated with activity durations, has an extensive array of literature [10]. Hazir et al. [92] also use the RO approach for robust scheduling of the discrete time/cost trade-off problem often seen in project scheduling.

A well developed application of RO is to airline scheduling. This was also found to be a common application of the SPP in Section 4.1.6. Problems in airline scheduling involve airline fleet planning and airline crew scheduling. Burke et al. [41] use a multi-objective approach to robust airline scheduling that focuses on reliability and flexibility as measures of robustness in real world schedules. The crew pairing problem, where flight and connection times are assumed to vary within an interval, are considered within an RO framework by Lu and Gzara [109]. Simulation experiments were used to confirm that this robust approach led to more robust crew pairing solutions. Gao et al. [82] consider an integrated approach that addresses both the fleet planning and crew scheduling problems simultaneously.

RO Applied to Healthcare Problems

In 2010, Rais and Viana [132] comment that, 'considerably less work appears to have been carried out with potentially promising methodologies', such as RO, rather than more traditional methods of dealing with uncertainty, for applications in healthcare. In a more recent discussion paper on the use of RO in healthcare management, it was noted that the approach due to Bertsimas and Sim [31] has rarely been applied to healthcare problems [6]. Many aspects of decision making in healthcare settings are subject to a high level of uncertainty, and the small number of papers that apply RO techniques to an application in healthcare are discussed here.

Robust appointment scheduling has been investigated by Mittal et al. [122], since the need for well-designed appointment systems is relevant to many aspects of healthcare delivery, from outpatient clinics to scanners. The service times of patients are uncertain, and an RO approach to assigning service slots to patients in advance has been shown to improve the utilisation of expensive personnel and medical equipment, and to reduce the waiting times of patients.

Perhaps more relevant to this research, Meng et al. [118] propose an RO approach to managing hospital beds for both emergency and planned inpatients, and Addis et al. [6] discuss the use of a robust methodology for patient scheduling. Meng et al. [118] use a distributionally robust optimization approach for managing elective admissions to determine the required quotas of elective patients given the unscheduled and urgent nature of emergency patients. The level of uncertainty the admission system can withstand, as opposed to the worst-case performance, is maximised without breaching the expected bed shortfall limit. Simulation of the resulting quotas suggest that improvements to the bed shortfalls can be achieved. Addis et al. [6] do not give details of their model for assigning surgery cases to blocks of operating theatre time when the surgery duration is uncertain, however, they comment on the trade-off between increasing robustness for a reduction in the number of patients scheduled per operating theatre block. Having an increased level of robustness was found to decrease the number of cancellations which improves the quality of the solution from the point of view of the patient.

Denton et al. [58] also use a robust counterpart model to allocate surgeries to operating theatre blocks. The results are compared to those from a two-stage stochastic program with recourse, and conclude favourably that the RO model performs approximately as well as, and is much faster, than solving the stochastic recourse model, whilst having the benefit of limiting the worst-case outcome.

A methodology to construct an MSS when the demand, i.e. the number of patients for each specialty, is considered uncertain from week to week, is developed by Holte and Mannino [95]. Mannino et al. [113] previously worked on an RO model for the construction of an MSS that aims to balance patient queue lengths among the different specialties, and to minimise the likelihood of using operating theatre overtime. Their investigations into robustness found that, in order to gain a more robust schedule, the amount of allowable overtime should increase.

Banditori et al. [17] group patients based on surgery resource requirements and maximise patient throughput taking into consideration patient's surgery due dates. Surgery durations are taken to be uncertain, so RO is used to find solutions that allow for a satisfactory number of surgeries without incurring overtime or excessive cancellations. The resulting MSSs are tested using a simulation model of patient's uncertain surgery duration and LoS. They also present a combined optimisation-simulation approach that allows the fine-tuning of the optimisation model to trade-off robustness and efficiency.

6.4 Scenario-Based Optimisation

6.4.1 Approaches to Scenario-Based Optimisation

Scenario-based optimisation takes its name from the many, often infinite, possible realisations of uncertainty, or scenarios, that are associated with stochastic optimisation problems. A scenario is an instance of an optimisation problem in which the uncertain data realise certain values. Many approaches to optimisation using scenarios exist and are discussed in this section.

A popular approach to optimisation using scenarios was first developed by Calafiore and Campi [44] in which a finite set of constraints are sampled at random from all possible constraint instances to construct the 'sampled program'. This computationally efficient methodology is seen as an alternative to RO. Sampled programs provide a less restrictive framework than RO by requiring that the risk of failure of the solution is small in a probabilistic sense. This scenario-based optimisation technique is also closely related to chance-constrained programming in which constraints are required to be satisfied by at least a certain level of probability. Benefits of using scenario-based optimisation as opposed to chance-constrained programming include not having to assume a probability distribution for the uncertain parameters in the constraints, and it being a computationally tractable methodology.

A key decision in the scenario-based optimisation methodology concerns how many scenarios to include in the scenario program. Several bounds for the number of scenarios have been developed [42, 43, 44, 45, 46] that ensure that a solution is optimal among all but a few constraint instances. 'Tuning parameters' in these bounds allow the decision maker to trade the probability of violation of the omitted constraints for performance. The theory developed for scenario programs initially related to convex optimisation problems, however, Esfahani et al. [74] later extended this approach to non-convex problems. A more in depth discussion of this scenario-based optimisation approach is given in Section 8.1.

Other approaches to optimisation under uncertainty using scenarios as representations of a subset of the realisations of uncertainty include that of Mulvey et al. [123], Berstimas and Brown [27] and Dembo [57]. Mulvey et al. [123] use scenarios within a two-stage SP model and formulate a robust counterpart to find a robust solution. This work bridges the gap between SP and RO, however, it has the same dimensionality issues as RO since the robust counterpart is nonlinear. Bertsimas and Brown [27] try to bridge the gap between RO and scenario-based optimisation by developing a data-driven approach to constructing uncertainty sets for RO based on a finite set of sampled constraints. Coherent risk measures are used to ensure that the optimal solutions remain feasible for all realisations of uncertainty, however, this approach can only be applied to problems with multiple constraints in a constraint-wise fashion. Dembo [57] presents an approach to solving stochastic problems through a series of deterministic sub-problems, each representing a different scenario of the uncertain constraints. All scenario solutions are then combined into a single, feasible policy.

6.4.2 Applications of Scenario-Based Optimisation

One of the first applications of scenario-based optimisation was to the area of robust control design [45]. Indeed, much of the theory developed by Calafiore and Campi was in relation to robust control in which uncertainty is inherent in the feedback from systems. Pagnoncelli et al. [131] use a scenario-based optimisation approach for portfolio optimisation. The uncertain returns on investments are handled using the scenario approach modified for chance-constrained programming problems. They also use a sampling and discarding approach to selecting the scenarios for the scenario program, as developed by Campi and Garatti [47].

Denton et al. [58] apply the sample average approximation method for SP to surgery sequencing and scheduling, assuming that the surgery durations are uncertain. They conclude that scheduling models that consider uncertainty in the surgery durations have the potential to improve operating theatre schedules. The scenario-based optimisation method of Calafiore and Campi [44] does not appear to have been applied to healthcare problems.

6.5 Summary

A review of the literature on RO has revealed the potential that exists in applying this optimisation technique to many different areas that deal with uncertainty. A popular application of RO is to scheduling problems; ranging from machine scheduling to airline scheduling. Of particular interest here is the great potential that has been demonstrated by applying RO to scheduling in healthcare. A handful of papers have used RO on aspects that relate to generating an MSS by taking uncertain surgery duration into consideration. To the best of our knowledge, it is believed that RO has not been applied to the construction of the MSS in the specific case when uncertain patient LoS or uncertain post-operative bed requirements are taken into account. The use of simulation has also been demonstrated to be a useful tool for testing the robustness of the resulting solutions. Development of a RO model for the construction of the MSS is presented in Chapter 7.

Scenario-based optimisation offers an alternative approach to incorporating uncertainty into an optimisation model. This is achieved by including multiple instances, or scenarios, in the optimisation programme. However, a key research question concerns how many scenarios are required to provide a sufficient level of feasibility, without excessive computational complexity. Scenario-based optimisation has been used for a limited number of applications, however, it has been used successfully in the scheduling of individual patients for surgery, showing promise that it can be applied successfully to the construction of the MSS. Development of a scenario-based optimisation model for the construction of the MSS is presented in Chapter 8.

Chapter 7

Robust Optimisation of the MSS

Due to the importance of constructing a 'good' MSS that affects many expensive hospital resources such as beds, staff and the operating theatres themselves, it is desirable to develop a modelling framework that pro-actively guards against the uncertainty inherent in these resources. As discussed in Section 6.3, the methodologies analogous to RO provide a framework to include uncertainty of model parameters in the optimisation process when information on the stochastic behaviour of the uncertainty is unknown.

In particular, the RO approach due to Bertsimas and Sim [31] seems particularly appealing to modellers due to the ability to vary the level of conservatism of the robust solution, while keeping the problem tractable. It also provides scope for using probabilistic bounds of constraint violation which could be important for decision makers.

Due to our application of RO to the construction of the MSS requiring the use of binary decision variables, and because the robust counterparts of ellipsoidal uncertainty sets are non-linear, a polyhedral uncertainty set will be used. Although this approach adds decision variables and constraints to the original problem (via the robust counterpart), the benefits of problem linearity and computational tractability are deemed to outweigh these slight drawbacks. A polyhedral uncertainty set based on ranges of values for the uncertain parameters has a particularly intuitive interpretation that will help hospital decision makers understand the modelling concepts.

7.1 Robust Counterpart Optimisation

The approach developed by Bertsimas and Sim [31] will be used to develop a robust counterpart optimisation framework to construct a robust MSS.

Consider the following deterministic nominal linear optimisation problem:

$$\begin{aligned} \max \quad & \mathbf{c^T x} \\ s.t. \quad & \mathbf{Ax} \le \mathbf{b} \\ & \mathbf{l} \le \mathbf{x} \le \mathbf{u} \end{aligned} \tag{7.1}$$

In this model, data uncertainty is assumed only for elements of $\mathbf{A}$. All other parameters will be assumed to be certain.

Consider row i of $\mathbf{A}$, and let

$$J_i = \text{set of coefficients in row } i \text{ that are subject to uncertainty.}$$

The model of uncertainty that is adopted in this approach, assumes that each element $a_{ij}, j \in J_i$, is modelled as a symmetric, independent and bounded random variable $\tilde{a}_{ij}, j \in J_i$. It is then assumed that this random variable, $\tilde{a}_{ij}$, takes values in the range $[a_{ij} - \hat{a}_{ij}, a_{ij} + \hat{a}_{ij}]$, where $\hat{a}_{ij}$ is a user-defined amount.

For every row i in $\mathbf{A}$, a parameter Γ_i is introduced and is used to adjust the robustness of the proposed model against the level of conservatism of the solution. Γ_i is known as the 'protection level', or alternatively as the 'budget of uncertainty', and its value can be chosen to make the solution more or less conservative by taking values in the range $\Gamma_i \in [0, |J_i|]$, however, is not necessarily integer. Essentially, Γ_i specifies how many of the uncertain coefficients in constraint i we would like to protect the solution against.

The values of the two parameters, Γ_i and $\hat{a}_{ij}$, that are used in this RO approach are independent of one another. That is to say, that a higher value of $\hat{a}_{ij}$, which implies there is more uncertainty associated with the value that $\tilde{a}_{ij}$ takes, does not imply that a larger value of Γ_i should be used to protect against more of the uncertainty. The value of $\hat{a}_{ij}$ should be informed by data analysis or knowledge of the application, however, the value of Γ_i is chosen by the decision maker and should reflect his views on how much to protect against the uncertainty. Although the parameters take independent values, there is a special case when $\hat{a}_{ij} = 0 \ \forall \ j$ for constraint i. If $\hat{a}_{ij} = 0 \ \forall \ j$ for constraint i, this is equivalent to saying that the random variable $\tilde{a}_{ij}$ takes its point estimate a_{ij} for constraint i, since the width of the interval is specified as zero. Hence we are certain of the value of $\tilde{a}_{ij}$ for constraint i, so the only sensible choice of Γ_i is also zero because there are no uncertain coefficients to protect against.

In most circumstances, it is unlikely that all of the $a_{ij}, j \in J_i$, will change, but through the use of Γ_i we have the ability to be protected against up to $\lfloor \Gamma_i \rfloor$ of the $a_{ij}, j \in J_i$, changing values. Only one other coefficient, say $a_{it}, t \in J_i$, is allowed to change at most by the amount $(\Gamma_i - \lfloor \Gamma_i \rfloor)\hat{a}_{it}$. Here, $\lfloor \cdot \rfloor$ denotes the floor function which returns the largest integer less than or equal to its argument.

Under these conditions, Bertsimas and Sim [31] have shown that this approach to RO has the properties that:

- the robust solution will be feasible deterministically;
- even if more than $\lfloor \Gamma_i \rfloor$ coefficients change, then the robust solution will be feasible with very high probability.

The robust counterpart of Model 7.1 involves the use of a protection function for each uncertain constraint as follows:

$$\begin{array}{lll}
\max & \mathbf{c}^{\mathbf{T}}\mathbf{x} & \\
s.t. & \sum_j a_{ij}x_j + \max\limits_{\{S_i \cup \{t_i\} | S_i \subseteq J_i, |S_i| = \lfloor \Gamma_i \rfloor, t_i \in J_i \setminus S_i\}} \left\{ \sum_{j \in S_i} \hat{a}_{ij} y_j + (\Gamma_i - \lfloor \Gamma_i \rfloor)\hat{a}_{it} y_t \right\} \leq b_i & \forall \ i \\
& -y_j \leq x_j \leq y_j & \forall \ j \\
& \mathbf{l} \leq \mathbf{x} \leq \mathbf{u} & \\
& \mathbf{y} \geq \mathbf{0} &
\end{array} \tag{7.2}$$

As can be seen in Model 7.2, the objective function remains the same as in the nominal problem (Model 7.1) because it is assumed there are no uncertain data that affects the objective function coefficients. If there exists some uncertainty in the objective function coefficients, then the objective function can be transformed into a constraint and included into $\mathbf{Ax} \leq \mathbf{b}$ in Model 7.1. A protection function, $\beta_i(\mathbf{x}, \Gamma_i)$, has been added to the left hand side of each constraint which is used to account for the desired level of robustness. For each constraint i, the protection function is the maximum amount the uncertain coefficients can change. Dummy decision variables, y, are included in the protection function to represent the uncertainty set.

An important point to note is that when $\Gamma_i = 0$, the protection function, $\beta_i(\mathbf{x}, \Gamma_i)$, also equals zero implying that the constraints in Model 7.2 are equivalent to those in the nominal problem (Model 7.1). In the other extreme, if $\Gamma_i = |J_i|$ to ensure full protection against uncertainty, the robust problem (Model 7.2) becomes that of Soyster's method [145].

The set of coefficients over which the protection function is maximised, is $S_i \cup \{t_i\}$, where

- S_i is a subset of J_i, whose number of elements equals $\lfloor \Gamma_i \rfloor$;
- t_i is an element of J_i that is not in S_i.

This is illustrated in Figure 7.1, where J_i is the set of of uncertain coefficients from $a_{i1}, ..., a_{in}$. In this example, $\Gamma_i \in [0, 3]$, and let $\Gamma_i = 1.5$. Hence $\lfloor \Gamma_i \rfloor = 1$. Note that it is not always the case that $S_i \cup t_i = J_i$. Therefore the protection function is maximised over all combinations of S_i and $\{t_i\}$ in J_i.

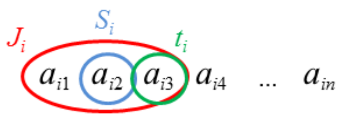


Figure 7.1: Illustrative example of the sets used in the protection function

By considering all combinations of S_i and $\{t_i\}$ in J_i, the maximum amount that the uncertain coefficients can vary by $(\sum_{j \in S_i} \hat{a}_{ij} y_j)$, and the maximum amount that the one other coefficient can vary by $((\Gamma_i - \lfloor \Gamma_i \rfloor)\hat{a}_{it} y_t)$ are found. By taking the maximum value, the constraint ensures that the 'worse-case' scenario is satisfied.

As a result of including a protection function in each constraint, Model 7.2 becomes a bi-level optimisation problem. In order to get a linear formulation of the robust counterpart, we need to use the fact that the protection function is equivalent to a linear optimisation problem. Specifically, given a solution vector $\mathbf{x}^*$, the protection function of the ith constraint,

$$\beta_i(\mathbf{x}^*, \Gamma_i) = \max_{\{S_i \cup \{t_i\} | S_i \subseteq J_i, |S_i| = \lfloor \Gamma_i \rfloor, t_i \in J_i \setminus S_i\}} \left\{ \sum_{j \in S_i} \hat{a}_{ij} |\mathbf{x}^*| + (\Gamma_i - \lfloor \Gamma_i \rfloor) \hat{a}_{it} |\mathbf{x}^*| \right\}$$

equals the objective function of the following linear optimisation problem:

$$\begin{aligned}\beta_i(\mathbf{x}^*, \Gamma_i) = \quad & \max \quad \sum_{j \in J_i} \hat{a}_{ij} |\mathbf{x}^*| z_{ij} \\ & s.t. \quad \sum_{j \in J_i} z_{ij} \leq \Gamma_i \\ & \qquad\;\; 0 \leq z_{ij} \leq 1 \qquad \forall \; j \in J_i\end{aligned} \tag{7.3}$$

This can be shown to be true by inspecting the optimal solution to Model 7.3. The optimal solution consists of $\lfloor \Gamma_i \rfloor$ of the z_{ij} decision variables equalling 1, and one z_{ij} decision variable equalling $(\Gamma_i - \lfloor \Gamma_i \rfloor)$; giving the sum of the decision variables as:

$$\lfloor \Gamma_i \rfloor \times 1 + 1 \times (\Gamma_i - \lfloor \Gamma_i \rfloor) = \Gamma_i = \sum_{j \in J_i} z_{ij}.$$

Using duality theory, the dual of sub-problem Model 7.3 is:

$$\begin{aligned} \min \quad & \Gamma_i q_i + \sum_{j \in J_i} p_{ij} \\ s.t. \quad & q_i + p_{ij} \geq \hat{a}_{it} |\mathbf{x}^*| \\ & p_{ij} \geq 0 \\ & q_i \geq 0 \end{aligned} \tag{7.4}$$

where p_{ij} and q_i are the dual decision variables.

By the strong duality theorem [55], since Model 7.3 is feasible and bounded for all $\Gamma_i \in [0, |J_i|]$, the dual of the problem is also feasible and bounded. The dual and primal will also have identical optimal values of the objective function. Using this fact and substituting the dual problem, Model 7.4, back into Model 7.2, a linear formulation of the robust counterpart of Model 7.1 is obtained:

$$\begin{aligned} \max \quad & \mathbf{c^T x} & \\ s.t. \quad & \sum_{j} a_{ij} x_j + \Gamma_i q_i + \sum_{j \in J_i} p_{ij} \leq b_i & \forall \; i \\ & q_i + p_{ij} \geq \hat{a}_{ij} y_i & \forall \; i, j \in J_i \\ & -y_j \leq x_j \leq y_j & \forall \; j \\ & l_j \leq x_j \leq u_j & \forall \; j \\ & p_{ij} \geq 0 & \forall \; i, j \in J_i \\ & q_i \geq 0 & \forall \; i \\ & x_j \geq 0 & \forall \; j \\ & y_j \geq 0 & \forall \; j \end{aligned} \tag{7.5}$$

We now have a linear optimisation problem that includes the original decision variables, x_j, to indicate which plans are chosen for each specialty, and new decision variables p_{ij} and q_i that are used to reflect the desired robustness of the final solution.

If the decision variables x_j are binary decision variables, then the linear robust formulation becomes:

$$\begin{array}{lll} \max & \mathbf{c^T x} & \\ s.t. & \sum_j a_{ij}x_j + \Gamma_i q_i + \sum_{j \in J_i} p_{ij} \leq b_i & \forall\ i \\ & q_i + p_{ij} \geq \hat{a}_{ij} y_i & \forall\ i, j \in J_i \\ & p_{ij} \geq 0 & \forall\ i, j \in J_i \\ & q_i \geq 0 & \forall\ i \\ & x_j \in \{0,1\} & \forall\ j \end{array} \tag{7.6}$$

Bertsimas and Thiele [32] show how, in some cases, the optimal solution to this binary robust problem can be found by solving n subproblems of the same size and structure as the original deterministic problem, and selecting the one with the highest objective value as the optimal solution. This approach exploits the nature of the binary variables, while preserving the computational tractability of this approach to RO. It is not deemed applicable in the case of the construction of the MSS optimisation model due to the restrictive nature of the GUB constraints.

7.2 Developing a Robust Optimisation Model for the MSS

Recall the nominal formulation for the construction of the MSS from Section 4.2.1 is as follows:

$$\begin{array}{lll} \min & \sum_{k=1}^{p}\sum_{l=1}^{q}(d_k^{(l)} - \sum_{j=1}^{n} b_{kj}^{(l)} x_j) & \\ \text{s.t.} & \sum_{j=1}^{n} a_{ij}x_j = 1 & \forall \quad i = 1,...,s \\ & \sum_{j=1}^{n} a_{ij}x_j \leq 1 & \forall \quad i = s+1,...,m \\ & \sum_{j=1}^{n} b_{kj}^{(l)} x_j \leq d_k^{(l)} & \forall \quad k = 1,...,p, \;\; l = 1,...,q \\ & x_j \in \{0,1\} & \forall \quad j = 1,...,n \end{array} \tag{7.7}$$

The above nominal problem differs from the nominal problem in Section 7.1 since it is a maximisation problem, and there are both equality and inequality constraints. In order to be confident that a robust counterpart can be found using the approach presented in Section 7.1, Model 7.7 is re-formulated to match the format of the nominal problem in Model 7.1.

The nominal problem for the construction of the MSS in the same format as Model 7.1 is as follows:

$$
\begin{aligned}
\max \quad & \sum_{k=1}^{p}\sum_{l=1}^{q}\Big(\sum_{j=1}^{n} b_{kj}^{(l)} x_j - d_k^{(l)}\Big) && && (7.8)\\
\text{s.t.} \quad & \sum_{j=1}^{n} a_{ij} x_j \le 1 && \forall \quad i = 1, ..., s && (7.9)\\
& -\sum_{j=1}^{n} a_{ij} x_j \le -1 && \forall \quad i = 1, ..., s && (7.10)\\
& \sum_{j=1}^{n} a_{ij} x_j \le 1 && \forall \quad i = s+1, ..., m && (7.11)\\
& \sum_{j=1}^{n} b_{kj}^{(l)} x_j \le d_k^{(l)} && \forall \quad k = 1, ..., p, \;\; l = 1, ..., q && (7.12)\\
& x_j \in \{0, 1\} && \forall \quad j = 1, ..., n && (7.13)
\end{aligned}
$$

The objective function in (7.8) is now being maximised. Initially we were minimising the difference between the beds available and the beds required, i.e. the number of empty beds, however, now we are maximising the difference between the number of beds required and the beds available, i.e. the number of used beds. Both objectives aim to increase the throughput of patients through the wards. The GUB constraints in Model 7.7 are equality constraints, so are transformed into two equivalent sets of constraints; one less than (constraint 7.9) and one greater than (constraint 7.10) which was multiplied by -1 in order to be in the form of a less than or equal constraint. The nominal problem for the construction of the MSS is now in the same form as Model 7.1, allowing us to continue to formulate a robust counterpart.

The nominal problem for the construction of the MSS can be summarised in the following matrix form:

$$
\begin{aligned}
\max \quad & \mathbf{Bx} - \mathbf{d} \\
s.t. \quad & \mathbf{Mx} \leq \mathbf{e} \\
& \mathbf{x} \in \{0,1\}^n
\end{aligned}
\tag{7.14}
$$

where M is the combined matrix of the A and B matrices from Model 7.7, and $\mathbf{e}$ is the combined vector of the right-hand side values of the GUB constraints, operating theatre constraints and bed constraints. M and $\mathbf{e}$ are summarised in Figure 7.2.

$$
\mathbf{M} = \begin{pmatrix} \text{GUB constraints} \\ \hdashline \text{OT constraints} \\ \hdashline \text{Bed constraints} \end{pmatrix} \qquad \mathbf{e} = \begin{pmatrix} 1 \\ -1 \\ \hdashline 1 \\ \hdashline d_k^{(l)} \end{pmatrix}
$$

Figure 7.2: Combined matrix and vector for constraints in the nominal problem

In Model 7.14, data uncertainty only affects elements of matrix M. Indeed, the only uncertain data in the model are the $b_{kj}^{(l)}$ coefficients of the B matrix, i.e. we are uncertain about the number of beds required in ward k on day l for plan j.

Let

J_i = set of coefficients in constraint i in matrix M that are subject to uncertainty.

We assume that each uncertain coefficient in M, $m_{ij}, j \in J_i$, is modelled as a symmetric, independent and bounded random variable $\tilde{m}_{ij}, j \in J_i$, taking values in $[m_{ij} - \hat{m}_{ij}, m_{ij} + \hat{m}_{ij}]$.

For each constraint i, we introduce a parameter, Γ_i, that represents the budget of uncertainty and can be used to control the conservatism of the robust solution. Let

$$\Gamma_i \in [0, |J_i|] \quad \forall \quad i.$$

The robust counterpart of Model 7.14, which includes a protection function for each constraint i, is therefore:

$$\begin{aligned} \max \quad & \mathbf{Bx} - \mathbf{d} \\ s.t. \quad & \sum_{j}^{n} m_{ij}x_j + \max_{\{S_i \cup \{t_i\} | S_i \subseteq J_i, |S_i| = \lfloor \Gamma_i \rfloor, t_i \in J_i \setminus S_i\}} \left\{ \sum_{j \in S_i} \hat{m}_{ij}x_j + (\Gamma_i - \lfloor \Gamma_i \rfloor)\hat{m}_{it}x_t \right\} \leq e_i \forall i \\ & \mathbf{x} \in \{0,1\}^n \end{aligned} \tag{7.15}$$

Model 7.15 is a bi-level optimisation problem, so following the linearisation approach as in Section 7.1, the linear formulation of the robust counterpart of Model 7.14 may be written:

$$\begin{aligned} \max \quad & \mathbf{Bx} - \mathbf{d} & \\ s.t. \quad & \sum_{j}^{n} m_{ij}x_j + \Gamma_i q_i + \sum_{j \in J_i} p_{ij} \leq e_i & \forall \ i \\ & q_i + p_{ij} \geq \hat{m}_{ij}x_j & \forall \ i, j \in J_i \\ & p_{ij}, q_i \geq 0 & i, j \in J_i \\ & x_j \in \{0,1\} & \forall \ j = 1, ..., n \end{aligned} \tag{7.16}$$

where p_{ij} and q_i are dual decision variables from the linearisation process.

Given that we know that the only uncertain data in the model are the $b_{kj}^{(l)}$ coefficients of the B matrix, i.e. we are uncertain about the number of beds required in ward k on day l for plan j, we can specify some values of the parameters $\hat{m}_{ij}$ and Γ_i.

If we are certain about the value of a coefficent in a constraint, the size of the interval $[m_{ij} - \hat{m}_{ij}, m_{ij} + \hat{m}_{ij}]$ is zero, and the random variable $\tilde{m}_{ij}$ takes the value of its point estimate m_{ij}. Therefore, we can set $\hat{m}_{ij} = 0$ for all certain coefficients. If a constraint does not contain any uncertain data, then $|J_i| = 0$ and there are no uncertain coefficients to protect against in the robust solution. Therefore, we can also set $\Gamma_i = 0$ for all constraints that do not contain any uncertain data.

In the problem of the construction of the MSS, we are certain about the coefficient values in the GUB constraints and the operating theatre constraints. Therefore $\hat{m}_{ij} = 0$ and $\Gamma_i = 0$ for these constraints.

For the bed constraints that contain uncertain data, let

J_k = set of bed requirement coefficients for ward k that are subject to uncertainty.

It is assumed that the set of uncertain coefficients, J_k, is constant for all days l on each ward k. This is because there are the same number of plans that result in uncertain bed requirements for all days on each ward in the B matrix.

We assume that each $b_{kj}^{(l)}, j \in J_k$ is modelled as a symmetric, independent and bounded random variable $\tilde{b}_{kj}^{(l)}, j \in J_k$, taking values in the interval $[b_{kj}^{(l)} - \hat{b}_{kj}^{(l)}, b_{kj}^{(l)} + \hat{b}_{kj}^{(l)}]$. Due to the lack of bed count data available from UHW, we cannot infer any information on the shape of the uncertainty set that the uncertain $b_{kj}^{(l)}$ coefficients belong to. Therefore, it is considered reasonable to assume that the uncertain bed count coefficients belong to a symmetric interval around a point estimate. We must ensure that the $\tilde{b}_{kj}^{(l)}$ take integer values so that they correspond to a whole number of beds required. A discussion of how the $b_{kj}^{(l)}$ and $\hat{b}_{kj}^{(l)}$ values are chosen is given in Section 7.3.1 and 7.3.2.

For every ward k, we introduce a parameter $\Gamma_k \in [0, |J_k|]$ that is not necessarily integer and is used to adjust the robustness of the proposed model with respect to the level of conservatism of the solution. The robust solution will be protected against up to $\lfloor \Gamma_k \rfloor$ of the $b_{kj}^{(l)}, j \in J_k$ changing values. Only one other coefficient, $b_{kt}^{(l)}$, is allowed to change at most by the amount $(\Gamma_k - \lfloor \Gamma_k \rfloor)\hat{b}_{kt}^{(l)}$. The higher the value of Γ_k, the more protection there is against the uncertain bed requirement for ward k. Γ_k is assumed constant for all days l on each ward k; a different $\Gamma_k^{(l)}$ could be specified for each day l on each ward k, however, this is deemed too detailed for the data available from UHW.

Having defined the parameters for the uncertain bed constraints, Model 7.16 can now be separated into the GUB, operating theatre and bed constraints, and values can be set for Γ_k and $\hat{m}_{ij}$. We can set $\Gamma_i = \Gamma_k$ and $\hat{m}_{ij} = \hat{b}_{kj}^{(l)}$ which results in the following forms of $\hat{\mathbf{m}}$ and $\mathbf{\Gamma}$ in Figure 7.3.

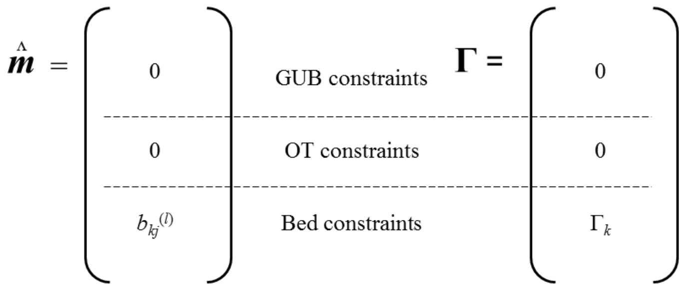


Figure 7.3: Specific values of Γ_k and $\hat{b}_{kj}^{(l)}$

Separating the constraints of Model 7.16 into their constituent parts of the A and

B matrices, we get the following formulation:

$$\begin{aligned}
\max \quad & \mathbf{Bx} - \mathbf{d} \\
s.t. \quad & \sum_{j=1}^{n} a_{ij}x_j + \Gamma_i q_i + \sum_{j \in J_i} p_{ij} \leq 1 \ \forall \quad i = 1, ..., s \\
& -\sum_{j=1}^{n} a_{ij}x_j + \Gamma_i q_i + \sum_{j \in J_i} p_{ij} \leq -1 \ \forall \quad i = 1, ..., s \\
& \sum_{j=1}^{n} a_{ij}x_j + \Gamma_i q_i + \sum_{j \in J_i} p_{ij} \leq 1 \quad \forall \quad i = s+1, ..., m \\
& \sum_{j=1}^{n} b_{kj}^{(l)}x_j + \Gamma_k q_k^{(l)} + \sum_{j \in J_k} p_{kj}^{(l)} \leq d_k^{(l)} \quad \forall \quad k = 1, ..., p, \ \ l = 1, ..., q \\
& q_i + p_{ij} \geq \hat{m}_{ij}x_j \quad \forall \quad i = 1, ..., m, j \in J_i \\
& q_k^{(l)} + p_{kj}^{(l)} \geq \hat{b}_{kj}^{(l)}x_j \quad \forall \quad k, l, j \in J_k \\
& p_{ij}, q_i \geq 0 \quad \forall \quad i, j \in J_i \\
& p_{kj}^{(l)}, q_k \geq 0 \quad \forall \quad k, l, j \in J_k \\
& x_j \in \{0, 1\} \quad \forall \quad j = 1, ..., n
\end{aligned} \tag{7.17}$$

We can now substitute the above chosen values for $\hat{m}_{ij}$ and Γ_i into Model 7.17 to simplify the formulation:

$$\max \quad \mathbf{Bx} - \mathbf{d}$$

$$s.t. \quad \sum_{j=1}^{n} a_{ij}x_j + \sum_{j \in J_i} p_{ij} \leq 1 \ \forall \quad i = 1, ..., s \tag{7.18}$$

$$-\sum_{j=1}^{n} a_{ij}x_j + \sum_{j \in J_i} p_{ij} \leq -1 \ \forall \quad i = 1, ..., s \tag{7.19}$$

$$\sum_{j=1}^{n} a_{ij}x_j + \sum_{j \in J_i} p_{ij} \leq 1 \quad \forall \quad i = s+1, ..., m \tag{7.20}$$

$$\begin{aligned}
\sum_{j=1}^{n} b_{kj}^{(l)}x_j + \Gamma_k q_k^{(l)} + \sum_{j \in J_k} p_{kj}^{(l)} &\leq d_k^{(l)} \quad \forall \quad k = 1, ..., p, \ \ l = 1, ..., q \\
q_i + p_{ij} &\geq \hat{m}_{ij}x_j \quad \forall \quad i = 1, ..., m, j \in J_i \\
q_k^{(l)} + p_{kj}^{(l)} &\geq \hat{b}_{kj}^{(l)}x_j \quad \forall \quad k, l, j \in J_k \\
p_{ij}, q_i &\geq 0 \quad \forall \quad i, j \in J_i \\
p_{kj}^{(l)}, q_k &\geq 0 \quad \forall \quad k, l, j \in J_k \\
x_j &\in \{0, 1\} \quad \forall \quad j = 1, ..., n
\end{aligned} \tag{7.21}$$

If a constraint does not contain any uncertain data, then $|J_i| = 0$ and the sum $\sum_{j \in J_i} p_{ij} = 0$, since J_i is an empty set. This affects constraints 7.18, 7.19 and 7.20

that relate to the GUB and operating theatre constraints, so this summation term can be removed from these constraints, and constraint 7.21 becomes redundant. The linear robust counterpart of the construction of the MSS problem can therefore be simplified to:

$$\max \quad \sum_{k=1}^{p}\sum_{l=1}^{q}(\sum_{j=1}^{n} b_{kj}^{(l)}x_j - d_k^{(l)}) \tag{7.22a}$$

$$s.t. \quad \sum_{j=1}^{n} a_{ij}x_j \leq 1 \quad \forall \quad i=1,...,s \tag{7.22b}$$

$$-\sum_{j=1}^{n} a_{ij}x_j \leq -1 \quad \forall \quad i=1,...,s \tag{7.22c}$$

$$\sum_{j=1}^{n} a_{ij}x_j \leq 1 \quad \forall \quad i=s+1,...,m \tag{7.22d}$$

$$\sum_{j=1}^{n} b_{kj}^{(l)}x_j + \Gamma_k q_k^{(l)} + \sum_{j\in J_k} p_{kj}^{(l)} \leq d_k^{(l)} \quad \forall \quad k=1,...,p, \;\; l=1,...,q \tag{7.22e}$$

$$q_k^{(l)} + p_{kj}^{(l)} \geq \hat{b}_{kj}^{(l)}x_j \quad \forall \quad k,l,j \in J_k^{(l)} \tag{7.22f}$$

$$p_{kj}^{(l)}, q_k \geq 0 \quad \forall \quad k,l,j \in J_k \tag{7.22g}$$

$$x_j \in \{0,1\} \quad \forall \quad j=1,...,n \tag{7.22h}$$

The combined terms in the uncertain bed constraint (Constraint 7.22e) that include the dual decision variables can be interpreted as a safety buffer of beds reserved on each ward on each day. As Γ_k increases, i.e. the decision maker becomes more conservative against the uncertainty associated with the number of required beds, the size of the safety buffer increases. This intuitively matches the interpretation as the decision maker becomes more conservative against the uncertainty associated with the number of required beds.

Through the development of Model 7.22, we have shown that the RO approach developed by Bertsimas and Sim [31] can successfully be applied to construct a robust counterpart of Model 7.14. It has also been shown that if there is a mixture of certain and uncertain constraints in the nominal problem, then it is possible to apply the protection function to the uncertain constraints only, thereby extending the original model formulation of Bertsimas and Sim [31].

7.3 Applying the Robust Optimisation Model to the Case Study

The selection of values for the two parameters in the robust counterpart for the construction of the MSS (Model 7.22) is discussed in the following sections.

7.3.1 Parameter Values: Point Estimate of Bed Requirements, $b_{kj}^{(l)}$

We have assumed that the uncertain coefficients, $b_{kj}^{(l)}$, representing the bed requirements in the bed constraints for each ward on each day are modelled as symmetric, independent and bounded random variables $\tilde{b}_{kj}^{(l)} \in [b_{kj}^{(l)} - \hat{b}_{kj}^{(l)}, b_{kj}^{(l)} + \hat{b}_{kj}^{(l)}] \ \forall \ j \in J_k$.

The interval is centred on a point estimate of the bed requirement, which is found using the approach described in Section 4.2.4. The values must be a whole number of beds that are required for each ward on each day, given the conditional probability of leaving the hospital on each consecutive day after surgery. One instance of the B matrix generated in this way can be thought of as one instance of the estimate of the bed requirement on each ward on each day for each plan. Of course, different instances of the B matrix will result in different values of $b_{kj}^{(l)}$. However, this uncertainty is inherently taken care of through the use of the protection level Γ_k and $\hat{b}_{kj}^{(l)}$ in the robust counterpart of the problem.

7.3.2 Parameter Values: Width of Uncertainty Set, $\hat{b}_{kj}^{(l)}$

The $\hat{b}_{kj}^{(l)}$ values represent the amount by which the point estimates of bed requirements can vary, i.e. creating the interval $[b_{kj}^{(l)} - \hat{b}_{kj}^{(l)}, b_{kj}^{(l)} + \hat{b}_{kj}^{(l)}]$. In order to ensure that the random variable $\tilde{b}_{kj}^{(l)}$ has an integer value, the $\hat{b}_{kj}^{(l)}$ values need to be chosen to also take integer values.

The choice of $\hat{b}_{kj}^{(l)}$ is left to the decision maker to choose appropriate values. In order to make an informed choice of $\hat{b}_{kj}^{(l)}$, variation in the observed bed count data on each ward in UHW is used to obtain information on what values $\hat{b}_{kj}^{(l)}$ should take.

Given that the conditional probability of leaving hospital on each consecutive day after surgery is fixed for each ward (as found from the LoS data), the only way that the bed requirement can vary is due to the number of patients that have surgery per operating theatre session. The distribution of the number of operations per session that took place in 2012/13 for each specialty is shown in Figure 7.4.

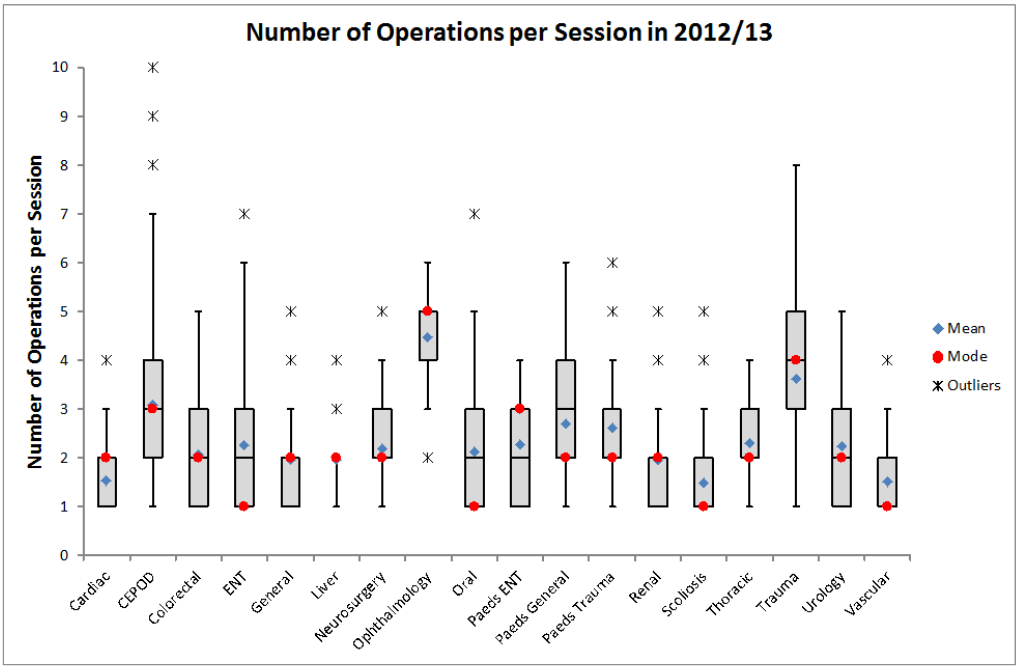


Figure 7.4: Number of operations per session in 2012/13

A discussion of the distribution of the number of operations per session observed in 2012/13 was given in Section 3.3.3. In order to use this information for the number of operations that took place per session to inform the values of $\hat{b}_{kj}^{(l)}$, we will let the interval $[b_{kj}^{(l)} \pm \hat{b}_{kj}^{(l)}]$ be represented by the interquartile range (IQR). The data on the number of operations per session relates to each surgical specialty and is shown in Table 7.1. To find values for $\hat{b}_{kj}^{(l)}$, the IQR is rounded down after being divided by two, i.e. $\left\lfloor \frac{\text{IQR}}{2} \right\rfloor$, since a whole number of operations per session is required. If half the IQR is a fraction of an operation, then for surety the largest integer number of operations that is not larger than half the IQR can be performed in a session.

Specialty	**Operations per session IQR**	$\left\lfloor \frac{\text{IQR}}{2} \right\rfloor$
Cardiac	1	0
CEPOD	2	1
Colorectal	1	0
ENT	2	1
General	1	0
Liver	1	0
Neurosurgery	1	0
Ophthalmology	1	0
Oral	3	1
Paeds ENT	2	1
Paeds General	2	1
Paeds Trauma	3	1
Renal	1.5	0
Scoliosis	1	0
Thoracic	1	0
Trauma	3	1
Urology	2	1
Vascular	1	0

Table 7.1: Value of $\hat{b}_{kj}^{(l)}$ for each specialty

These values of $\hat{b}_{kj}^{(l)}$ for each specialty need to be translated into $\hat{b}_{kj}^{(l)}$ values for each ward, since $\hat{b}_{kj}^{(l)}$ relates to each ward k. This is calculated by finding the average half IQR of all specialties that use ward k, and the $\hat{b}_{kj}^{(l)}$ values for each ward are given in Table 7.2.

Ward	**Average operations per session IQR**	$\hat{b}_{kj}^{(l)}$
Paediatric	0.8	1
ENT/Oral	0.75	1
Vascular	0.5	1
Trauma	1	1
Renal	0	0
General/Liver	0.33	0
Urology	1	1
Colorectal	0	0
Cardiothoracic	0	0
Neurosurgery	0	0
Critical Care	0.4	0

Table 7.2: Values of $\hat{b}_{kj}^{(l)}$ for each ward

Note that all values of $\hat{b}_{kj}^{(l)}$ are zero, except for the Paediatric, ENT/Oral, Vascular, Trauma and Urology wards for which $\hat{b}_{kj}^{(l)} = 1$. If $\hat{b}_{kj}^{(l)} = 0$ for ward k, the random variable $\tilde{b}_{kj}^{(l)}$ will take the value of the point estimate $b_{kj}^{(l)}$ because the symmetric interval uncertainty set has zero width.

However, from the analysis of the bed count in UHW, it is clear that there is variation and hence uncertainty associated with the bed requirement on all wards under consideration. Therefore, $\hat{b}_{kj}^{(l)} \geq 1$ will be assumed for all wards in order to be able to use a protection function and to investigate varying levels of uncertainty associated with the bed requirements. Unless stated otherwise, $\hat{b}_{kj}^{(l)} = 1$ will be used for all wards k since this is the smallest integer value that $\tilde{b}_{kj}^{(l)}$ can take, and is deemed reasonable from the results of the above data analysis of the number of operations per session.

7.3.3 Parameter Values: Protection Level, Γ_k

As discussed in Section 7.1, the parameter Γ is used to adjust the robustness of the proposed model against the level of conservatism of the solution. The protection level, Γ, controls the *price of robustness* which is defined as 'the trade-off between the probability of constraint violation and the effect to the objective function of the nominal problem' [31]. The optimal value of the objective function typically worsens in order to have a more robust model that attempts to reduce the probability of constraint violation. Hence the choice of Γ is important in this compromise.

Bertsimas and Thiele [32] call the parameter Γ the 'budget of uncertainty' since it relates to the number of uncertain coefficients in each constraint that are protected against the uncertainty. Theoretically, the budget of uncertainty can take values in the range $\Gamma_i \in [0, |J_i|]$, i.e. you can protect against all of the uncertain coefficients in constraint i, none of them, or a subset of the uncertain coefficients. The values of Γ_i chosen for model implementation are chosen by the decision maker. They can be chosen to reflect the decision maker's attitude to uncertainty, or based on their knowledge of the application.

In the robust counterpart model for the generation of the MSS (Model 7.22), a value of Γ_k needs to be chosen for each ward k. Γ_k will reflect the amount of uncertain coefficients in the bed constraints, i.e. the bed requirement on ward k on day l for plan j. It is assumed that Γ_k is constant for all days for each ward k considered in the bed constraints.

Due to the enumeration of all possible plans for each specialty, given a variety of scheduling rules, there are a different number of possible plans for each specialty in the model. This results in a different number of uncertain coefficients for each ward in the bed constraints, i.e. $|J_k|$ is different for each ward k. Hence, the values of Γ_k are chosen in relation to $|J_k|$ for each ward k.

For all wards k, Γ_k will be assigned a value that is a certain proportion of $|J_k|$ so that all wards are protected against the same proportion of uncertainty, regardless of how many actual uncertain coefficients there are for each ward. The same proportion of uncertainty will be protected for all wards, i.e. $\Gamma_k = x\%$ of $|J_k|$ where $x \in [0, 100]$ and is constant for all wards k. This ensures that Γ_k will take proportionate values in the range $[0, |J_k|]$ for each ward k. Illustrative examples of the values of Γ_k for different proportions of $|J_k|$ are given in Table 7.3. If the decision maker in the hospital was able to quantify whether one ward is more uncertain than another ward, x could be varied across wards. Since this information is not available, the level of Γ_k will remain constant for all wards k in this model.

	$\Gamma_k = x\%$ of $\lvert J_k \rvert$					
Ward	5%	10%	25%	50%	80%	100%
Paediatric	1.6	3.1	7.8	15.5	24.8	31.0
ENT/Oral	26.1	52.1	130.3	260.5	416.8	521.0
Vascular	10.6	21.1	52.8	105.5	168.8	211.0
Trauma	0.3	0.6	1.5	3.0	4.8	6.0
Renal	22.8	45.5	113.8	227.5	364.0	455.0
General/Liver	10.8	21.6	54.0	108.0	172.8	216.0
Urology	0.1	0.1	0.3	0.5	0.8	1.0
Colorectal	0.3	0.5	1.3	2.5	4.0	5.0
Cardiothoracic	0.3	0.6	1.5	3.0	4.8	6.0
Neurosurgery	0.1	0.2	0.5	1.0	1.6	2.0
Critical Care	21.4	42.7	106.8	213.5	341.6	427.0

Table 7.3: Illustrative values of Γ_k for different proportions of $|J_k|$

An alternative approach to choosing values of Γ_k for all wards k that ensures that the probability of constraint violation is bounded by a specified amount is discussed in Section 7.6.

7.4 Results: Investigating Different Values of $\hat{b}_{kj}^{(l)}$

It is up to the decision maker to decide on the value of $\hat{b}_{kj}^{(l)}$ to use in the robust counterpart. Therefore, it is of interest to see the affect of different values of $\hat{b}_{kj}^{(l)}$ on the optimal schedules and their performance measures.

From further inspection of the distributions of the number of operations per session for each specialty in Figure 7.4 and Table 7.1, it can be seen that half the interquartile range of the number of operations per session does not exceed 1.5 operations for all specialties. Therefore, since $\hat{b}_{kj}^{(l)}$ must be integer, experiments will be carried out in order to investigate what happens when $\hat{b}_{kj}^{(l)} = 2$ for the uncertain coefficients in the bed constraints for all wards k. An extreme case of demand for beds, $\hat{b}_{kj}^{(l)} = 3$, will also be considered. These values of $\hat{b}_{kj}^{(l)}$ correspond to an additional two or three patients requiring a bed on each day on each ward than the point estimate $b_{kj}^{(l)}$. This is deemed a sensible and arguably realistic additional demand for beds; any more than three additionally required beds is considered unlikely to occur in reality.

In order to see the effect of changing the values of $\hat{b}_{kj}^{(l)}$, the other parameter in the robust model, Γ_k, will be kept constant for all experiments. In the robust counterpart model of the MSS problem (Model 7.22), $\hat{b}_{kj}^{(l)}$ affects the values of the dual decision variables. These dual decision variables are also present in the bed constraints, and considering the fact that the bed constraints are already quite 'tight' as found from the deterministic model in Section 5.1.2, it is considered the protection level should be kept quite low to ensure that the bed constraints are not violated as $\hat{b}_{kj}^{(l)}$ is increased. Hence, Γ_k will take values of 10% of $|J_k|$ for all wards k, i.e. 10% of the uncertain coefficients will be protected in the model and will be kept at this level for all experiments.

An initial analysis found that the current MSS used in UHW is not a feasible solution to the robust counterpart (Model 7.22) when Γ_k is 10% of $|J_k|$ for all wards k, and $\hat{b}_{kj}^{(l)} = 1$, 2 or 3. This supports the conclusion from Section 5.1.1, that the current MSS is not a feasible solution to the deterministic optimisation problem when the bed constraints are also included in the model. It also implies that the current MSS used in UHW is not robust against likely variations of bed requirements as described.

7.4.1 The Effect of $\hat{b}_{kj}^{(l)}$ on the Optimal Value

The first thing to note is the reduction in the number of feasible instances of the robust counterpart as the value of $\hat{b}_{kj}^{(l)}$ increases. Out of 1000 instances of the problem, 38.7% resulted in feasible solutions when $\hat{b}_{kj}^{(l)} = 1$, whereas only 2% resulted in feasible solutions when $\hat{b}_{kj}^{(l)} = 2$, and no feasible solutions were found when $\hat{b}_{kj}^{(l)} = 3$. This supports the theory that as $\hat{b}_{kj}^{(l)}$ increases, the bed constraints become 'tighter' due to the fact that the dual decision variables need to increase in order to satisfy constraint 7.22f. Hence, as $\hat{b}_{kj}^{(l)}$ increases, the uncertainty about the values of the bed requirements increases and it becomes harder to find feasible solutions and schedules for the problem. Subsequent experiments will therefore be performed for $\hat{b}_{kj}^{(l)} = 1$ and $\hat{b}_{kj}^{(l)} = 2$.

The results for twenty instances of the robust counterpart with $\hat{b}_{kj}^{(l)}$ taking different values are given in Table 7.4. Recall that the objective function is now maximising the difference between the number of beds required and the number of beds available, so a negative objective function value implies that there are unused beds in the system.

	Optimal value	
$\hat{b}_{kj}^{(l)}$	**Mean**	**Std Dev**
1	-903.8	36.0
2	-1003.2	43.2

Table 7.4: Results of twenty instances for different values of $\hat{b}_{kj}^{(l)}$

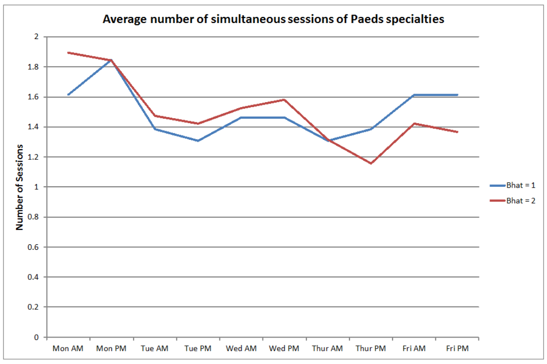
As can be seen in Table 7.4, when $\hat{b}_{kj}^{(l)}$ increases from 1 to 2, the average optimal value of solutions is reduced. A matched pairs t-test was performed at the 5% significance level on these results, which found that the average optimal value when $\hat{b}_{kj}^{(l)} = 2$ is statistically significantly lower than when $\hat{b}_{kj}^{(l)} = 1$ ($t = 9.895, p < 0.0005$).

A higher value of $\hat{b}_{kj}^{(l)}$ implies that there is more uncertainty associated with the bed requirement coefficients since the random variable will take values from a wider interval. Hence, if there is more uncertainty, then we can expect to see worse (lower) optimal solutions because different plans will have to be chosen to enable the tighter bed constraints to be satisfied.

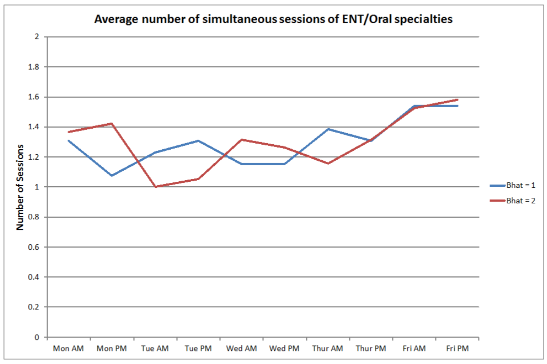
7.4.2 The Effect of $\hat{b}_{kj}^{(l)}$ on the Optimal Schedule

The optimal schedules for the same twenty instances were investigated to see if they are affected by different values of $\hat{b}_{kj}^{(l)}$. Specifically, the spread of simultaneous sessions of specialties that send their patients to one of the shared wards will be analysed. As in Sections 5.1.1 and 5.1.2, the five shared wards that have been identified as pinch-points in the system are investigated. Only specailties that are not assigned to fixed sessions are considered.

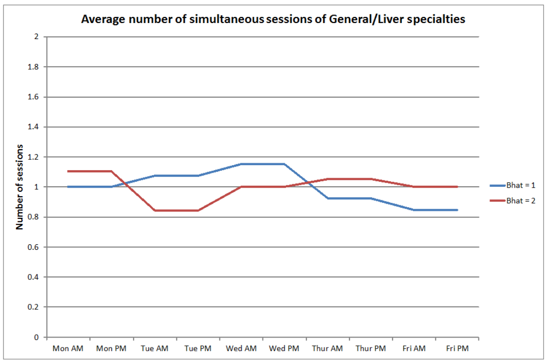
Figure 7.5 shows the average number of sessions that are scheduled simultaneously throughout the week for the specialties that send their patients to the shared wards, for when $\hat{b}_{kj}^{(l)} = 1$ and $\hat{b}_{kj}^{(l)} = 2$. These results represent the average across twenty instances of the robust counterpart.



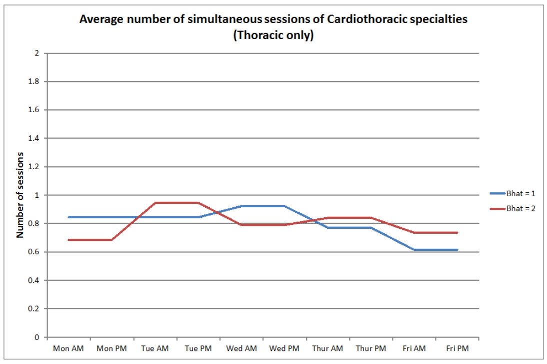
(a) Paediatric ward



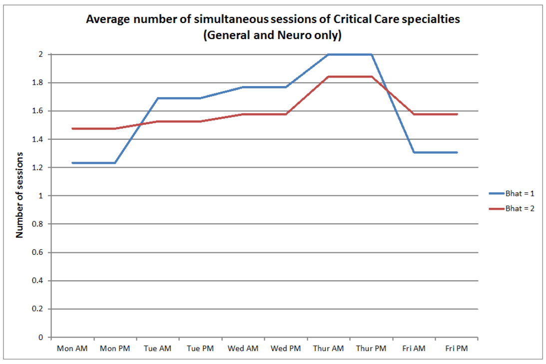
(b) ENT/Oral ward



(c) General/Liver ward



(d) Cardiothoracic ward



(e) CCU

Figure 7.5: Number of specialties that are scheduled simultaneously

As seen in Figures 7.5(a), 7.5(b), 7.5(c) and 7.5(e), more simultaneous sessions are scheduled at the start of the week (in particular on Monday) when $\hat{b}_{kj}^{(l)} = 2$ than when $\hat{b}_{kj}^{(l)} = 1$. For the Cardiothoracic ward, the majority of sessions are shifted to be earlier in the week when $\hat{b}_{kj}^{(l)} = 2$, as shown in Figure 7.5(d).

The higher the value of $\hat{b}_{kj}^{(l)}$, the wider the interval of possible bed requirements, which could imply that there are more bed days on the ward. It would appear that on average, the optimal schedules for a higher value of $\hat{b}_{kj}^{(l)}$ schedule more simultaneous sessions at the beginning of the week. This allows for the greater volume of bed days to be serviced throughout the week.

As well as a general trend in scheduling more sessions at the beginning of the week for a higher value of $\hat{b}_{kj}^{(l)}$, there also appears to be cyclic patterns in all graphs in Figure 7.5. For both values of $\hat{b}_{kj}^{(l)}$, the peaks in Figures 7.5(a) and 7.5(b) occur roughly every 2 days. This ties-in with the average LoS for the Paediatric ward and ENT/Oral wards which is 1.95 days and 2.4 days respectively. For the other wards, cyclic patterns are evident for both values of $\hat{b}_{kj}^{(l)}$, however, there is no apparent trend in the change in the cyclic pattern in relation to the different values of $\hat{b}_{kj}^{(l)}$. It is also not evident that the cycle length is related to the average LoS on each of these wards as shown in Table 7.5.

Ward	**Cycle length**		**Ward average LoS**
	$\hat{b}_{kj}^{(l)} = 1$	$\hat{b}_{kj}^{(l)} = 2$	**(days)**
Paediatric	2	2	2.0
ENT/Oral	1	2	2.4
General/Liver	5	2 – 5	5.4
Cardiothoracic	5	2 – 3	5.7
Critical Care	5	5	8.8

Table 7.5: Comparison of cycle length and average length of stay for different values of $\hat{b}_{kj}^{(l)}$

Overall findings from the graphs in Figure 7.5, indicate that for larger values of $\hat{b}_{kj}^{(l)}$, when there is increased uncertainty in the values of the bed requirement coefficients, more sessions are scheduled simultaneously at the start of the week. This could be in order to allow enough time for the patients to leave hospital, given the LoS distributions for each ward, ready for the next week to start and the cyclic MSS to repeat. The graphs in Figure 7.5 also exhibit a cyclic nature of the number of simultaneous sessions, however, this appears to be independent of $\hat{b}_{kj}^{(l)}$ and is related to the average LoS on each ward.

7.4.3 The Effect of $\hat{b}_{kj}^{(l)}$ on the Expected Bed Shortage

Table 7.6 contains the average expected bed shortage, obtained from the simulation of 100 instances for when $\hat{b}_{kj}^{(l)} = 1$ and 2. The expected bed shortage is, on average, slightly higher when $\hat{b}_{kj}^{(l)} = 2$ than when $\hat{b}_{kj}^{(l)} = 1$, and the standard deviation remains quite similar for the different levels of $\hat{b}_{kj}^{(l)}$. This can be interpreted that as $\hat{b}_{kj}^{(l)}$ increases, more cancellations are expected if the optimal schedules were to be implemented. A higher value of $\hat{b}_{kj}^{(l)}$ effectively means there is more uncertainty associated with the values of the bed requirement coefficients, so it is not surprising that an MSS that includes this additional uncertainty does not result in as few expected cancellations as for a lower value of $\hat{b}_{kj}^{(l)}$.

$\hat{b}_{kj}^{(l)}$	Expected bed shortage	
	Mean	**Std Dev**
1	7.8	0.9
2	8.2	0.9

Table 7.6: Expected bed count for different values of $\hat{b}_{kj}^{(l)}$

A paired samples t-test was performed at the 5% significance level on the data on expected bed shortage from the 100 instances. The difference in expected bed shortage for when $\hat{b}_{kj}^{(l)} = 1$ and 2 was not found to be significantly different ($t = -1.475$, $p - value = 0.158$). Hence, although the expected bed shortage is slightly higher for a higher value of $\hat{b}_{kj}^{(l)}$, it is not statistically significantly higher.

7.5 Results: Using $\Gamma_k = x\%$ of $|J_k|$

As discussed in Section 7.3.3, values for Γ_k will be chosen to reflect a certain percentage, $x\%$, of $|J_k|$ where x is a constant percentage for all wards k.

For all experiments that will investigate the parameter Γ_k, the value of $\hat{b}_{kj}^{(l)}$ is assumed to be one for all wards k. This reasonable assumption was based on a mixture of data analysis and the inherent meaning of having uncertain coefficients in the constraints as discussed in Section 7.3.2.

Similar to the $\hat{b}_{kj}^{(l)}$ experiments, an initial check for whether the current MSS was a feasible solution to the RO problem found that it was not a feasible solution for varying values of $\Gamma_k \in [0, |J_k|]$. This is consistent with results from the deterministic model in Section 5.1.1.

7.5.1 The Effect of the Protection Level on the Optimal Value

Figure 7.6 illustrates the effect of the protection level on the optimal value of the objective function for one instance of the MSS problem. Γ_k is increased for each ward in 1% increments from 0 to $|J_k|$. It can be seen that as the protection level increases, the optimal value decreases. This illustrates the pay-off, or price of robustness, that as the model becomes more robust, the optimal value worsens (until it reaches a plateau).

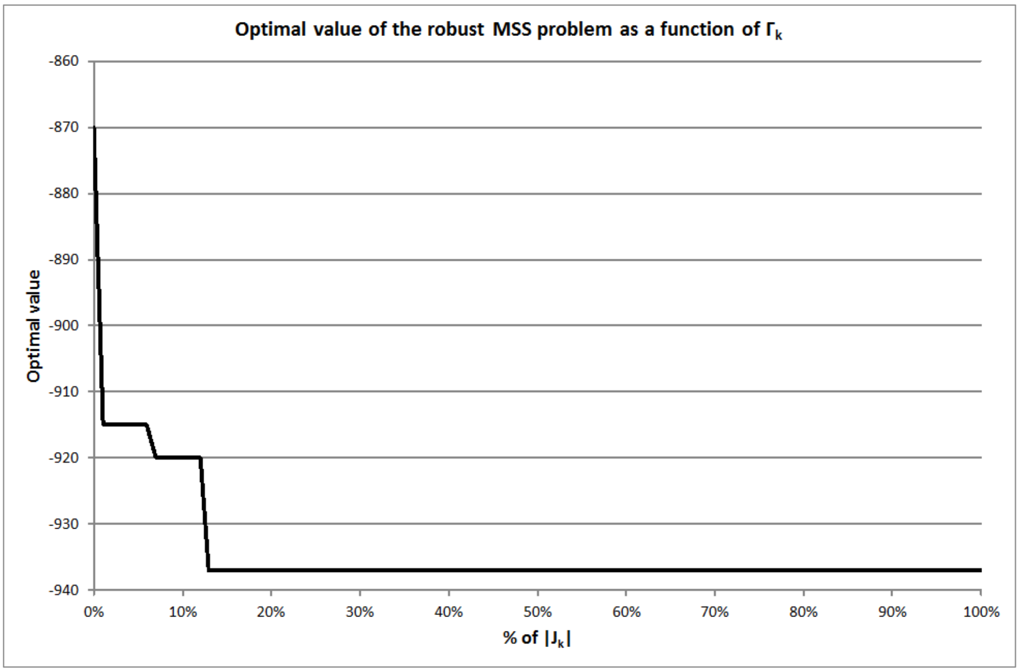


Figure 7.6: Optimal value as a function of Γ_k

An interesting characteristic of Figure 7.6 is the existence of 'phase transitions' of the optimal value as the level of protection increases. As soon as the protection level is set to 1% of $|J_k|$, the optimal value decreases rapidly from its value when $\Gamma_k = 0$ for all wards k. The optimal value then remains constant until 6% of $|J_k|$, after which it decreases again to another optimal value when the protection level is between 7% and 12% of $|J_k|$. A further jump to a decreased optimal value is present at 13% of $|J_k|$, where it remains constant for all levels of protection up to the maximum protection level $\Gamma_k = |J_k|$. When the graph remains constant for varying protection levels, it implies that the optimal value is insensitive to Γ_k, resulting in only a finite number of optimal values of the objective function.

Further instances of the model were investigated in order to determine whether this trend was specific to the above instance, or is a typical feature of the robust problem. Figure 7.7 shows how the optimal value varies with the protection level for twenty instances of the MSS problem.

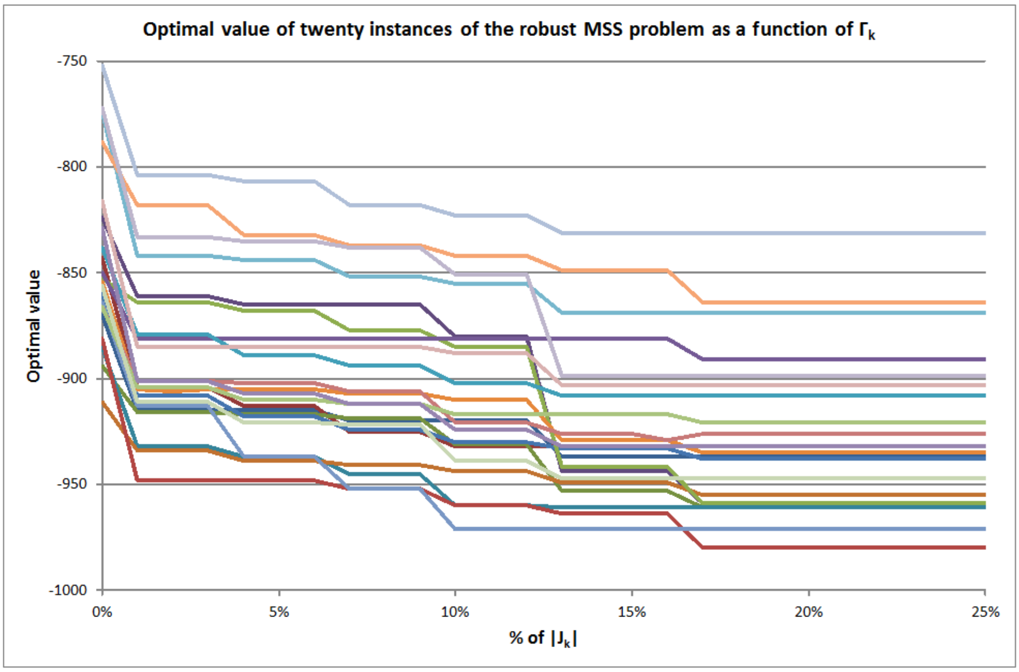


Figure 7.7: Optimal value as a function of Γ_k for 20 instances

The results of twenty instances of the robust MSS problem, as shown in Figure 7.7, confirm that the optimal value is a non-increasing function of the level of protection and step changes exist as the protection level increases. For all twenty instances of the MSS problem in Figure 7.7, there are several optimal values when the protection level takes values between 0 and 20% of $|J_k|$. However, the optimal value remains constant when the protection level is higher than 20% of $|J_k|$ for these twenty instances. This implies that it is not necessary to implement full protection, or indeed any more protection than 20% of $|J_k|$, against the uncertain bed requirement coefficients, since the optimal value is insensitive to changes in higher values of Γ_k. The size of the effect of the protection level on the optimal value is shown in Table 7.7 for the same problem instance as in Figure 7.6.

$\Gamma_k = x\%$ **of** $\lvert J_k \rvert$	**Optimal value**	**% reduction**
0%	-870	0.00
1%	-915	5.17
6%	-915	5.17
7%	-920	5.75
12%	-920	5.75
13%	-937	7.70
100%	-937	7.70

Table 7.7: Optimal value reduction as a function of the protection level

When the protection level is set to 0% of $|J_k|$, there is no protection applied to any of the uncertain bed constraints. If $\Gamma_k = 0$ for all wards k, the optimal value of the robust counterpart is the same as that of the nominal problem. This is due to the fact that there is no protection function in any of the bed constraints in Model 7.22, causing them to become equivalent to the bed constraints in the nominal problem. For validation purposes, this has been checked for many instances of the MSS problem for which the optimal value of the objective function and chosen plans for each specialty are the same for when $\Gamma_k = 0$ for all wards k as the nominal problem.

As soon as the protection level is increased to 1% of $|J_k|$ for all wards k, the optimal value is reduced by 5.17% for this particular instance. However, when the protection level is further increased to the maximum protection, the optimal value is only marginally worsened to 7.70%. This is a relatively small reduction in optimal value for full protection against the uncertain bed constraints. Bertsimas and Thiele [32] discuss that this is an advantage of this approach to RO.

7.5.2 The Effect of the Protection Level on the Feasibility of Solutions

Given that there are a finite number of optimal schedules for each level of Γ_k, it is of interest to check whether these different schedules remain optimal for different values of Γ_k. The same instance of the MSS problem as in Figure 7.6 will be used for this experiment. There are four different optimal schedules for this instance as Γ_k ranges from 0 to $|J_k|$ for all wards k. The solution to the nominal problem (when $\Gamma_k = 0$ for all wards k) was found to be an infeasible solution to the robust counterpart problem for all values of $\Gamma_k > 0$. The next optimal solution that was found for values of Γ_k between 1% and 6% of $|J_k|$ was found to be an infeasible solution to the problem when values of $\Gamma_k > 7\%$ of $|J_k|$ were used. However, this solution was found to be feasible for the problem when $\Gamma_k = 0$ for all wards k.

Similar comparisons were made for all four of the optimal solutions found for this instance of the MSS problem, and an interesting trend emerged. Solutions found for a certain protection level were also feasible solutions for lower levels of protection, but were infeasible for higher levels of protection. This trend is depicted in Figure 7.8 for this instance of the MSS problem.

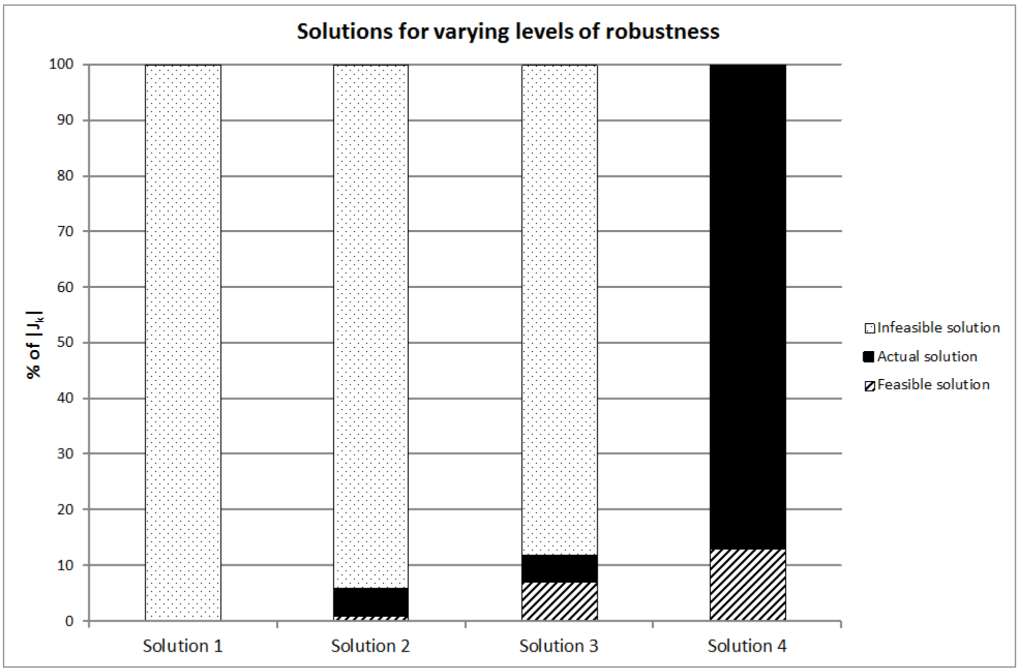


Figure 7.8: Feasibility of solutions for varying levels of Γ_k

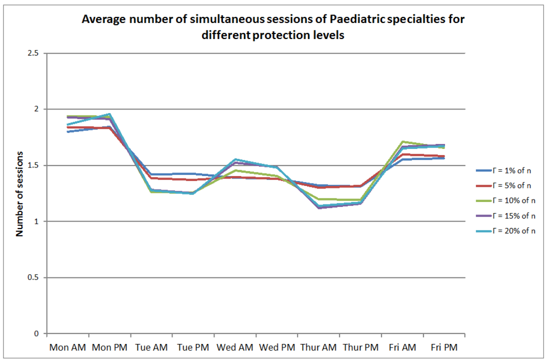
For twenty instances of the problem, it was found that only one optimal solution existed for each level of Γ_k from 5% to 20% of $|J_k|$.

7.5.3 The Effect of the Protection Level on Optimal Schedules

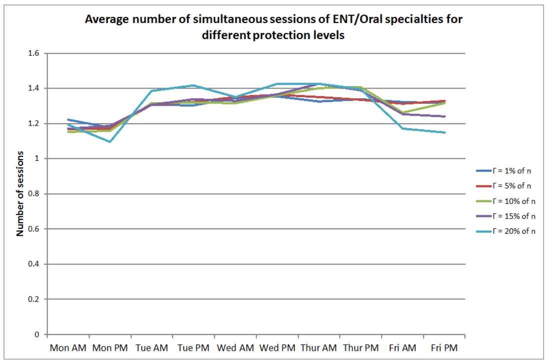
As seen in Sections 7.5.1 and 7.5.2, different optimal solutions are obtained for different protection levels, so it is of interest to investigate how the optimal schedules actually change for different protection levels. In particular, the spread of the simultaneous sessions throughout the week is analysed.

The levels of protection that were chosen for this analysis were 1%, 5%, 10%, 15% and 20% of $|J_k|$. $\Gamma_k >$ 20% of $|J_k|$ was not investigated because it was found in Section 7.5.1 that the optimal solutions become insensitive to Γ_k once $\Gamma_k >$ 20% of $|J_k|$.

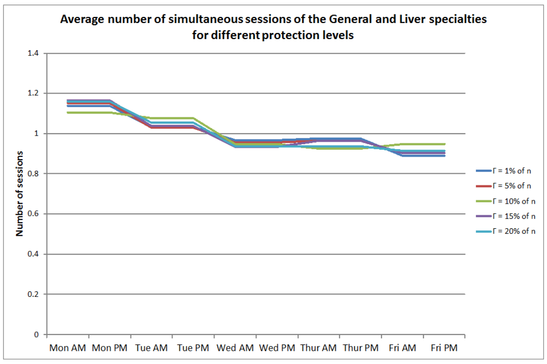
Figure 7.9 shows the average number of sessions (over 100 instances) that are scheduled simultaneously throughout the week for the specialties that send their patients to shared wards. The number of simultaneous sessions does not include the specialties that have fixed sessions in the MSS.



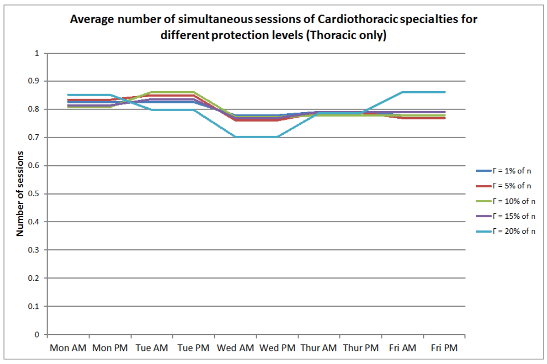
(a) Paediatric ward



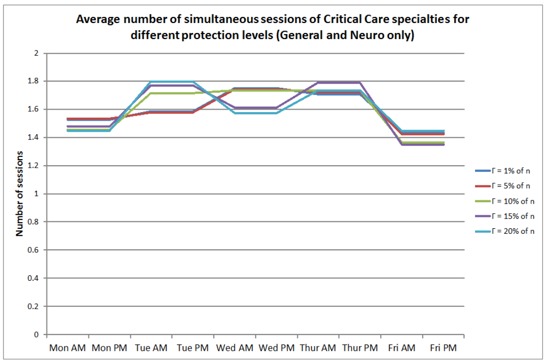
(b) ENT/Oral ward



(c) General/Liver ward



(d) Cardiothoracic ward



(e) CCU

Figure 7.9: Number of specialties that are scheduled simultaneously for different protection levels

For all wards in Figure 7.9, an increase in Γ_k does not seem to affect the overall trend of the number of simultaneous sessions throughout the week as was observed for an increase in $\hat{b}_{kj}^{(l)}$ in Section 7.4.2. However, an increase in Γ_k does seem to affect the magnitude of the variation in the number of simultaneous sessions throughout the week. Particularly for the Paeds, ENT/Oral and Cardiothoracic wards, the cyclic pattern in the graphs becomes more pronounced as Γ_k is increased from 1% of $|J_k|$ through to 20% of $|J_k|$. Figure 7.9(c) for the General/Liver ward does not show much change in the cyclic pattern when Γ_k is increased. For the

CCU in Figure 7.9(e), the shape of the cyclic pattern differs for higher values of Γ_k, but there is clearly a more pronounced cyclic shape with peaks on Tuesday and Thursday when $\Gamma_k = 15\%$ and 20% of $|J_k|$.

As Γ_k increases, more protection is applied to the uncertainty in the constraints, and it would appear from the graphs in Figure 7.9 that a more pronounced cyclic pattern in the number of simultaneous sessions is adopted by the optimisation model in order to be able to provide this additional protection.

These peaks in the number of simultaneous sessions, however, do not seem to correspond to the average LoS for each ward. This is illustrated in Table 7.8 for when $\Gamma_k = 20\%$ of $|J_k|$.

Ward	**Cycle length (days)**	**Ward average LoS (days)**
Paediatric	2	2
ENT/Oral	2	2.4
General/Liver	-	5.4
Cardtiothoracic	-	5.7
Critical Care	2	8.8

Table 7.8: Simultaneous sessions cycle length and ward length of stay when $\Gamma_k = 20\%$ of $|J_k|$

7.5.4 The Effect of the Protection Level on the Expected Bed Shortage

Table 7.9 contains the average expected bed count that was obtained from the simulation of 100 instances for different protection levels. As the level of protection increases, the average expected bed shortage decreases.

Γ_k	5%	10%	15%	20%
Average expected bed shortage	8.36	8.07	7.94	7.98

Table 7.9: Expected bed shortage for different protection levels

A Friedman test was conducted in order to determine whether there is a difference in expected bed count for different levels of protection (Γ_k). The conclusion that, at the 5% significance level, the expected bed count is statistically significantly different between different levels of protection (ranging from 5% to 20% of $|J_k|$) can be drawn, $\chi^2(3) = 30.155, p - value < 0.0005$.

Post hoc analysis using a series of Wilcoxon signed-rank tests was conducted at the 5% significance level in order to examine where the difference(s) in expected bed count actually occur. A Bonferroni correction was used, giving each pairwise comparison a significance level of 0.8%. The results of this analysis are summarised in Figure 7.10.

	5% of $\|J_k\|$	10% of $\|J_k\|$	15% of $\|J_k\|$	20% of $\|J_k\|$
5% of $\|J_k\|$		$p = 0.002$ **Difference**	$p < 0.0005$ **Difference**	$p < 0.0005$ **Difference**
10% of $\|J_k\|$			$p = 0.138$ **No Difference**	$p = 0.492$ **No Difference**
15% of $\|J_k\|$				$p = 0.754$ **No Difference**
20% of $\|J_k\|$				

Figure 7.10: Results of post hoc tests for differences in expected bed shortage between different levels of Γ_k

As can be seen in Figure 7.10, the post hoc analysis indicates that there is a significant difference in the expected bed shortage between the lowest value of Γ_k considered (5% of $|J_k|$) and all values of Γ_k higher than 10% of $|J_k|$. These results suggest that in order to significantly affect the expected bed shortage, the value of Γ_k should either be less than 5% or higher than 10%.

7.6 Results: Choosing Γ_k for a Given Bound on the Probability of Constraint Violation

The choice of Γ_k for each ward k is the responsibility of the decision maker, and can reflect the subjective views on uncertainty held by the decision maker. In the absence of any information on the uncertain system, it is not clear how to choose the values of Γ_k. An approach is proposed in [31] whereby the value of Γ_i for each constraint i is based on bounding the probability of constraint violation. This approach is less subjective, and will be investigated in this section.

7.6.1 Bound for the Probability of Constraint Violation

Following the theory of Bertsimas and Sim [31], under the assumption that the uncertain coefficients take values in the interval $[a_{ij} - \hat{a}_{ij}, a_{ij} + \hat{a}_{ij}]$, they prove that the probability of constraint violation, $P(a_i'x > b_i)$, is bounded by a function of Γ_i such that:

$$P(a_i'x > b_i) \leq B(n, \Gamma_i)$$

where the bound $B(n, \Gamma_i)$ is defined as:

$$B(n, \Gamma_i) = \frac{1}{2^n}\left\{(1-\mu)\binom{n}{\lfloor \nu \rfloor} + \sum_{l=\lfloor \nu \rfloor + 1}^{n} \binom{n}{l}\right\} \tag{7.23}$$

where $n = |J_i|$, $\nu = (\Gamma_i + n)/2$, and $\mu = \nu - \lfloor \nu \rfloor$.

Although this bound is the best possible, it is suggested in [31] to use another bound, since Bound (7.23) could involve computational difficulties in evaluating the sum of combination functions for large n. The bound suggested in [31] takes the form:

$$B(n, \Gamma_i) \leq (1-\mu)C(n, \lfloor \nu \rfloor) + \sum_{l=\lfloor \nu \rfloor + 1}^{n} C(n, l) \tag{7.24}$$

where

$$C(n, l) = \begin{cases} \frac{1}{2^n}, \text{ if } l = 0 \text{ or } l = n, \\ \frac{1}{\sqrt{2\pi}}\sqrt{\frac{n}{(n-l)l}} \exp\left(n \log\left(\frac{n}{2(n-l)}\right) + l \log\left(\frac{n-l}{l}\right)\right), \text{ otherwise.} \end{cases}$$

The decision maker may wish to define that the probability of constraint violation for constraint i should not exceed ϵ_i. By letting $B(n, \Gamma_i) = \epsilon_i$, a lower bound for the value of Γ_i for constraint i can therefore be found that ensures that the probability of constraint violation is at most ϵ_i.

For the robust counterpart, values of Γ_k will be chosen for each ward k in order that the probability of constraint violation is less than a specified bound ϵ_k. There can be a different bound for each ward, allowing the hospital decision maker to be more confident of constraint satisfaction on one ward than another. However, in the absence of information on bounds of probability of constraint violation for each individual ward, the same ϵ_k will be chosen for all wards k.

7.6.2 Values of Γ_k for Bounds on the Probability of Constraint Violation

It is not possible to rearrange Bound 7.24 to find an equation for Γ_k, so values of Γ_k will be selected to one decimal place in order to obtain a suitable bound on the probability of constraint violation. Values of ϵ_k ranging from 0.01 to 0.85 will be investigated in this section since $\Gamma_k = 0$ for all wards when $\epsilon_k > 0.85$. The values of Γ_k for each ward k that correspond to a certain value of ϵ_k are shown in Figure 7.11 below.

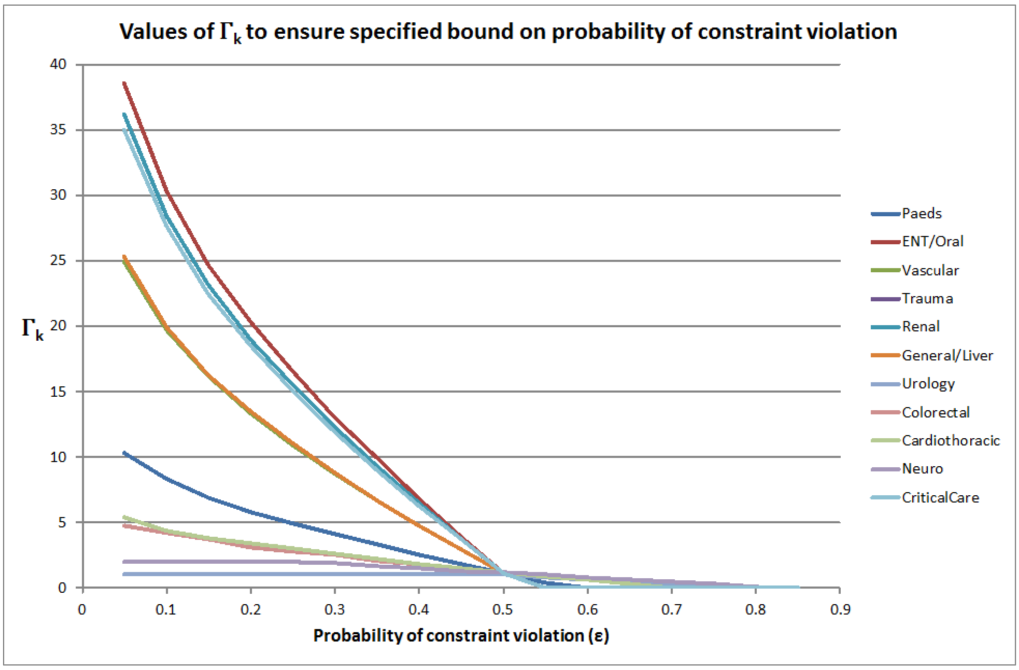


Figure 7.11: Values of Γ_k for varying bounds on probability of constraint violation

As can be seen in Figure 7.11, the general trend for each ward is that as ϵ (the bound on the probability of constraint violation) increases, the value of Γ_k decreases. This implies that as the decision maker becomes more willing to accept constraint violation, the model requires protection against fewer of the uncertain coefficients in each of the bed constraints for each ward. Wards with a higher number of uncertain coefficients in their bed constraints have higher values of Γ_k for lower bounds on the probability of constraint violation. The values of Γ_k for all wards steadily decrease as the bound on the probability increases, and eventually converge to $\Gamma_k = 0$ for all wards k once ϵ reaches 0.85.

There are two wards that are slight exceptions to this trend; Urology and Neurosurgery. Notice that in Figure 7.11, the value for Urology remains constant at $\Gamma_k = 1$ until ϵ exceeds 0.5, and the value for Neurosurgery remains constant at

$\Gamma_k = 2$ until ϵ exceeds 0.2. In order to achieve lower values of ϵ, higher values of Γ_k are required, however, for these two wards, Γ_k is capped at its upper bound n: one for Urology and two for Neurosurgery. Hence, the lowest possible value for the bound on the probability of constraint violation is $\epsilon = 0.5$ for Urology and $\epsilon = 0.25$ for Neurosurgery. These two 'special case' wards both have a very small number of uncertain coefficients in their bed constraints, which is a consequence of the very restrictive scheduling rules for these specialties that only allow for the generation of one and two possible plans respectively.

There appears to be a cross-over point in Figure 7.11 at $\epsilon = 0.5$, where the value of Γ_k decreases for some wards faster than other wards causing an intersection of the lines. This is shown in more detail in Figure 7.12. As ϵ exceeds 0.5, five out of the eleven wards have Γ_k values of zero. Higher values of ϵ cannot be used for these wards, since Γ_k is at its minimum allowed value.

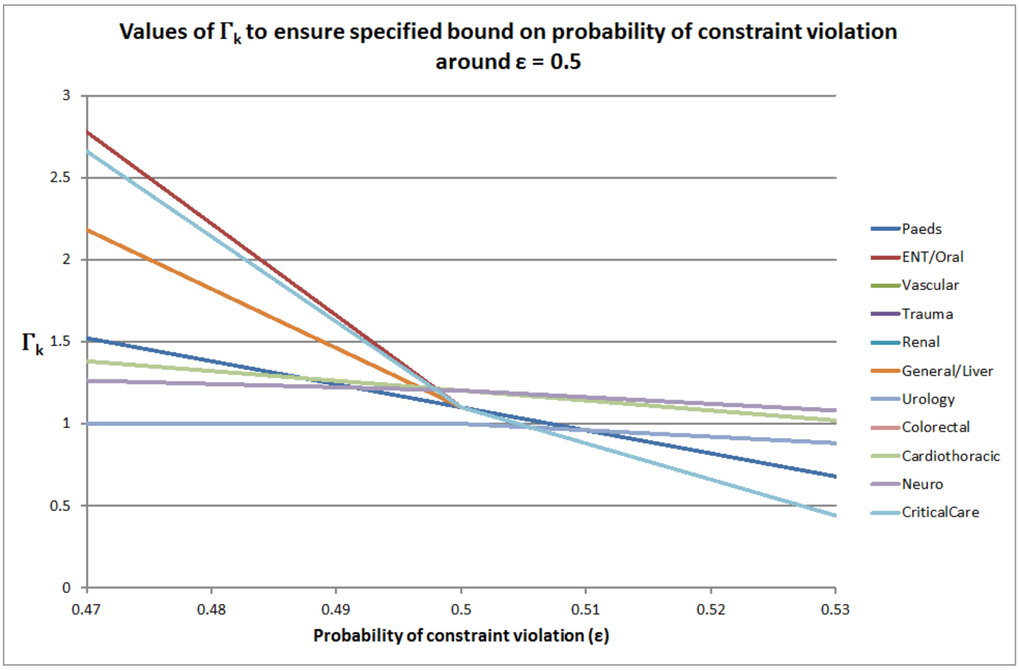


Figure 7.12: Values of Γ_k for varying bounds on probability of constraint violation for $\epsilon \approx 0.5$

7.6.3 The Effect of Epsilon on the Optimal Value

We have seen what effect the required bound on the probability of constraint violation has on the value of Γ_k, however, now its effect on the objective function will be investigated. In Figure 7.13, the optimal value of the objective function for one instance of the MSS problem is plotted for various bounds on the probability of constraint violation. It can be seen that as the bound on the probability of constraint violation increases, the optimal value of the MSS problem increases. In other words, as the decision maker becomes more willing to accept constraint violation, better optimal values are found since we are interested in maximising the objective function. Step-changes are also evident in this graph due to the finite number of optimal solutions for varying values of Γ_k.

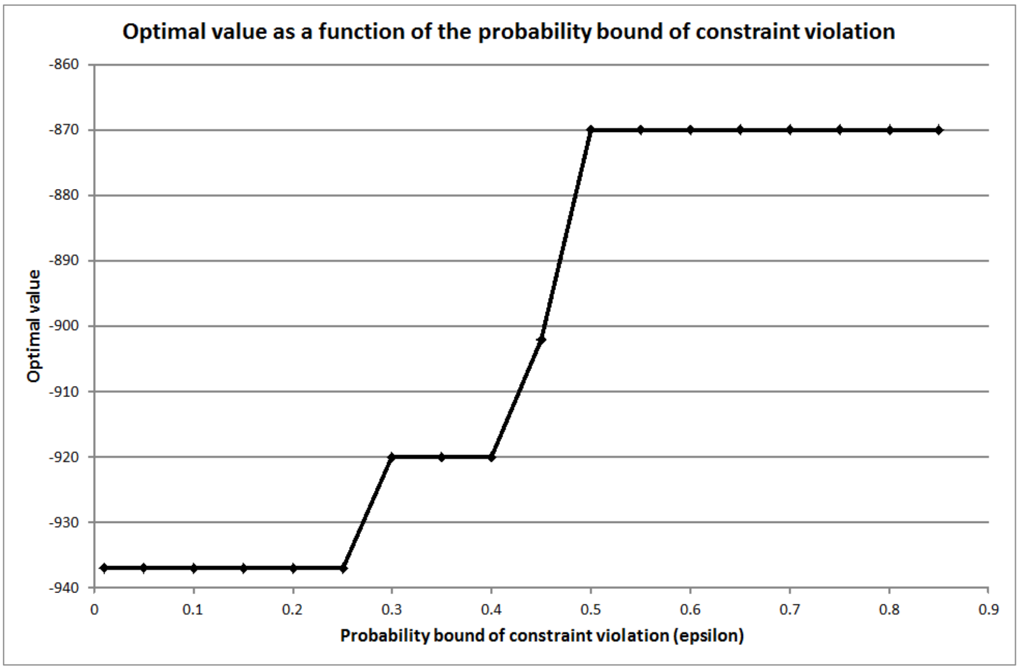


Figure 7.13: Optimal value as a function of the bound on the probability of constraint violation

Again, there is a pay-off to be struck between bounding the probability of constraint violation and achieving 'good' optimal values. A smaller level of ϵ implies a more robust solution will be produced, since as ϵ decreases, the required values of Γ_k increase, implying that more uncertainty is being protected against in the solution. Table 7.10 shows that although the optimal value only becomes compromised for values of $\epsilon \leq 0.5$, even to have a probability guarantee of at most 1% chance of constraint violation, the optimal objective value is only reduced by 7.7%. We can conclude that the quality of the optimal solutions are only marginally affected when ensuring low bounds on the probability of constraint violation.

ϵ	**Optimal value**	**Reduction (%)**
$(\Gamma_k = 0)$	-870	
0.85	-870	0.00
0.80	-870	0.00
0.75	-870	0.00
0.70	-870	0.00
0.65	-870	0.00
0.60	-870	0.00
0.55	-870	0.00
0.50	-902	3.68
0.45	-920	5.75
0.40	-920	5.75
0.35	-920	5.75
0.30	-937	7.70
0.25	-937	7.70
0.20	-937	7.70
0.15	-937	7.70
0.10	-937	7.70
0.05	-937	7.70
0.01	-937	7.70

Table 7.10: Optimal value reduction as a function of the bound on the probability of constraint violation, ϵ

7.6.4 An Evaluation of the Bound on the Probability of Constraint Violation

The proposed bound on the probability of constraint violation, Bound 7.24, is compared to the 'observed' probability of constraint violation from the simulation of the optimal schedules for varying levels of protection. Γ_k will take values ranging from 1% of $|J_k|$ to 20% of $|J_k|$; no higher since the problem becomes insensitive to Γ_k, resulting in the same optimal solutions being simulated. The probability of constraint violation is found for the bed constraints for all wards over all days, and is calculated for each simulation in the following way:

$$\text{Probability of constraint violation} = \frac{\text{No. of violated bed constraints}}{\text{No. of bed constraints}}$$

In Figure 7.14, the bound on the probability of constraint violation as calculated using Bound 7.24 and the probability of constraint violation from the simulation results are compared.

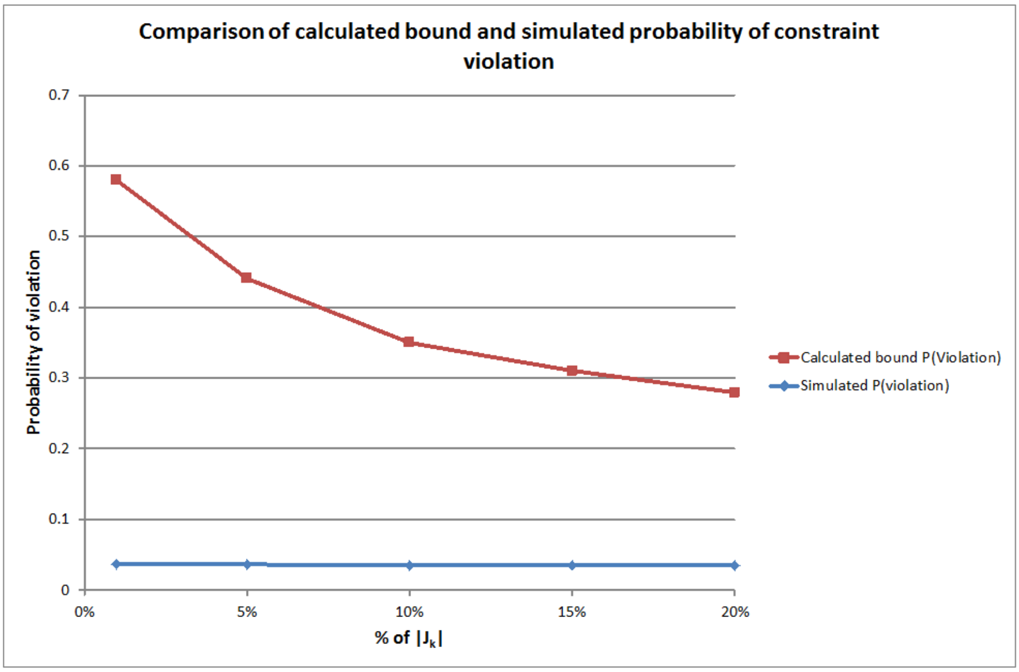


Figure 7.14: Calculated bound and simulated probability of constraint violation

As expected, the theoretical bound for the probability of constraint violation decreases as Γ_k increases, implying that a more robust model will result in fewer expected violated constraints. However, the probability of constraint violation found from the simulation is significantly lower than that predicted by Bound 7.25. This is not unexpected based on insights from [31], however, the scale of the difference between the theoretical and simulated results is surprising, as is the relative constancy of the simulated probabilities. The simulated probability of constraint violation has a mean value of 0.04, with a standard deviation of 0.004.

The simulated probability of constraint violation was found to be small and insensitive to Γ_k. This could be due to the fact that there are multiple protection levels in the robust counterpart, each of which have different absolute values on each ward k. Previous implementations of this bound in [31] involve just one constraint and thus one value of Γ. The probability of constraint violation has also been analysed for just one ward, for example the Paediatric ward, by using different values of $\Gamma_{\text{Paeds}} \in [0, |J_{\text{Paeds}}|]$ and $\Gamma_k = 0$ for all other wards k. This analysis also resulted in the probability of constraint violation for the Paediatric ward to remaining virtually constant for all values of Γ_{Paeds}.

It is suspected that the interaction between the different absolute values of

Γ_k for each ward k affects the overall probability of constraint violation for all wards over all days, rendering the comparison with Bound 7.24 inappropriate. For this reason, a related measure of how prevalent constraint violation was in the simulations was also investigated. The percentage of simulations in which at least one bed constraint was violated was analysed for 100 instances, the results of which are shown in Table 7.11.

Gamma = x% of J_k	% simulations
1%	87.7%
5%	87.5%
10%	86.8%
15%	86.3%
20%	86.5%

Table 7.11: Average percentage of simulations with at least one violated bed constraint as a function of Γ_k

As can be seen in Table 7.11, the average percentage of simulations in which at least one bed constraint is violated decreases slightly as the protection level increases. This implies that as a more robust model is adopted, it becomes less likely that any bed constraints will be violated in the simulation. It does not, however, provide any information on how many constraints are violated, and thus it does not give a measure of the scale of the problem of violated bed constraints in the simulation.

In conclusion, Bound 7.24 may be used as a guide for the decision maker when choosing values of Γ_k for all wards k. However, it has been shown not to be a tight bound on the resulting simulated probability of constraint violation, and thus should not be relied upon.

7.7 Conclusion

This chapter has presented the development of a robust counterpart formulation of the deterministic optimisation model for the construction of the MSS. This RO technique has been based on the theory developed by Bertsimas and Sim [31], and has been shown to be a suitable method for incorporating uncertainty about model coefficients within the optimisation model as discussed in Section 7.2.

An advantage to this RO technique is the use of a tuning parameter, Γ, within the model. This tuning parameter can reflect the decision maker's attitude toward uncertainty, thus varying the conservatism of the solution. A number of experiments have been carried out in Sections 7.4 and 7.5 in order to determine how the parameters chosen by the decision maker affect the solutions obtained, and the impact on key output measures relevant to the hospital. Key findings include the existence of a trade-off between obtaining a more robust solution versus the detrimental affect on the optimal objective function value, more robust schedules result in a lower expected bed shortage in the simulations, and a characteristic of more robust schedules is a clearly defined cyclic pattern of simultaneous sessions in the MSS for specialties that send their patients to shared wards.

Chapter 8

Scenario-Based Optimisation of the MSS

Scenario-based optimisation has been identified in Section 6.4 as a data-driven and logical way to deal with uncertainty in optimisation problems. When there exist a large number, possibly an infinite number, of possible realisations of uncertain parameters, a natural approach is to use 'data-driven' techniques that use these observations as 'scenarios' of uncertainty. In particular, the approach first developed by Calafiore and Campi [44] finds a solution that is optimal for all instances of uncertainty included in the optimisation problem (represented by known scenarios), and can be shown to remain feasible for the other omitted instances of uncertainty with high probability. This chapter explores the use of this scenario-based optimisation approach for the construction of the MSS, and is compared with the robust counterpart optimisation method used in Chapter 7.

8.1 Scenario-Based Constraint Sampling Optimisation

The approach first proposed by Calafiore and Campi in [44] considers scenarios that are based on constraint sampling in order to perform optimisation involving uncertainty. A finite set of constraints, chosen at random from all possible realisations of uncertainty, are included in the optimisation model. The model and theory developed by Calafiore and Campi is presented and extended in this section.

8.1.1 Sampled Convex Programs

Consider the following convex linear optimisation problem for which data uncertainty is assumed only in the constraints and is parameterised by an uncertainty parameter, $\delta \in \Delta \subseteq \mathbb{R}^l$

$$\begin{aligned} \min_{x\in\mathbb{R}^n} \quad & \mathbf{c^T x} \\ s.t. \quad & f(x,\delta) \leq 0 \\ & x \in \mathcal{X} \end{aligned} \tag{8.1}$$

where $\mathcal{X}$ is a convex and closed set. A single constraint is considered here without loss of generality, since multiple constraints $f_i(x,\delta) \leq 0$, $i = 1,..,m$ can be converted into a single constraint by taking $f(x,\delta) = \max_{i=1,..,m} f_i(x,\delta) \leq 0$.

Each realisation of uncertainty, δ, results in a different constraint, and hence could result in a different optimal solution to Problem (8.1) being found. It is assumed that δ is a random variable with probability P, implying that different realisations of uncertainty occur with a known probability. From all possible instances, N samples, $\delta^{(1)},...,\delta^{(N)}$, are chosen randomly and used to construct the so-called sampled convex program, SCP_N:

$$\begin{aligned} \min_{x\in\mathbb{R}^n} \quad & \mathbf{c^T x} \\ \text{s.t.} \quad & f(x,\delta^{(i)}) \leq 0, \quad \forall\ i = 1,...,N \\ & x \in \mathcal{X} \end{aligned} \tag{8.2}$$

The probability, P, according to which the uncertainty is sampled, may have different meanings in different problems. It could simply be the probability of occurrence of the different instances of δ, or it could reflect the importance placed on the different instances by the decision maker. Either way, the probability P may not be explicitly known, in which case the N sampled constraints are found directly from observations of the uncertainty.

By constructing the SCP_N, the one uncertain constraint in the original problem (8.1) is now represented by N linear constraints. The SCP_N is therefore a deterministic representation of the original stochastic problem and remains in standard convex program form. This is an advantage over some approaches in RO in which the robust counterparts do not maintain computational tractability (see Section 6.3).

Despite this advantage over RO, a price is paid in this scenario approach due to the fact that a solution found from considering N random scenarios of uncertainty is feasible for many instances of uncertainty, δ, but not all. Hence a critical question is: how many scenarios need to be included in the SCP_N in order to guarantee that the resulting optimal solution violates only a small proportion of the constraints that represent all instances of δ?

Calafiore and Campi [44] use statistical learning techniques to provide a bound on the number of total, omitted constraints that are possibly violated by the SCP_N. It was shown in [44] that the number of violated constraints rapidly decreases as the number of included constraints increases. Let the probability of constraint violation be defined by,

$$V(x) = P(\delta \in \Delta : f(x, \delta) > 0).$$

The main result from [44] is an upper bound of the probability of constraint violation in terms of the number of scenarios, N. The result is presented as follows.

Let $\hat{x}_N$ be the unique optimal solution to the SCP_N. $\hat{x}_N$ is itself a random variable due to the fact that the constraints $f(x, \delta^{(i)}) \leq 0$ are randomly selected, and hence depend on the chosen $\delta^{(1)}, ..., \delta^{(N)}$. Thus it was proved in [44] that the expected probability of constraint violation can be bounded as follows:

$$E_{P^N}[V(\hat{x}_N)] \leq \frac{n}{N+1}, \tag{8.3}$$

where n is the dimension of the decision variable x, and P^N is the probability measure in the space Δ^N. It can be seen that the expected probability of violation of $\hat{x}_N$ is proportional to the dimension of the decision variable n, and tends to zero linearly as N increases.

A parameter, $\epsilon \in [0, 1]$, is introduced as an upper bound for the probability of constraint violation. A solution $x \in \mathcal{X}$ is defined to be an ϵ-level robustly feasible solution if $V(x) \leq \epsilon$. Using Bound (8.3), it was shown in [44] that the optimal solution $\hat{x}_N$ of SCP_N is ϵ-level robustly feasible with probability of at least $1 - \beta$, where $\beta \in [0, 1]$, when

$$N \geq \frac{n}{\epsilon\beta} - 1. \tag{8.4}$$

Bound (8.4) is independent of the probability, P, of each scenario occurring and has been shown in [44] to hold irrespective of P. Hence, this bound still applies if P is unknown, as is common in many applications.

In later work, Calafiore and Campi [45] generalise Bound (8.3) further, since

Bound (8.4) is linear in β^{-1} and increases quickly for smaller, more desirable, values of β. The more generalised bound for the required number of scenarios is:

$$N \geq \left\lceil \frac{2}{\epsilon} \ln \frac{1}{\beta} + 2n + \frac{2n}{\epsilon} \ln \frac{2}{\epsilon} \right\rceil \tag{8.5}$$

Bound (8.5) results in a smaller number of required scenarios than for Bound (8.4), whilst maintaining the desired probabilistic level of the solution to the SCP_N through the use of ϵ and β. An exact formulation of the number of required scenarios that ensures that the probability of constraint violation does not exceed its desired level was later found by Campi and Garatti [46], and is given by Bound (8.6).

$$\mathbb{P}(V(\hat{x}_N) > \epsilon) \quad \leq \quad \sum_{i=0}^{n-1} \binom{N}{i} \epsilon^i (1-\epsilon)^{N-1} \tag{8.6}$$

A trade-off exists between the desire to set the probability of constraint violation, ϵ, at a small level, and the optimal performance of the SCP_N. Prior to running the optimisation, a-priori parameters ϵ and β are recommended in [44] to not be chosen too small due to limitations on the number of constraints that optimisation software can handle. Before running the optimisation, it is guaranteed by the above bounds that if N samples are drawn, the solution of the sampled convex program will be ϵ-level robustly feasible, with probability at least $1 - \beta$. The closer the value of β to zero, the higher the number of required scenarios, N. However, Calafiore and Campi [45] comment that β plays a very marginal role in practice. This is due to β appearing under the sign of a logarithm in the Bound (8.5). Hence β can be chosen as small as 10^{-10} or even 10^{-20}. In numerical examples in [44, 45], β takes values ranging from 0.1 to 0.0001. Once a solution to the problem has been obtained, an a-posteriori calculation of the feasibility level can be made through the use of simulation techniques as discussed in [44].

Calafiore [42] manipulates Bound (8.6) to provide a sufficient bound for the number of scenarios, N, that are required in order to guarantee that the probability of violation does not exceed ϵ with low probability, β, where $\epsilon, \beta \in [0, 1]$. Let the right-hand side of Bound (8.6) be denoted by the function $B_C(\epsilon)$. Then the aim is to find N such that:

$$\mathbb{P}(V(\hat{x}_N) > \epsilon) \quad \leq \quad B_C(\epsilon) = \sum_{i=0}^{n-1} \binom{N}{i} \epsilon^i (1-\epsilon)^{N-1} \quad \leq \quad \beta. \tag{8.7}$$

The Chernoff bound on the lower binomial tail [52] was then used to approximate $B_C(\epsilon)$ to:

$$B_C(\epsilon) \leq \mathrm{e}^{-\frac{(N\epsilon-n+1)^2}{2N\epsilon}}, \quad \text{for } N\epsilon > n. \tag{8.8}$$

An explicit and sufficient bound for the number of scenarios, N, can then be found to be:

$$N \geq \frac{2}{\epsilon}\left(\ln\frac{1}{\beta} + n\right) \tag{8.9}$$

8.1.2 Sampled Non-Convex Programs

All results and bounds on the number of required scenarios discussed in Section 8.1.1 relate to convex optimisation problems. Model 4.18 for the construction of the MSS is, however, a non-convex combinatorial optimisation problem due to the binary nature of the decision variables. Esfahani et al. [74] extend the analysis to non-convex, scenario-based optimisation. The reader is referred to Esfahani et al. for much of the underlying set-theoretic details which have been omitted here for sake of brevity.

Given a sampled convex program, SCP_N, as in Problem (8.2), consider a family of m solutions to the SCP_N which are indexed by k, i.e.,$(\mathcal{X}_k, f_k, \epsilon_k)_{k=1}^{m}$. Here $\mathcal{X}_k$ are closed and convex sets for $k = 1, .., m$, f_k are convex functions for $k = 1, .., m$, and ϵ_k is the constraint violation level for $\mathrm{SCP}_N^{(k)}$, where $\epsilon_k \in [0, 1]$. The sampled non-convex program (SNCP) is given in Model 8.10.

$$\begin{aligned} \min \quad & c^T x \\ \text{s.t.} \quad & x \vDash \bigcup_{k=1}^{m} \mathrm{SCP}_N^{(k)} \end{aligned} \tag{8.10}$$

where $x \vDash \mathrm{SCP}_N$ means that x is a feasible solution of SCP_N (conversely, $x \nvDash \mathrm{SCP}_N$ means that x is not a feasible solution of the SCP_N). This means that the SNCP seeks an optimal solution, $x^*_{N,k}$ say, that is feasible for at least one of the m SCP_N subprograms.

Adapting the methodology taken from Theorem 4.1 of Esfahani et al. [74] to the SCP_N problem, we extend the Bound (8.6) for a non-convex problem as follows:

Theorem: Feasibility of the SNCP.
The probability of constraint violation, $V(x^*_{Nk})$, at the ϵ-level, $\epsilon = (\epsilon_1, \ldots, \epsilon_m)$ where $\epsilon \in [0, 1]^m$, for the SNCP with the optimal solution x^*_{Nk}, where $x \in \mathbb{R}^n$, when N

constraints are included is,

$$\mathbb{P}\left(V(x^*_{Nk}) > \epsilon\right) \quad \leq \quad \sum_{k=1}^{m}\sum_{i=0}^{n-1}\binom{N}{i}\epsilon_k^i\left(1-\epsilon_k\right)^{N-i} \tag{8.11}$$

Proof. Let x^*_{Nk} be the optimal solution of $\mathrm{SCP}_N^{(k)}$, then

$$\begin{aligned}\mathbb{P}\left(x^*_{Nk} \nvDash \mathrm{SCP}_N^{(k)}\right) &= \mathbb{P}\left(V\left(x^*_{Nk}\right) > \epsilon\right) \\ &\leq \sum_{i=0}^{n-1}\binom{N}{i}\epsilon^i\left(1-\epsilon\right)^{N-i},\end{aligned}$$

from Theorem 1 of Campi and Garatti [46].
Let x^*_N be the optimal solution of SNCP_N, then $x^*_N \in (x^*_{Nk})_{k=1}^m$ since the optimal solution for SNCP_N is feasible for at least one of the k subproblems $\mathrm{SCP}_N^{(k)}$, $(k = 1, \ldots, m)$.
Thus,

$$\begin{aligned}\mathbb{P}\left(x^*_N \nvDash \mathrm{SNCP}_N\right) &\leq \mathbb{P}^N\left(\exists k \in (1,\ldots,m)\,|x^*_{Nk} \nvDash \mathrm{SCP}_N^{(k)}\right) \\ &\leq \sum_{k=1}^{m}\mathbb{P}^N\left(x^*_{Nk} \nvDash \mathrm{SCP}_N^{(k)}\right) \\ &\leq \sum_{k=1}^{m}\sum_{i=0}^{n-1}\binom{N}{i}\epsilon_k^i\left(1-\epsilon_k\right)^{N-i}\end{aligned}$$

□

Using the results from this scenario approach for non-convex problems, an explicit bound on the number of required scenarios, similar to that of Bound (8.9) is now derived. In the non-convex setting, let the right-hand side of Bound (8.11) be denoted by $B_{NC}(\epsilon)$ and bounded by β similarly as in Bound (8.7).

$$\mathbb{P}(V(\hat{x}_N) > \epsilon) \quad \leq \quad B_{NC}(\epsilon) = \sum_{k=1}^{m}\sum_{i=0}^{n-1}\binom{N}{i}\epsilon^i(1-\epsilon)^{N-1} \quad \leq \quad \beta \tag{8.12}$$

Note that the function for the non-convex problem, $B_{NC}(\epsilon)$, is the sum of the function for the convex problem, $B_C(\epsilon)$, for m subprograms, since

$$\begin{aligned} B_{NC} &= \sum_{k=1}^{m}\sum_{i=0}^{n-1}\binom{N}{i}\epsilon_k^i\left(1-\epsilon_k\right)^{N-i} \\ &= \sum_{k=1}^{m} B_C(\epsilon) \\ &= mB_C(\epsilon) \end{aligned}$$

where m is the number of convex subprograms used to approximate the original non-convex program. Using the same approximation of $B_C(\epsilon)$ as in Bound (8.8), $B_{NC}(\epsilon)$ is therefore approximated by,

$$B_{NC}(\epsilon) \leq m\mathrm{e}^{-\frac{(N\epsilon-n+1)^2}{2N\epsilon}}, \quad \text{for } N\epsilon > n. \tag{8.13}$$

Substituting this into Bound (8.12), an explicit bound for the number of scenarios, N, can be derived as follows:

$$\begin{aligned} m\mathrm{e}^{-\frac{(N\epsilon-n+1)^2}{2N\epsilon}} \leq \beta \quad &\Leftrightarrow \quad \frac{(N\epsilon-n+1)^2}{2N\epsilon} \geq \ln\left(\frac{\beta}{m}\right)^{-1} \\ &\Leftarrow \quad \frac{1}{2}N\epsilon + \frac{(n-1)^2}{2N\epsilon} + 1 \geq \ln\left(\frac{\beta}{m}\right)^{-1} + n \\ &\Leftarrow \quad \frac{1}{2}N\epsilon \geq \ln\left(\frac{\beta}{m}\right)^{-1} + n \\ &\Leftarrow \quad N \geq \frac{2}{\epsilon}\left(\ln\left(\frac{\beta}{m}\right)^{-1} + n\right) \end{aligned} \tag{8.14}$$

Hence, the number of scenarios, N, to include in the sampled non-convex program in order to guarantee that the probability of constraint violation does not exceed ϵ with probability at most β is given by Bound (8.14).

8.2 Developing a Scenario-Based Optimisation Model for the MSS

Scenarios will be used to represent possible realisations of the uncertain values of bed requirements in the bed constraints. If N scenarios are used in the optimisation model, it can be thought of as considering N random weeks of bed requirements in order to construct an optimal MSS. Multiple scenarios will be used in the optimisation model to generate robust schedules with bounds on the probability of constraint violation.

If more than one scenario is being considered, the objective function and bed constraints in Model 4.18 need to be modified in order to take into account the additional bed constraints for each scenario. Model 8.15 is the scenario-based optimisation model for the construction of the MSS that includes the scope for bed transference. Note that when the number of scenarios, N, is equal to one (i.e. $\sigma = 1$), the scenario-based model reduces to Model 4.18.

$$
\begin{aligned}
\min \quad & \sum_{\sigma=1}^{N}\sum_{k=1}^{p}\sum_{l=1}^{q}\Big(d_k^{(l)} - \sum_{j=1}^{n} b_{kj\sigma}^{(l)} x_j\Big) & & \qquad (8.15)\\
\text{s.t.} \quad & \sum_{j=1}^{n} a_{ij}x_j = 1 & \forall \quad & i = 1, ..., s\\
& \sum_{j=1}^{n} a_{ij}x_j \leq 1 & \forall \quad & i = s+1, ..., m\\
\sum_{j=1}^{n} b_{kj\sigma}^{(l)} x_j - \sum_{v=1}^{p} w_{kv} z_{vk\sigma}^{(l)} + & \sum_{v=1}^{p+1} w_{vk} z_{kv\sigma}^{(l)} = d_k^{(l)} & \forall \quad & k = 1, ..., p,\\
& & & l = 1, ..., q,\\
& & & \sigma = 1, ..., N\\
\sum_{k=1}^{p}\sum_{v=1}^{p} w_{kv} z_{vk\sigma}^{(l)} \leq & \sum_{k=1}^{p}\sum_{v=1}^{p+1} w_{vk} z_{kv\sigma}^{(l)} & \forall \quad & l = 1, ..., q,\\
& & & \sigma = 1, ..., N\\
& x_j \in \{0, 1\} & \forall \quad & j = 1, ..., n\\
& z_{kv\sigma}^{(l)} \geq 0 \text{ and integer} & \forall \quad & k = 1, ..., p, \;\; v = 1, ..., p+1,\\
& & & l = 1, ..., q, \;\; \sigma = 1, ..., N
\end{aligned}
$$

The scenario-based approach is the equivalent of having multiple B matrices in a single optimisation model. Each B matrix is generated using the conditional probability of failure approach as described in Section 4.2.4. The algorithm used to generate and append multiple B matrices onto the bed constraints in the optimisation model is given in Algorithm 7.

Algorithm 7 Generation of B Matrix for multiple scenarios

Let N = number of scenarios
for $i = 1$ **to** N **do**
  Generate the B matrix, B_i, using Algorithm 6, for scenario i
  Append B_i matrix onto the overall B matrix
end for

8.2.1 Number of Scenarios Required

The number of scenarios, N, that are required in the sampled program to ensure that the probability of constraint violation does not exceed $\epsilon \in [0, 1]$ with probability at least $1 - \beta, for \beta \in [0, 1]$, according to Bound (8.14) is investigated in this section.

Recall that Bound (8.14) is given by:

$$N \geq \frac{2}{\epsilon}\left(\ln\left(\frac{\beta}{m}\right)^{-1} + n\right)$$

The parameters defined by the decision maker are ϵ and β. Let the upper bound for the probability of constraint violation be set at $\epsilon = 0.1$, and let $\beta = 0.001$.

The values of the other parameters in the bound are found from the structure of the optimisation program. Since Model 8.15 is a non-convex optimisation problem, it will be approximated by m convex subproblems, as discussed in Section 8.1.2. For non-convex scenario problems involving binary decision variables, Esfahani et al. [74] set the number of subproblems to be $m = 2^l$, where l is the dimension of the decision variable. When applied to the construction of the MSS in UHW, the dimension of the binary decision variable, x, is $n = 1449$. Hence, the number of subprograms, m, required to approximate the sampled non-convex program into a union of sampled convex programs is $2^n = 2^{1449}$.

Substituting these values into Bound (8.14), the number of required scenarios is:

$$\begin{aligned} N &\geq \frac{2}{0.1}\left(\ln\left(\frac{0.001}{2^{1449}}\right)^{-1} + 1449\right) \\ N &\geq 49,203 \end{aligned} \tag{8.16}$$

Bound (8.16) implies that an extremely large number of scenarios are required in order to ensure the chosen levels of ϵ and β are guaranteed in the solution. In the sampled program, a scenario relates to a B matrix, so in practice there would be approximately $77 \times N$ constraints in the optimisation problem. This leads to very constrained problems, as well as large computer memory requirements. As will be discussed in Section 8.3.1, even when there are twenty scenarios ($N = 20$) included in the optimisation, the probability of finding an instance with a feasible solution reaches as small as 0.002. Therefore, it is considered impractical to use the number of scenarios that are suggested by Bound (8.16) in the sampled program for the construction of the MSS.

8.3 Results

The results of optimal schedules found from 100 instances were collected to investigate the effects of increasing the number of scenarios in the scenario-based optimisation model. The same values for model parameters that were used in the baseline scenario in Section 5.1.2 are used here when multiple scenarios are considered. We have checked and validated that the results for the one scenario problem are the same for the baseline scenario in Section 5.1.2. The current MSS used in UHW was not found to be a feasible solution to the single scenario problem in Section 5.1.1, so will not be investigated here for multiple scenarios.

8.3.1 The Effect of N on the Feasibility of Solutions

The results are analysed in order to examine the feasibility of the problem when multiple scenarios are included in the optimisation. Figure 8.1 shows the relationship between the number of scenarios and the resulting percentage of problem instances that did not have a feasible solution.

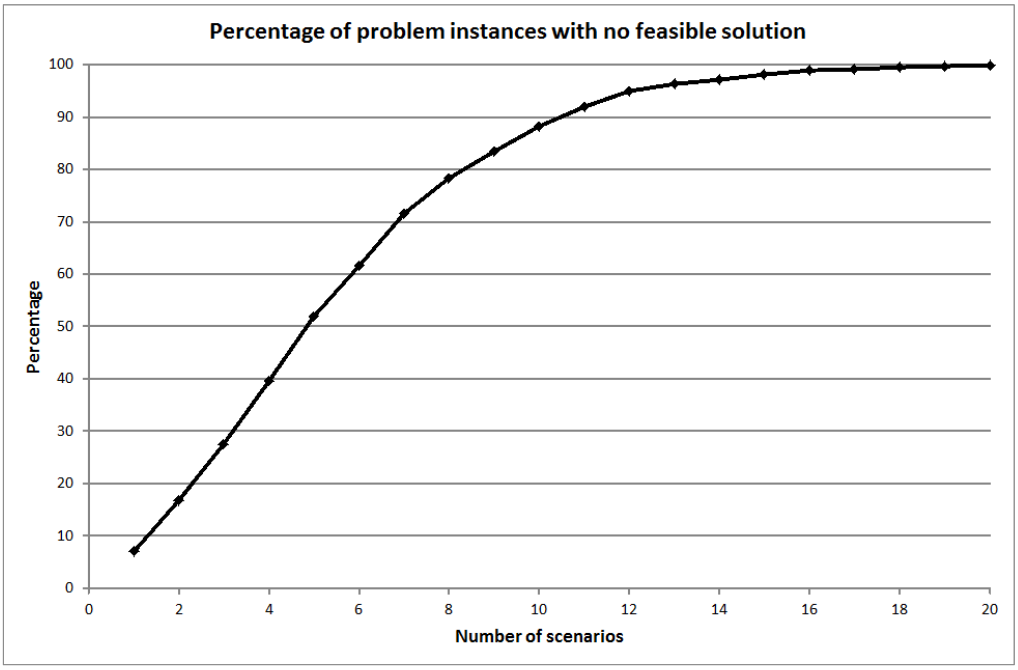


Figure 8.1: Percentage of problem instances with no feasible solutions

As can be seen in Figure 8.1, the percentage of problem instances that result in no feasible solutions increases as the number of scenarios increase. This is as expected due to the increase in the number of constraints when additional scenarios are included in the optimisation. As more scenarios are included in the optimisation model, a smaller feasible solution space is formed, causing a feasible solution less likely to be found.

When more than 10 scenarios are included in the optimisation model, a feasible solution is found less than 10% of the time. An instance in which a feasible solution exists is very rarely found (approximately around 0.2% of the time) when 20 scenarios are used. Hence, a maximum of 20 scenarios will be considered in all subsequent experiments.

8.3.2 The Effect of N on the Computational Time

The effect of including multiple scenarios in the optimisation model on the computational run time to perform the optimisation and simulation is shown in Figure 8.2.

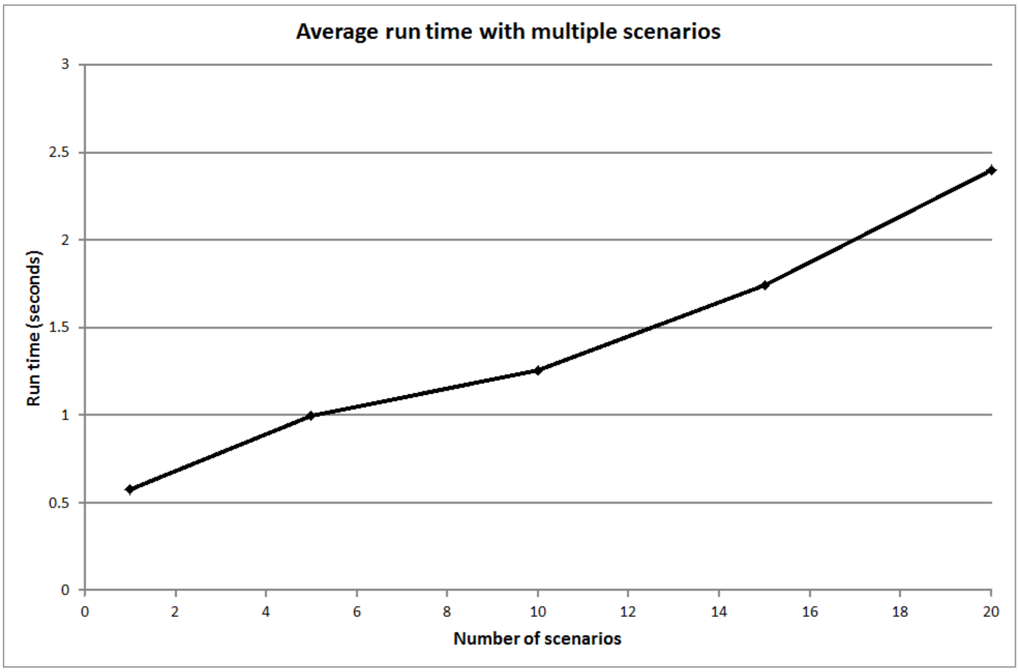


Figure 8.2: Run time as a function of the number of scenarios

As can be seen in Figure 8.2, the run time, normalised to one scenario, increases virtually linearly as the number of scenarios in the optimisation increases. Despite the reduced feasible solution space resulting from using additional scenarios, the run time of the model remains acceptable even when the maximum number of 20 scenarios is used (average of 2.4 seconds on a PC running Intel Core i3-2100 at 3.10 GHz with 4 GB RAM).

In order to determine a statistical relationship between the number of scenarios and the run time, a test on the Spearman correlation coefficient was performed. It is clear from Figure 8.2 that a positive relationship exists between the two, so a one-tailed test at the 5% significance level was performed. A correlation coefficient of $r = 0.863$ was found to be statistically significant at the 5% level.

8.3.3 The Effect of N on the Optimal Value

The average optimal value of the objective function per scenario, or equivalently the average number of unused bed days per week, is shown in Figure 8.3 for an increasing number of scenarios. The 95% confidence intervals are also shown to give an indication of the variation in the optimal value throughout the 100 instances investigated.

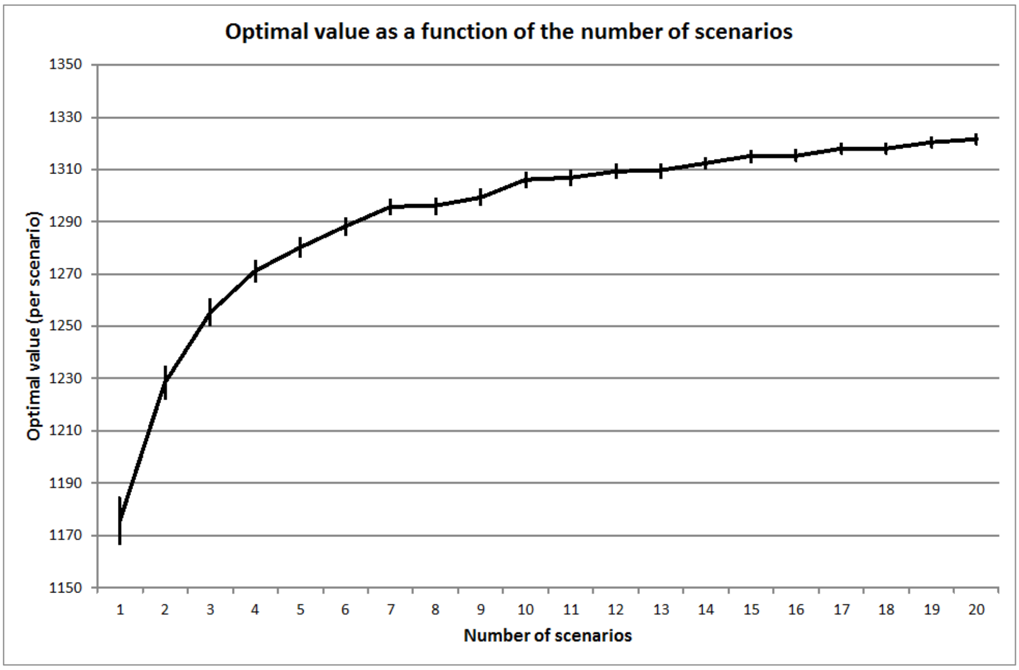


Figure 8.3: Optimal value as a function of the number of scenarios

As the number of scenarios, N, increases, the average optimal value per scenario of the objective function increases, with an increasing tendency to flatten off with increasing N. It is more intuitive to discuss the objective value in terms of the objective value per scenario since this reflects the number of unused bed days over all of the wards over one week, not N weeks. Using the definition of the objective function, this is equivalent to the average number of unused bed days over all wards increasing as the number of scenarios increases. Even though we are minimising the objective function, an increase in the number of unused bed days is desirable as this gives the potential for more patients to use the remaining bed days. In each of these multiple scenario experiments, the throughput of patients remains the same since the number of patients operated on per session remains constant. The LoS distributions are unaltered for each of these scenario experiments. The increase in unused bed days must therefore be a result of better chosen plans that means that fewer beds are required on the wards throughout the week.

As can be seen in Figure 8.3, the width of the 95% confidence intervals around the average optimal value per scenario get smaller as the number of scenarios increases. This is desirable, since it implies that the more scenarios that are included in the optimisation, the smaller the variation in the optimal value throughout the instances investigated.

Although the average optimal value per scenario increases with the number of scenarios, the rate of increase in the optimal value appears to decrease from around seven and more scenarios. Table 8.1 gives the percentage increase in the optimal value per scenario when more than one scenario is used when compared to using just one scenario.

Number of scenarios	**% increase in optimal value**
2	4.5
3	6.8
4	8.1
5	8.9
6	9.6
7	10.2
8	10.2
9	10.5
10	11.1
11	11.2
12	11.4
13	11.4
14	11.6
15	11.9
16	11.9
17	12.1
18	12.1
19	12.3
20	12.4

Table 8.1: Optimal value increase as a function of the number of scenarios

As can be seen in Table 8.1, the benefit of including multiple scenarios in the optimisation ranges from a 4.5% increase with two scenarios, to a 12.4% increase with twenty scenarios. It can be seen that in order to obtain a decent increase of 10% in optimal value per scenario, seven scenarios are required. Higher numbers of scenarios do not have such a great effect on the percentage increase in the optimal value, especially considering the increase in run time and infeasibility as discussed in Sections 8.3.1 and 8.3.2.

An increase in unused bed days can be seen as a positive outcome; if the existing levels of patient numbers can be accommodated in the wards using fewer bed days, then there is the potential to increase the throughput of patients in these wards. This could be done by increasing the number of patients operated on per operating session, if operating theatre resources allowed.

8.3.4 The Effect of N on Constraint Violation

As can be seen in Figure 8.4, the average percentage of simulations in which there is at least one violated bed constraint decreases as the number of scenarios increases. The more bed constraints included in the optimisation indicates a greater resilience of the optimal schedules to uncertainty.

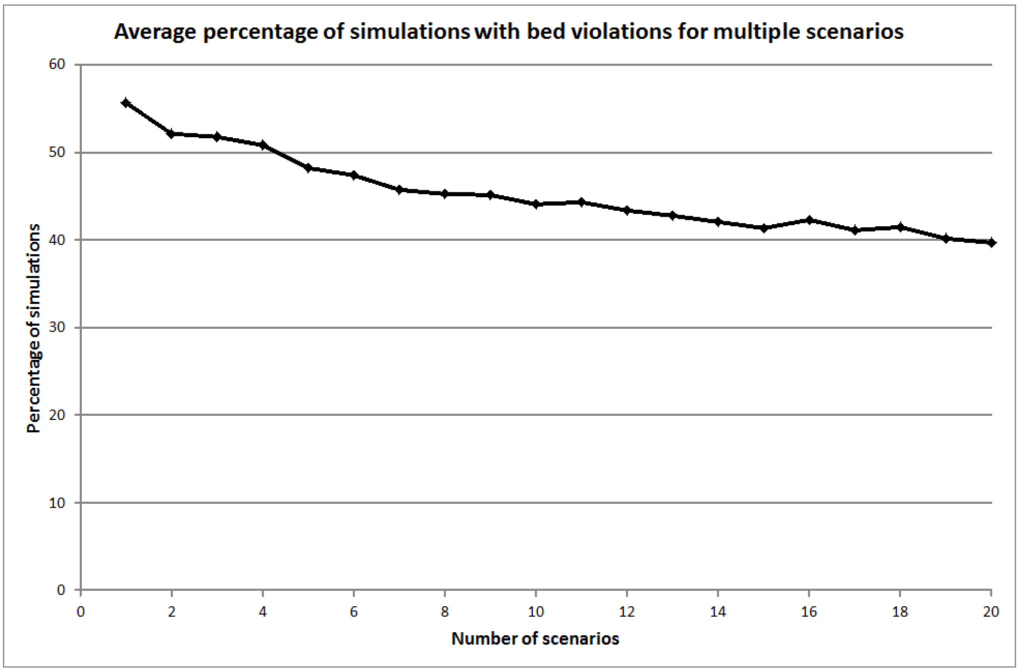


Figure 8.4: Percentage of simulations with violated bed constraints as a function of the number of scenarios

In order to investigate how common the violations of constraints are within the simulations, the proportions of the number of constraints that are violated in the simulations for each of the experiments of multiple scenarios are shown in Figure 8.5. It would appear that as the number of scenarios increases, fewer constraints are violated, on average, in each simulation of the optimal schedules.

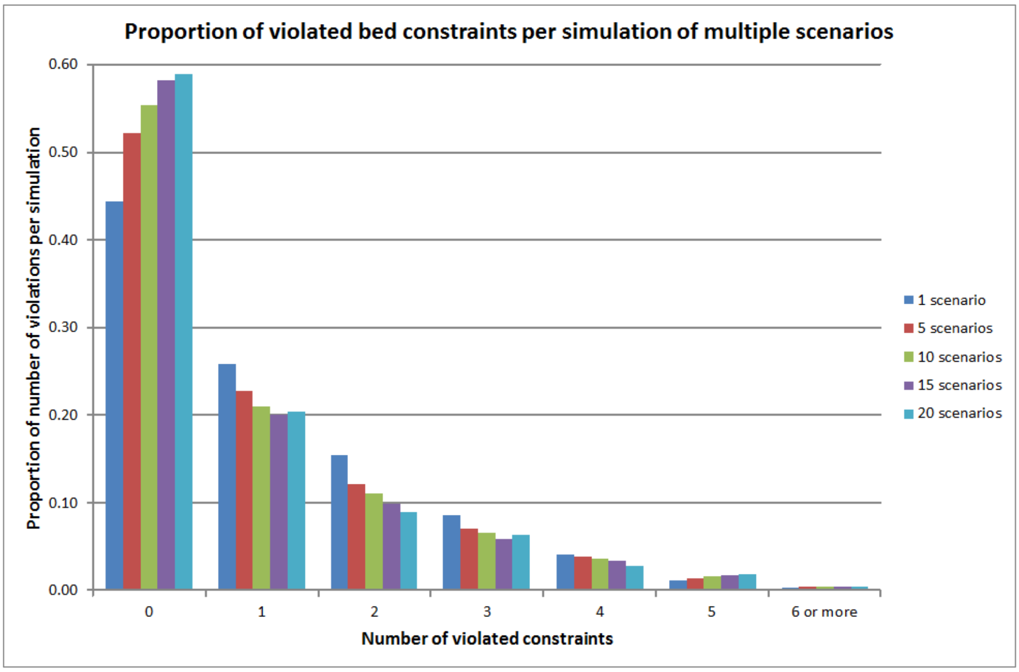


Figure 8.5: Violated bed constraints per simulation as a function of the number of scenarios

Overall, fewer simulations involve violated bed constraints, and in those that do, there is a trend of fewer constraints that are violated in each simulation as the number of scenarios increases. Hence, the schedules that are constructed using more scenarios are more likely to result in the bed requirements being satisfied by the beds available on the hospital wards.

As discussed in Section 8.2.1, the number of scenarios can be chosen such that the probability of constraint violation does not exceed ϵ with probability at most β. Calafiore and Campi [44] discuss how it is possible to compare the a-priori chosen value of β with the a-posteriori 'observed' values of β, however, it is not possible to do this here due to the size of the problem as suggested by Bound (8.16). The probability of constraint violation, in particular for the bed constraints, can, however, be examined and is shown in Figure 8.6.

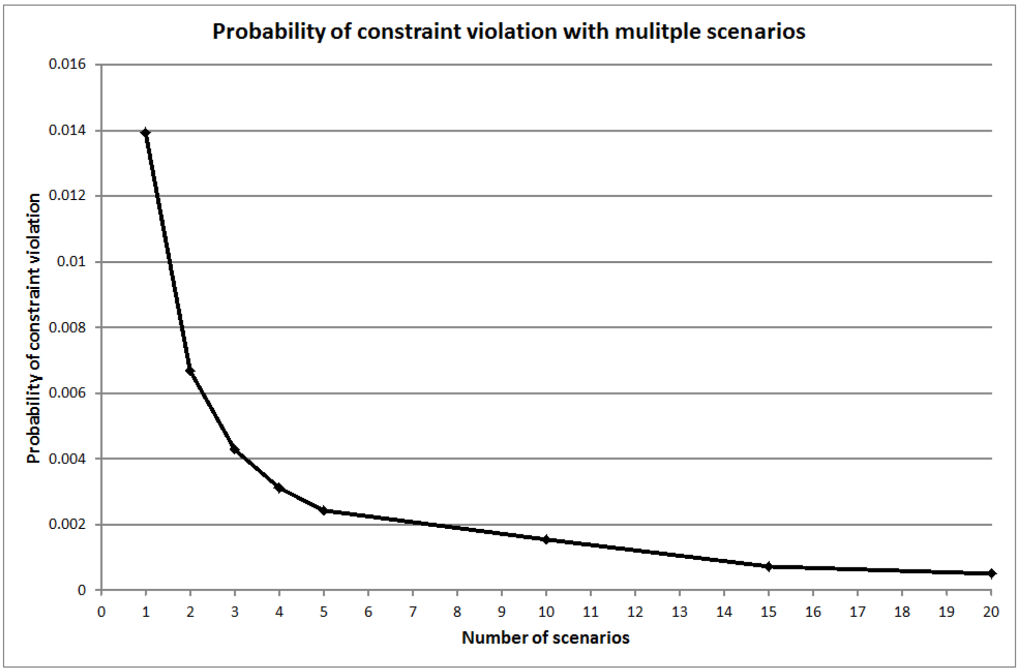
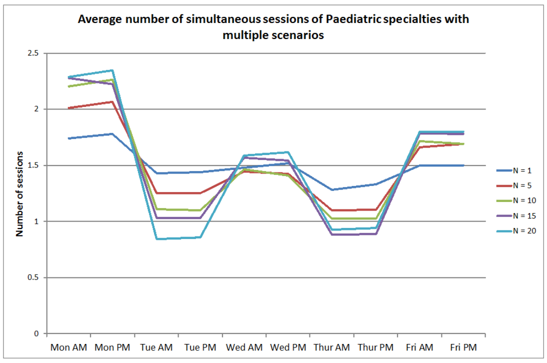


Figure 8.6: Probability of constraint violation as a function of the number of scenarios

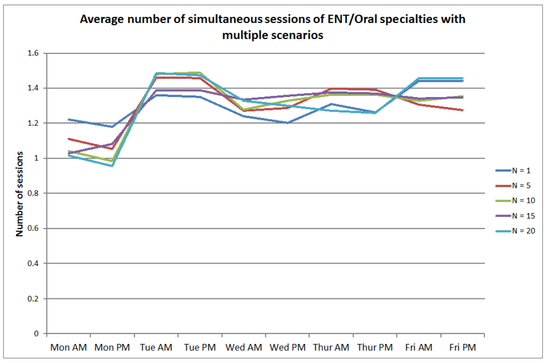
It can be seen in Figure 8.6 that the probability of constraint violation decreases as the number of scenarios in the scenario program increases. This is a desirable effect of including more scenarios in the optimisation and would indicate that the more bed constraints included in the optimisation, the greater resilience of the optimal schedules to uncertainty.

8.3.5 The Effect of N on Optimal Schedules

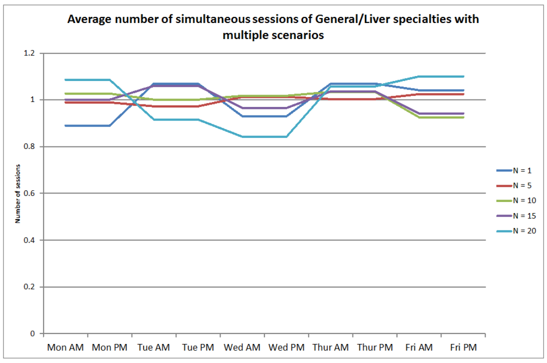
It is of interest to determine the effect of increasing the number of scenarios on the optimal schedules that are found from the optimisation. As in Section 5.1.1, the spread of the simultaneous sessions of specialties that send their patients to shared wards is analysed. Shared wards have been identified as 'pinch-points' in the system in Section 5.1.1, where violated bed constraints are more likely to occur. Figure 8.7 shows the average number of sessions, over 100 instances, that are scheduled simultaneously throughout the week for the specialties that send their patients to shared wards. The number of simultaneous sessions does not include the specialties that have fixed sessions in the MSS, as discussed in Section 5.1.1.



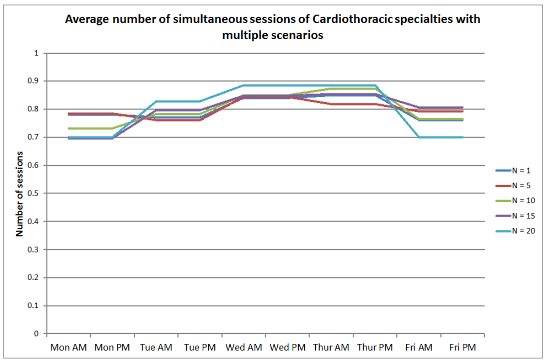
(a) Paediatric ward



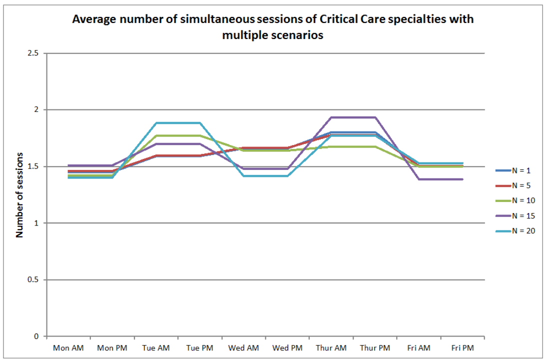
(b) ENT/Oral ward



(c) General/Liver ward



(d) Cardiothoracic ward



(e) CCU

Figure 8.7: Number of specialties that are scheduled simultaneously as a function of the number of scenarios

For most wards in Figure 8.7, as the number of scenarios in the sampled program increases, the trend in the graphs of the average number of simultaneous sessions appears to be that the peaks and troughs become more defined throughout the week. That is, the peaks get higher, i.e. more sessions are scheduled simultaneously at certain points in the week, and the troughs get deeper, i.e. fewer sessions scheduled simultaneously at other points in the week. This trend is evident in the graphs for all of the shared wards, except for the General/Liver ward in Figure 8.7(c). There does not appear to be any trend in this graph as the number of scenarios increases, except perhaps that fewer sessions are scheduled simultaneously during the middle of the week (Tuesday and Wednesday) and more at the beginning and end of the week (Monday, Thursday and Friday) for higher numbers of scenarios.

In all graphs for the shared wards in Figure 8.7, the number of simultaneous sessions follows the pattern from the baseline scenario (when $N = 1$) for all experiments that have multiple scenarios. This suggests that the results of the baseline scenario can be improved upon by including more scenarios in the optimisation, resulting in more pronounced pattern of variation in the schedule.

8.3.6 The Effect of N on the Expected Bed Shortage

Figure 8.8 shows that the trend in the expected bed shortage, averaged over the simulations of optimal schedules, decreases as the number of scenarios increases.

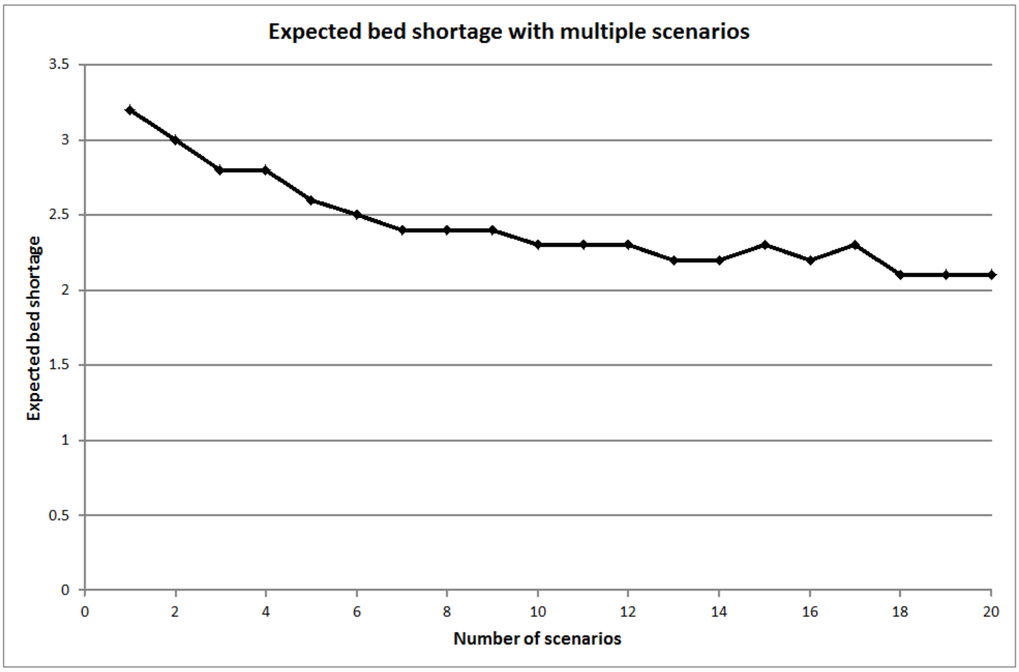


Figure 8.8: Expected bed shortage as a function of the number of scenarios

A lower expected bed shortage is desirable, as this implies that fewer cancelled operations might occur if an optimal schedule was to be implemented at UHW. These results indicate that as more scenarios are considered in the optimisation model, the resulting optimal schedules become more resilient to uncertainty, thus reducing the likelihood of cancellations.

8.4 Comparison of Optimisation Under Uncertainty Models

Two approaches to optimisation under uncertainty have been investigated in this research. In order to be able to make recommendations regarding which approach should be used for the construction of the MSS, a comparison of the two approaches is presented in this section. An evaluation of the strengths and weaknesses and performance measures of both approaches will be presented.

An advantage of the RO technique is the ability to explicitly set the level of protection through the parameter Γ_k. By being able to set the level of conservatism, the decision maker has a certain degree of control on the resulting MSS. This control is not present in the scenario-based technique. Despite this being an advantage to RO, it is also the technique's weakness – how is the value of Γ_k chosen? It can be hard to interpret the meaning of Γ_k for hospital managers without relying on the mathematical formulation of the optimisation model. Bertsimas and Sim [31] derived a method of choosing Γ_k that is based on the probability of constraint violation, however, it has been shown in Section 7.6.4 that this theory does not work well with the formulation of the RO model for the construction of the MSS. Bed transference between wards was not included in the RO model, since the protection function in each bed constraint had a similar function to the slacks and surpluses in the deterministic model. They both provide a safety buffer in each bed constraint to safe-guard against the uncertainty in bed requirements, so it was not considered appropriate to include both in the RO model. This has resulted in a loss of information with respect to the movement of patients between wards in the model.

The scenario-based optimisation technique is a natural extension to the deterministic model in the sense that multiple snapshots of bed requirement can be used simultaneously within the optimisation model for use in the bed constraints. This strength of the model lends itself to a simpler explanation of the optimisation model to hospital managers than for the RO model. This technique can also include bed transference in the bed constraints, further reflecting reality for implementation in the case study hospital. A disadvantage to using this technique, however, is the rapidly increasing complexity of the solution space as the number of scenarios increases. This results in a feasible solution being less likely to be found, or requiring a longer run time in order to find a feasible solution. Given that the construction of the MSS should be carried out every 3–6 months, a run time of a few hours is not considered unreasonable. Experiments showed that it was realistic to include up to 20 scenarios which is significantly less than the theory suggests.

A summary of the experimental results for both approaches, together with results from the deterministic model, is given in Table 8.2. The effects of increasing the level of robustness on the main performance measures for optimal MSSs are given for the RO and scenario-based models. This corresponds to the protection level, Γ_k increasing in the RO model, and the number of scenarios, N, increasing in the scenario-based optimisation model.

Performance Measure	**Deterministic (Baseline)**	**Robust Optimisation**	**Scenario-Based**
Unused bed days	1175.6	750 $\longrightarrow$ 1000	1175.6 $\longrightarrow$ 1330
% simulations with violated bed constraints	55.6%	87.7% $\longrightarrow$ 86.6%	55.6% $\longrightarrow$ 39.7%
Expected bed shortage	3.2	8.4 $\longrightarrow$ 8.0	3.2 $\longrightarrow$ 2.1
Instances with no feasible solution	7%	45.8% $\longrightarrow$ 81.4%	7% $\longrightarrow$ 99.8%

Table 8.2: Comparison of deterministic and optimisation under uncertainty methods

As can be seen in Table 8.2, an increase in the level of robustness has the same effect on the optimal value, expected bed shortage, the pattern of simultaneous sessions, and the percentage of instances with no feasible solutions in both models. The only performance measure for which the effect of robustness differs is the percentage of simulations with violated bed constraints. The level of robustness does not appear to have an affect on the percentage of simulations with violated bed constraints in the RO model, however, it is reduced in the scenario-based model. A reduction in the percentage of simulations with violated bed constraints is desirable, since it can be thought of as the failure rate of the MSS: will there be more beds required than available in a typical week? A lower failure rate is desirable because it implies that cancellations might also be lower.

Based on the strengths and weaknesses of both approaches, and the comparison of the performance measures, we can conclude that the scenario-based optimisation model is the better of the two approaches considered for the construction of the

MSS. Given this conclusion concerning optimisation under uncertainty, it is now of interest to ask whether it is better to use optimisation under uncertainty techniques to construct the MSS, or to continue with a more traditional deterministic model.

As can be seen in Table 8.2, the values of the performance measures are substantially worse for the RO model than for the deterministic model. There are fewer unused bed days, more simulations involve violated bed constraints, there is a higher expected bed shortage, and it is more likely to have an instance in which there is no feasible solution. Therefore, it is not advantageous to use the RO approach over using the deterministic model for the construction of the MSS. However, the values for the performance measure are better for the scenario-based optimisation model than for the deterministic model. There is at least the same number of unused bed days and percentage of instances with no feasible solution, and there is at most the same percentage of simulations with violated bed constraints and expected bed shortage. Hence, it is concluded that it is better to use the scenario-based optimisation model as opposed to the deterministic model. This is not surprising since the deterministic model is a special case of the scenario-based model with one scenario, and hence will be at least as good as the deterministic model.

Overall, we conclude that it is better to use scenario-based optimisation rather than RO for the construction of the MSS, and using at least two scenarios to improve the quality of the MSS in relation to cancellations and better flow of patients through the system.

8.5 Conclusion

This chapter has presented an alternative technique to that in Chapter 7 for dealing with uncertainty in optimisation problems. The scenario-based optimisation approach is data-driven and uses a subset of observations, or scenarios, of the uncertain values in an optimisation model. A review of the literature on this topic is given in Section 8.1, however, the previous literature mainly focuses on convex optimisation problems. An extension to the theory for non-convex problems is derived and presented here enabling the approach to be applied to the non-convex MSS problem. A scenario-based optimisation model was then developed for the construction of the MSS in Section 8.2, using scenarios to represent possible realisations of the uncertain bed requirements for the bed constraints.

Results from parameter experiments were collected and analysed in Section 8.3. It was found that including more scenarios in the optimisation problem resulted in a lower expected bed shortage in the simulations of optimal schedules, and produced a higher optimal objective function value. A higher objective value corresponds to a higher number of unused bed days, implying that the optimal schedules could accommodate the exisiting level of patient throughput in fewer bed days. This provides an opportunity for more patients to be operated on and use the beds on the wards, if capacity in the operating theatres allowed. However, a higher number of scenarios results in a more constrained solution space due to the increased number of bed constraints. This seems to be the price to pay for more robust schedules.

Chapter 9

Conclusions and Further Work

As discussed in Chapter 1, the objectives of this research concern the construction of the MSS when the demand for post-operative beds on wards is considered. This chapter aims to draw together the conclusions from preceding chapters, in addition to discussing the main outcomes of the research, and possible directions for future research.

Several optimisation models were developed and investigated in order to address the research objectives. Special consideration was made to include constraints on both the operating theatres and bed availability in the wards. Post-operative beds were largely ignored in the scheduling models reviewed in the literature, despite the authors commenting that bed availability affects the smooth running of the operating theatre. This was also highlighted by managers in UHW that bed availability, not theatre capacity, impacts on the MSS, and so was deemed important to include in the scheduling models developed here.

Novel formulations of a scheduling model for the construction of the MSS that are based on the set partitioning optimisation model have been developed. It is not known, to the best of our knowledge, that this technique has been used for this application prior to this research. The set partitioning technique has allowed for the sharing of bed between wards to be modelled, flexibility for the incorporation of soft and hard constraints relating to the operating theatres, and the enumeration of all allowable MSSs to be performed. The sharing of beds between wards, through the use of slack and surplus decision variables, is a particular feature of the model that serves to better reflect current practice in UHW. Optimal solutions can also be found in very quick run times.

Special consideration has also been made to incorporate the uncertainty associated with the post-operative bed requirements within the optimisation process. These optimisation under uncertainty techniques, again, not knowingly been previously used for the construction of the MSS, result in more robust MSSs which aid the reduction in the number of cancelled elective operations.

9.1 Conclusions

9.1.1 Operating Theatre Scheduling

It has been widely recognised throughout government reports, health board reviews and academic literature that careful planning and scheduling of the operating theatres is necessary in order to fully utilise these expensive resources, and to ensure any adverse affects on the rest of the hospital are minimised. Chapter 1 introduced the background to the scheduling problem associated with the operating theatres and the many factors that can affect their efficiency.

The tactical level of operating theatre planning – the construction of the MSS – is a challenging problem which has received much attention in academic literature. A comprehensive literature review of operational research publications is provided in Chapter 2, in which it was noted that mixed integer optimisation models are commonly used to address this problem. Simulation has also been demonstrated to be a useful tool in evaluating the performance of MSSs. Based on the opportunities for future research identified in the papers reviewed, it was considered that downstream hospital resources and the stochastic nature of many aspects of operating theatre scheduling should be included in any model developed for this research.

Data relating to operating theatres and surgical inpatient wards in UHW were provided by CaV UHB. The data analysis presented in Chapter 3 served to provide an understanding of how the operating theatres are currently used and provide a profile of the post-operative demand for beds on surgical wards. Findings from the data analysis were used as inputs into the developed models, including the number of operations per operating theatre session, post-operative LoS distributions, and the occurrence of emergency patients.

9.1.2 Deterministic Model for the Construction of the MSS

A deterministic model for the construction of the MSS was developed in Chapter 4. A set partitioning based optimisation model was derived that aims to find an MSS that minimises the number of unused bed days on wards over the planning horizon. This was achieved by assigning one surgical specialty to each operating theatre session, and ensuring that the number of beds required on each ward does not exceed the number of beds available.

A set partitioning optimisation model seems to be a natural model to use for the construction of the MSS since the ability to generate a number of candidate schedules as inputs into the optimisation model provides great scope for finding a suitable MSS for the hospital managers. The combinatorial nature of the scheduling problem also aligns with the characteristics of the set partitioning optimisation model. The developed model extends the basic formulation of a set partitioning optimisation model to include the novel use of bed constraints. The facility of bed transference between wards is also included in the bed constraints. This enables the modelling to reflect what happens in the case study hospital when beds are subject to particularly high demand, and patients are put on wards that are not necessarily assigned to their specialty.

The model formulation appears relatively simple when compared to other scheduling models reviewed in the literature, however, all important and necessary constraints that were relevant and identified by hospital managers are included in the model. Hard constraints concerning the operating theatres and bed availability are specified explicitly in the model formulation. Softer, preferential constraints are also included in the model via the generation of the operating theatre constraints. This prevents schedules that are deemed undesirable from the hospital manager's perspective from being generated by the optimisation. The simplicity of the model is considered an advantage over more complex models seen in the literature, particularly because the logical structure of the model can aid explanation of the modelling and the results to hospital managers. This attribute could therefore assist with the implementation of this scheduling model in the case study hospital.

The model was validated by comparing the patient throughput and bed count produced in the model with observations from the data. A number of experiments were then carried out in Chapter 5 to investigate the relationship between the MSS and the resulting bed requirements on surgical wards. One of the key findings was that the current MSS used in UHW is not a feasible solution to the deterministic optimisation model. The inclusion of the bed constraints resulted in this infeasibility, and certain wards were identified that had particularly high post-operative bed requirements. These were the wards that are shared between multiple specialties, and experienced an influx of demand when these specialties were scheduled simultaneously in the MSS. This insight lead to further investigations for optimal schedules in subsequent chapters.

Optimal MSSs were found when a baseline scenario that reflected current practice in UHW was used for model parameter values, and bed requirements were generated using the conditional probability of patients being discharged on each subsequent day after surgery. Optimal schedules using the baseline scenario parameter values were found to perform better than the current MSS used in UHW. The number of simultaneous sessions for the specialties that send their patients to a shared ward resulted in a flatter demand profile throughout the week. There were also fewer simulations that involved a higher predicted bed requirement than bed availability.

A series of 'what-if' scenarios were used to investigate the robustness of schedules and examine how they could cope with variation of the parameters in each experiment. Investigation into whether MSSs could be constructed under these 'what-if' scenario parameters was also undertaken. In particular, it was found that MSSs could be found for different levels of post-operative LoS, a different number of beds available on each ward, and different bed pools composed of wards sharing beds. The expected bed shortage was reduced when the post-operative LoS was reduced on the wards, the number of beds increased on each ward, and when more sharing between wards was allowed.

Despite the wealth of insights gained from the results, there are some limitations associated with this model. It is a deterministic model, since a 'snapshot' of bed demand is used to form the bed constraints. The post-operative bed requirements that form the bed constraints are, however, uncertain and cannot be known with surety, which is not reflected in the deterministic formulation. Investigation into the construction of the MSS under uncertainty was considered in later chapters.

9.1.3 Construction of the MSS Under Uncertainty

The availability of hospital resources, the length of time required for surgery or post-operative recovery is not always known with certainty. This can therefore make constructing the MSS a difficult process. Optimisation under uncertainty has been found to be a growing area of research, and a review of techniques was presented in Chapter 6. It was thought that incorporating uncertainty into the optimisation model should result in a more robust MSS, and thus provide more reassurance to hospital managers that an MSS will cope when bed requirements vary in reality.

Two optimisation techniques that embodied uncertainty in their formulation were applied to the problem. These were robust counterpart optimisation and scenario-based optimisation. Both techniques were identified in the literature as having been successfully applied to healthcare problems, with the ability to guard against the uncertainty associated with resources. Models were developed, in Chapters 7 and 8 respectively, using both of these techniques to account for the uncertainty associated with the bed requirements in the bed constraints. The resulting models are data-driven, in the sense that the stochastic behaviour of the uncertainty is unknown and so information from observed data can be used within the optimisation models.

The RO model developed in Chapter 7 involved the use of a user-specified uncertainty set for each parameter in the optimisation model. The uncertainty sets were informed from analysis of the UHW data and took the form of a range of values around a typical bed requirement value. An additional parameter was also introduced to each bed constraint to allow the decision maker to control the degree of conservatism of the solution. This protection level parameter can be thought of as controlling a safety buffer of beds within each bed constraint in order to protect the MSS from uncertainty. Both of these parameters were deemed quite relatable to the scheduling application. Hospital staff can easily state a range of values for the bed requirement on each ward given their experience, and managers can specify how 'protected' they would like the resulting MSS to be against the uncertainty in the bed requirements.

It was found that the wider the interval of the uncertainty set for each uncertain bed requirement in the model, the lower the optimal value of the objective function. This worsening of the optimal value implies that there are fewer unused beds on the wards, implying less slack in the system to cope with a possible increase in demand for beds. The expected bed shortage, however, seems to be insensitive to an increase in the width of the uncertainty interval.

As the protection level is increased, the optimal value of the objective function worsens. This is referred to as the 'price of robustness' in the literature, however, this trade-off can be off-set by the observation in the experiments that as the protection level is increased, the expected bed shortage decreases. A reduction in expected bed shortage can be interpreted as a reduction in the number of cancellations expected from an MSS. This provides great scope to achieve one of the research objectives specified in the introduction. It appears that the number of cancellations can be reduced by careful scheduling of the specialties that send their patients to one of the shared wards that have been identified as possible 'pinch-points' in the system. It was also found that as the protection level is increased, a more pronounced cyclic pattern that is aligned with the average LoS of each ward appears in the simultaneous sessions of these specialties within the MSS. The implication is that this cyclic pattern is a characteristic of a more robust schedule.

The scenario-based optimisation model developed in Chapter 8 naturally extends the deterministic model of Chapter 4 by including multiple instances, or scenarios, of the bed constraints within one optimisation model. Theory in the literature for convex optimisation problems suggests that the more scenarios included in the optimisation problem, the more robust an optimal solution will be towards uncertainty. This theory was extended to non-convex problems and applied to the construction of the MSS.

Results from experiments found that, by including more scenarios in the optimisation problem, a higher optimal objective function value was achieved. A higher objective value corresponds to a higher number of unused bed days, implying that the optimal schedules could accommodate more patients on the wards if other resources permitted. A lower expected bed shortage was also achieved by including more scenarios in the optimisation model, suggesting that fewer cancellations could be expected on implementation of an optimal MSS in the hospital. However, including a higher number of scenarios also implies a more constrained solution space due to the increased number of bed constraints in the optimisation model. This, again, is a price to pay for more robust schedules.

A comparison of the optimisation under uncertainty techniques was carried out in Section 8.4. Based on the strengths and weaknesses of both approaches, and the comparison of performance measures, it was concluded that it is better to use scenario-based optimisation rather than RO for the construction of the MSS. Using at least two scenarios, with a practical upper limit of 20, in the optimisation will improve the quality of the MSS in relation to cancellations and better flow of patients through the system.

9.1.4 Research Objectives: Revisited

In order to demonstrate that the research objectives of this project have been met, each objective, as outlined in Chapter 1, will be discussed in turn and references to specific sections within the thesis will be highlighted.

1. **Investigate the relationship between the MSS and the resultant bed demand on surgical wards.**

   Extensive data analysis was performed on the data provided by CaV UHB, and was presented and discussed in Chapter 3. The analysis showed the interdependency between the MSS and the availability of beds on the surgical wards, with many elective operations being cancelled due to a lack of available beds. The variability in the number of operations performed per session and in the LoS distributions was also highlighted, and can be seen as a contributing factor to the complexity of constructing an MSS. Information gleaned from the data analysis was used to inform the optimisation models that were developed in subsequent chapters.

2. **Understand the factors, if any, that affect why cancellations of elective operations occur, and identify whether they occur more frequently on particular wards.**

   Cancellation of elective operations has been found to be a problem in UHW, with over 18% being cancelled in 2012/13. Over half of these cancellations were caused by a lack of beds on surgical wards. As discussed in Section 3.6.2, this can be caused by patients of different specialties outlying on wards, thereby reducing the number of beds available to the assigned specialty. In Section 5.1.1, investigations into the current MSS used in UHW found that particularly high bed demand was experienced on wards that are shared by multiple surgical specialties. The cyclic pattern of operating theatre sessions that are scheduled simultaneously for specialties that send their patients to one of these shared wards was then investigated and found to be a key factor in influencing the bed demand. Schedules in which the simultaneous sessions were phased throughout the week resulted in a more levelled bed demand and fewer expected cancellations.

3. **Develop optimisation models to construct an MSS that satisfy constraints on both the operating theatres and bed availability on wards.**

   A deterministic optimisation model for the construction of the MSS was developed in Chapter 4. The model is based on a set partitioning optimisation model, and includes constraints on the operating theatre sessions and the bed availability on surgical wards. The novel use of slacks and surpluses within the bed constraints allowed for the sharing of beds between wards; this was later shown in Chapter 5 to help reduce the number of expected cancellations.

4. **Evaluate robust optimisation techniques for the construction of the MSS that incorporate the uncertainty associated with post-operative bed requirements.**

   In addition to the deterministic model developed in Chapter 4, two approaches to optimisation under uncertainty were applied to the MSS problem. A robust counterpart optimisation model was developed in Chapter 7 and a scenario-based optimsiation model extended the deterministic model in Chapter 8. It is believed that this is the first time these methods have been applied to the construction of the MSS taking into account post-operative bed requirements. The scenario-based optimisation model performed substantially better than the RO model, constructing schedules that resulted in a higher number of predicted unused bed days, fewer simulations with violated bed constraints, and a much lower expected bed shortage.

9.2 Further Work

As with any body of research, there is always scope for extensions to the work already carried out. Opportunities for further work associated with the research carried out in this thesis are proposed here.

The bed requirements predicted for each possible plan for the bed constraints in all optimisation models in this research have been calculated on a daily basis for each day after surgery. This was because the standard definition of a day in a hospital bed covers a 24 hour period with an overnight stay. In reality, however, patients can be discharged at any point throughout the day, meaning that if a patient is discharged in the morning, a new patient can be admitted to the same bed that afternoon. This is not currently reflected in the calculated bed requirements, as it is assumed that a bed is used by a patient for the whole day. Clearly this more detailed approach to calculating the bed requirements will affect the inputs for the optimisation models and the simulation of optimal schedules. It would be interesting to investigate whether this results in different optimal MSSs being found.

Given the data from CaV UHB, it was deemed most appropriate to schedule the specialties and predict their bed requirement based on the whole surgical specialty. It may however, be advantageous to model the surgical patients based on their specific surgical procedure, or at least by a group of procedures within each specialty. This would require further analysis of the data, and probably more discussion with hospital staff in relation to how sub-specialty groups can be determined. A more detailed picture of the time in surgery and LoS for each sub-group could again affect the optimal schedules that are found from the optimisation models. It would also be interesting to investigate whether the MSS should be specified in terms of whole specialties as is current practice, or by sub-specialty in which surgery of the same sub-group would take place in the operating theatre session.

There is additional scope to extend the theoretical work presented in Chapter 7 on RO to include bed transference between wards within each bed constraint. It was not included in the model developed due to the similarity between the interpretation of the slacks and surpluses for bed transference, and the protection function in the robust counterpart. Both act as a safety buffer in each bed constraint that can be seen to safeguard a number of beds in case of uncertainty associated with the number of beds required and those available. There are also concerns with including bed transference in the RO model in view of the already tight and problematic bed constraints for the shared wards.

A number of 'what-if' scenarios were considered in order to explore the affect of changes to parameters within the system. Namely, changes to post-operative LoS, the number of beds available on each ward, and the amount of sharing of beds between wards was explored. These scenarios were deemed the most relevant to hospital managers and also those that could be changed most easily or implemented in the case study hospital. There are, however, more scenarios that could be explored. For example, hospital managers have suggested that they might be moving towards the use of whole day operating theatre sessions as opposed to the current half-day sessions. This might allow for more operating time due to a reduction in set-up time throughout the day. Similarly, the UK government have indicated that elective patient services could be moved to a seven-day working week [3]. This would imply the MSS would have to span seven days as opposed to the current five-day working week. It would be interesting to examine how the wards would be able to cope with this extra demand for post-operative beds.

9.3 Final Reflections

The research contained in this thesis has identified the importance of systematic scheduling of the operating theatres within a hospital. The main research aim of this thesis was to develop a scheduling framework in which the affect of the MSS on other hospital resources and vice versa could be used to determine the best way in which specialties should be assigned to operating theatres.

A number of novel optimisation models, together with careful simulation of the resulting MSSs, have been developed and examined to investigate this research aim. Critical pinch-points within the system have been identified, namely the extreme demand experienced by wards shared by multiple specialties. The need to carefully schedule any simultaneous sessions that result in patients being sent to these wards is therefore of great importance, and the modelling approaches developed in this research appear to address this issue.

Finally, the importance of incorporating uncertainty associated with operating theatre scheduling has been demonstrated through the use of evaluating alternative optimisation under uncertainty techniques. To the best of our knowledge, this is the first time RO and scenario-based optimisation techniques have been applied to the construction of the MSS. It has been shown that better schedules can be produced using a robust, data-driven technique rather than a traditional deterministic model.

The insights gained from this research have the potential to aid the case study hospital, and indeed any other hospital that performs elective surgery, in the construction of an MSS that is robust to uncertainty associated with the demand for post-operative beds. Benefits of implementation of the modelling techniques developed here include a more levelled demand for ward beds throughout the week, and a reduction in the number of cancellations resulting from a shortage of beds. Findings from this research will be reported back to CaV UHB with the intention that this scheduling tool is piloted, resulting in an improved MSS being implemented for the main elective operating theatres.

Bibliography

[1] NHS beds by organisation and site, Welsh Government. https://statswales.wales.gov.uk/Catalogue/Health-and-Social-Care/NHS-Hospital-Activity/NHS-Beds/NHSBeds-by-Organisation-Site. Accessed: 21-03-2015.

[2] OPCS-4 Classification. http://systems.hscic.gov.uk/data/clinicalcoding/coding-standards/opcs4. Accessed: 20-03-2015.

[3] Seven day NHS pledge. http://www.theguardian.com/politics/2015/may/18/-cameron-seven-day-health-service-nhs-pledge-conservatives. Accessed: 18-05-2015.

[4] I. Adan, J. Bekkers, N. Dellaert, J. Jeunet, and J. Vissers. Improving operational effectiveness of tactical master plans for emergency and elective patients under stochastic demand and capacitated resources. *European Journal of Operational Research*, 213(1):290–308, 2011.

[5] I. J. B. F. Adan and J. M. H. Vissers. Patient mix optimisation in hospital admission planning: a case study. *International Journal of Operations & Production Management*, 22(4):445–461, 2002.

[6] B. Addis, G. Carello, A. Grosso, E. Lanzarone, S. Mattia, and E. Tnfani. Handling uncertainty in health care management using the cardinality-constrained approach: Advantages and remarks. *Operations Research for Health Care*, 4:1 – 4, 2015.

[7] S. Albers. Implicit enumeration algorithms for the set-partitioning problem. *OR Spectrum*, 2(1):23–32, 1980.

[8] A. Alonso-Ayuso, L. Escudero, and M. Ortuo. Modeling production planning and scheduling under uncertainty. In *Applications of Stochastic Programming*. MOS-SIAM Series on Optimization, 2005.

[9] T. Archer and A. Macario. The drive for operating room efficiency will increase quality of patient care. *Current Opinion in Anaesthesiology*, 19:171–176, 2006.

[10] C. Artigues, R. Leus, and F. Nobibon. Robust optimization for resource-constrained project scheduling with uncertain activity durations. *Flexible Services and Manufacturing Journal*, 25(1-2):175–205, 2013.

[11] T. Assavapokee, M. Realff, and J. Ammons. Min-max regret robust optimization approach on interval data uncertainty. *Journal of Optimization Theory and Applications*, 137:297–316, 2008.

[12] Audit Commission. Bed Management: Review of national findings, 2003.

[13] E. Balas and M. W. Padberg. Set partitioning: A survery. *SIAM Review*, 18(4):710–760, 1976.

[14] R. Baldacci, N. Christofides, and A. Mingozzi. An exact algorithm for the vehicle routing problem based on the set partitioning formulation with additional cuts. *Mathematical Programming*, 115(2):351–385, 2008.

[15] M. L. Balinski and R. E. Quant. On an integer program for a delivery problem. *Operations Research*, 12(2):300–304, 1964.

[16] C. Bandi and D. Bertsimas. Tractable stochastic analysis in high dimensions via robust optimization. *Mathematical Programming*, 134(1):23–70, 2012.

[17] C. Banditori, P. Cappanera, and F. Visintin. A combined optimization-simulation approach to the master surgical scheduling problem. *IMA Journal of Management Mathematics*, 24:155–186, 2013.

[18] J. Bard and H. Purnomo. Preference scheduling for nurses using column generation. *European Journal of Operational Research*, 2:510–534, 2005.

[19] J. Belien and E. Demeulemeester. Building cyclic master surgery schedules with leveled resulting bed occupancy. *European Journal of Operational Research*, 176(2):1185–1204, 2007.

[20] J. Belien, E. Demeulemeester, and B. Cardoen. Visualizing the demand for various resources as a function of the master surgery schedule: A case study. *Journal of Medical Systems*, 30(5):343–350, 2006.

[21] J. Belien, E. Demeulemeester, and B. Cardoen. A decision support system for cyclic master surgery scheduling with multiple objectives. *Journal of Scheduling*, 12(2):147–161, 2009.

[22] A. Ben-Tal, L. El-Ghaoui, and A. Nemirovski. *Robust optimization*. Princeton Series in Applied Mathematics. Princeton University Press, Princeton, NJ, 2009.

[23] A. Ben-Tal and A. Nemirovski. Robust convex optimization. *Mathematics of Operations Research*, 23(4):769–805, 1998.

[24] A. Ben-Tal and A. Nemirovski. Robust solutions of uncertain linear programs. *Operations Research Letters*, 25:1–13, 1999.

[25] A. Ben-Tal and A. Nemirovski. Robust solutions of linear programming problems contaminated with uncertain data. *Mathematical Programming*, 88:411–424, 2000.

[26] A. Ben-Tal and A. Nemirovski. Selected topics in robust convex optimization. *Mathematical Programming*, 112(1):125–158, 2008.

[27] D. Bertsimas and D. B. Brown. Constructing uncertainty sets for robust linear optimisation. *Operations Research*, 57(6):1483–1495, 2009.

[28] D. Bertsimas, D. B. Brown, and C. Caramanis. Theory and applications of robust optimisation. *SIAM Review*, 53:464–501, 2011.

[29] D. Bertsimas and V. Goyal. On the power of robust solutions in two-stage stochastic and adaptive optimisation problems. *Mathematics of Operations Research*, 35:284–305, 2010.

[30] D. Bertsimas and M. Sim. Robust discrete optimization and network flows. *Mathematical Programming B*, 98:49–71, 2003.

[31] D. Bertsimas and M. Sim. The price of robustness. *Operations Research*, 52(1):35–53, 2004.

[32] D. Bertsimas and A. Thiele. Robust and data-driven optimization: Modern decision making under uncertainty. In *INFORMS Tutorials in Operations Research*. INFORMS, 2006.

[33] D. Bertsimas and A. Thiele. A robust optimization approach to inventory theory. *Operations Research*, 54(1):150–168, 2006.

[34] H. Beyer and B. Sendhoff. Robust optimisation - a comprehensive survey. *Computer Methods in Applied Mechanics and Engineering*, 196(33-34):31903218, 2007.

[35] J. R. Birge and F. Louveaux. *Introduction to Stochastic Programming*. Springer, 2011.

[36] J. Blake, F. Dexter, and J. Donald. Operating room managers' use of integer programming for assigning block time to surgical groups: A case study. *Anesthesia & Analgesia*, 94:143–148, 2002.

[37] J. Blake and J. Donald. Mount sinai hospital uses integer programming to allocate operating room time. *Interfaces*, 32(2):63–73, 2002.

[38] J. T. Blake and M. W. Carter. Surgical process scheduling: a structured review. *Journal of the Society for Health Systems*, 5(3):17–30, 1997.

[39] J. T. Blake and M. W. Carter. A goal programming approach to strategic resource allocation in acute care hospitals. *European Journal of Operational Research*, 140(3):541–561, 2002.

[40] M. Bruni, P. Beraldi, and D. Conforti. A stochastic programming approach for operating theatre scheduling under uncertainty. *IMA Journal of Management Mathematics*, 26(1):99–119, 2015.

[41] E. K. Burke, P. D. Causmaecker, G. D. Maerea, J. Mulder, M. Paelinck, and G. V. Berghe. A multi-objective approach for robust airline scheduling. *Computers & Operations Research*, 37(5):822–832, 2010.

[42] G. Calafiore. On the expected probability of constraint violation in sampled convex programs. *Journal of Optimization Theory and Applications*, 143:405–412, 2009.

[43] G. Calafiore. Random convex programs. *SIAM Journal of Optimization*, 20(6):3427–3464, 2010.

[44] G. Calafiore and M. C. Campi. Uncertain convex programs: Randomised solutions and confidence levels. *Math. Programming Ser. A*, 102(1):25–46, 2005.

[45] G. Calafiore and M. C. Campi. The scenario approach to robust control design. *IEEE Transactions on Automatic Control*, 52(5):742–753, 2006.

[46] M. C. Campi and S. Garatti. The exact feasibilty of randomized solutions of uncertain convex programs. *SIAM Journal of Optimization*, 19(3):1211–1230, 2008.

[47] M. C. Campi and S. Garatti. A sampling-and-discarding approach to cnahce-constrained optimization: feasibility and optimisation. *SIAM Journal of Optimization*, 19(3):1211–1230, 2008.

[48] B. Cardoen, E. Demeulemeester, and J. Belien. Operating room planning and scheduling: A literature review. *European Journal of Operational Research*, 201(3):921–932, 2010.

[49] B. Cardoen, E. Demeulemeester, and J. Belien. Operating room planning and scheduling problems: a classification scheme. *International Journal of Health Management and Information*, 1(1):71–83, 2011.

[50] A. Charnes and W. W. Cooper. Chance-constrained programming. *Management Science*, 6(1):73–79, 1959.

[51] X. Chen, M. Sim, and P. Sun. A robust optimisation perspective on stochastic programming. *Operations Research*, 55:1058–1071, 2007.

[52] H. Chernoff. A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. *Annals of Mathematical Statistics*, 23:493–507, 1952.

[53] P. C. Chu and J. E. Beasley. Constraint handling in genetic algorithms: The set partitioning problem. *Journal of Heuristics*, 11(4):323–357, 1998.

[54] G. B. Dantzig. Linear programming under uncertainty. *Management Science*, 1:197–206, 1955.

[55] G. B. Dantzig. *Linear Programming and Extensions*. Princeton University Press, Princeton, NJ, 1963.

[56] E. Delage and Y. Ye. Distributionally robust optimization under moment uncertainty with application to data-driven problems. *Operations Research*, 58(3):595–612, 2010.

[57] R. Dembo. Scenario optimisation. *Annals of Operations Research*, 30(1):63–80, 1991.

[58] B. Denton, J. Viapiano, and A. Vogl. Optimization of surgery sequencing and scheduling decisions under uncertainty. *Health Care Management Science*, 10(1):13–24, 2007.

[59] J. Desrosiers, N. Mladenovi, and D. Villeneuve. Design of balanced mba student teams. *Journal of the Operational Research Society*, 56(1):60–66, 2005.

[60] F. Dexter, J. T. Blake, D. H. Penning, and D. A. Lubarsky. Calculating a potential increase in hospital margin for elective surgery by changing operating room time allocations or increasing nursing staffing to permit completion of more cases: A case study. *Anesthesia & Analgesia*, 94(1):138–142, 2002.

[61] F. Dexter, J. Ledolter, and R. E. Wachtel. Tactical decision making for selective expansion of operating room resources incorporating financial criteria and uncertainty in subspecialties' future workloads. *Anesthesia & Analgesia*, 100(5):1425–1432, 2003.

[62] F. Dexter, J. Ledolter, and R. E. Wachtel. Tactical decision making for selective expansion of operating room resources incorporating financial criteria and uncertainty in subspecialties' future workloads. *Anesthesia & Analgesia*, 100(5):1425–1432, 2005.

[63] F. Dexter and A. Macario. Changing allocations of operating room time from a system based on historical utilization to one where the aim is to schedule as many surgical cases as possible. *Anesthesia & Analgesia*, 94(5):1272–1279, 2002.

[64] F. Dexter, A. Macario, F. Qian, and R. D. Traub. Forecasting surgical groups' total hours of elective cases for allocation of block time. *Anesthesiology*, 91:1501–1508, 1999.

[65] F. Dexter, A. Macario, and R. D. Traub. Which algorithm for scheduling add-on elective cases maximizes operating room utilization? *Anesthesiology*, 91:1491–1500, 1999.

[66] F. Dexter, R. D. Traub, and P. Lebowitz. Scheduling a delay between different surgeons' cases in the same operating room on the same day using upper prediction bounds for case durations. *Anesthesia & Analgesia*, 92:943–946, 2001.

[67] G. Dullerud and F. Paganini. *Course in Robust Control Theory*. Springer-Verlag, New York, 2000.

[68] J. Dupačová, N. Gröwe-Kuska, and W. Römisch. Scenario reduction in stochastic programming: An approach using probability metrics. *Mathematical Programming, Series A*, 95:493–511, 2003.

[69] R. Düzgün and A. Thiele. Dynamic models for robust optimization. In *Wiley Encyclopedia of Operations Research and Management Science*. Wiley, 2011.

[70] R. Düzgün and A. Thiele. Robust optimisation with multiple ranges: Theory and application to pharmaceutical project selection. In *Proceedings of 14th INFORMS Computing Society Conference*, pages 103–118, 2015.

[71] L. El-Ghaoui and H. Lebret. Robust solutions to least-square problems to uncertain data. *SIAM Journal on Matrix Analysis and Applications*, 18(4):1035–1064, 1997.

[72] L. El-Ghaoui, F. Oustry, and H. Lebret. Robust solutions to uncertain semidefinite programs. *SIAM Journal on Optimization*, 9:33–52, 1998.

[73] E. A. Ernst, L. S. Lasdon, L. E. Ostrander, and S. S. Divell. Anesthesiologist scheduling using a set partitioning algorithm. *Computers and Biomedical Research*, 6(6):561–569, 1973.

[74] P. Esfahani, T. Sutter, and J. Lygeros. Performance bounds for the scenario approach and an extension to a class of non-convex programs. *IEEE Transactions on Automatic Control*, 60(1):46–58, 2015.

[75] P. Eveborn, P. Flisberg, and M. Rönnqvist. Laps care - an operational system for staff planning of home care. *European Journal of Operational Research*, 171(3):962–976, 2006.

[76] F. J. Fabozzi, D. Huang, and G. Zhou. Robust portfolios: contributions from operations research and finance. *Annals of Operations Research*, 176:191–220, 2010.

[77] J. E. Falk. Exact solutions of inexact linear programs. *Operations Research*, 24(4):783–787, 1976.

[78] H. Fei, C. Chu, and N. Meskens. Solving a tactical operating room planning problem by a column-generation-based heuristic procedure with four criteria. *Annals of Operational Research*, 166(1):91–108, 2009.

[79] M. L. Fisher. The lagrangian relaxation method for solving integer programming problems. *Management Science*, 27(1):1–18, 1981.

[80] B. A. Foster and D. M. Ryan. An integer programming approach to the vehicle scheduling problem. *Journal of the Operational Research Society*, 27(2):367–384, 1976.

[81] V. Gabrel, C. Murat, and A. Thiele. Recent advances in robust optimisation: An overview. *Eurpoean Journal of Operational Research*, 235(3):471–483, 2014.

[82] C. Gao, E. Johnson, and B. Smith. Integrated airline fleet and crew robust planning. *Transportation Science*, 43:2–16, 2009.

[83] M. Garey and D. Johnson. *Computers and Intractability: A Guide to the Theory of NP-Completeness*. W.H. Freeman and Company, New York, 1979.

[84] R. S. Garfinkel and G. L. Nemhauser. The set-partitioning problem: set covering with equality constraints. *Operations Research*, 17:848–856, 1969.

[85] Geer Mountain Software Corp. Stat::Fit, Version 2, 2015, South Kent, CT, http://www.geerms.com/.

[86] A. M. Geoffrion. Integer programming by implicit enumeration and balas' method. *SIAM Review*, 9(2):178–190, 1967.

[87] Y. Gerchak, D. Gupta, and M. Henig. Reservation planning for elective surgery under uncertain demand for emergency surgery. *Management Science*, 42(3):321–334, 1996.

[88] I. Gershkoff. Optimizing flight crew schedules. *Interfaces*, 19(4):29–43, 1989.

[89] J. Goh and M. Sim. Distributionally robust optimization and its tractable approximations. *Operations Research*, 58:902–917, 2010.

[90] F. Guerriero and R. Guido. Operational research in the management of the operating theatre: a survey. *Health Care Management Science*, 14:89–114, 2011.

[91] A. Guinet and S. Chaabane. Operating theatre planning. *International Journal of Production Economics*, 85(1):69–81, 2003.

[92] O. Hazir, M. Haouari, and E. Erel. Robust scheduling and robustness measures for the discrete time/cost trade-off problem. *Eurpoean Journal of Operational Research*, 207(2):633–643, 2010.

[93] M. Heng and J. Wright. Dedicated operating room for emergency surgery improves access and efficiency. *Canadian Journal of Surgery*, 56(3):167–174, 2013.

[94] K. L. Hoffman and M. W. Padberg. Solving airline crew scheduling problems by branch-and-cut. *Management Science*, 39(6):657–682, 1993.

[95] M. Holte and C. Mannino. The implementor/adversary algorithm for the cyclic and robust scheduling problem in health-care. *European Journal of Operational Research*, 226:551–559, 2013.

[96] A. Jebali, A. B. H. Alouane, and P. Ladet. Operating rooms scheduling. *International Journal of Production Economics*, 99(1-2):52–62, 2006.

[97] E. Kaplan and P. Meier. Nonparametric esitmation from incomplete observations. *Journal of the American Statistical Association*, 53(282):457–481, 1958.

[98] S. Kim and I. Horowitz. Scheduling hospital services:the efficacy of elective-surgery quotas. *Omega*, 30(5):335–346, 2002.

[99] A. Kleywegt, A. Shapiro, and T. H. de Mello. The sample average approximation method for stochastic discrete optimization. *SIAM Journal of Optimisation*, 12(2):479–502, 2001.

[100] P. C. Kuo, R. A. Schroeder, S. Mahaffey, and R. R. Bollinger. Optimization of operating room allocation using linear programming techniques. *Journal of the American College of Surgeons*, 197(6):889–895, 2003.

[101] M. Lamiri, X. Xie, A. Dolgui, and F. Grimaud. A stochastic model for operating room planning with elective and emergency demand for surgery. *European Journal of Operational Research*, 185(3):1026–1037, 2008.

[102] A. H. Land and A. G. Doig. An automatic method of solving discrete programming problems. *Econometrica*, 28(3):497–520, 1960.

[103] S. Lavoie, M. Minoux, and E. Odier. A new approach for crew pairing problems by column generation with application to air transportation. *European Journal of Operational Research*, 35:45–58, 1988.

[104] E. Lee and J. Wang. *Statistical Methods for Survival Data Analysis*. Wiley Series in Probability and Statistics, New Jersey, 2003.

[105] Y. H. Lee, J. I. Kim, K. H. Kang, and K. H. Kim. A heuristic for vehicle fleet mix problem using tabu search and set partitioning. *Journal of the Operational Research Society*, 59(6):833–841, 2008.

[106] D. Levine. *A parallel genetic algorithm for the set partitioning problem*. PhD thesis, Graduate School of the Illinois Institute of Technology, 1994.

[107] J. T. Linderoth, E. K. Lee, and M. W. P. Savelsbergh. A parallel, linear programming-based heuristic for large-scale set partitioning problems. *INFORMS Journal on Computing*, 13(3):191–209, 2001.

[108] C. Lu, S. Lin, and K. Ying. Robust scheduling on a single machine to minimize total flow time. *Computers & Operations Research*, 39:1682–1691, 2012.

[109] D. Lu and F. Gzara. The robust crew pairing problem: model and solution methodology. *Journal of Global Optimization, published online*, pages 1–26, 2014.

[110] R. M. Lusby, J. Larsen, M. Ehrgott, and D. M. Ryan. A set packing inspired method for real-time junction train routing. *Computers and Operations Research*, 2012.

[111] A. Macario, T. Vitez, B. Dunn, and T. McDonald. Where are the costs in perioperative care? analysis of hospital costs and charges for inpatient surgical care. *Anesthesiology*, 83(6):1138–1144, 1995.

[112] J. M. Magerlein and J. B. Martin. Surgical demand scheduling: A review. *Health Services Research*, 13:418–433, 1978.

[113] C. Mannino, E. Nilssen, and T. Nordlander. A pattern based, robust approach to cyclic master surgery scheduling. *Journal of Scheduling*, 15:553–563, 2012.

[114] E. Marcon and F. Dexter. Impact of surgical sequencing on post anesthesia care unit staffing. *Health Care Management Science*, 9(1):87–98, 2006.

[115] R. E. Marsten. An algorithm for large set partitioning problems. *Management Science*, 20:774–787, 1974.

[116] R. E. Marsten and F. Shepardson. Exact solution of crew scheduling problems using the set partitioning model: Recent successful applications. *Networks*, 11(2):165–177, 1981.

[117] J. May, W. Spangler, D. Strum, and L. Vargas. The surgical scheduling problem: Current research and future opportunities. *Production and Operations Management*, 20(3):392–405, 2011.

[118] F. Meng, J. Qi, M. Zhang, J. Ang, S. Chu, and M. Sim. A robust optimization model for managing elective admission in hospital. *Operations Research*, 2015 in press.

[119] P. Michaud. Exact implicit enumeration method for solving the set partitioning problem. *IBM Journal of Research and Development*, 16(6):573–578, 1972.

[120] A. Milburn. Operations research applications in home healthcare. In R. Hall, editor, *Handbook of Healthcare System Scheduling*, volume 168. Springer, 2012.

[121] D. Min and Y. Yih. Scheduling elective surgery under uncertainty and downstream capacity constraints. *European Journal of Operational Research*, 206(3):642–652, 2010.

[122] S. Mittal, A. Schulz, and S. Stiller. Robust appointment scheduling. In *Proceedings of the APPROX 2014 Conference*, 2014.

[123] J. M. Mulvey, R. J. Vanderbei, and S. A. Zenios. Robust optimization of largescale systems. *Operations Research*, 43(2):264–281, 1995.

[124] National Assembly for Wales. NHS Referral to Treatment Waiting Times: March 2015, 2015.

[125] NCEPOD. The report of the national confidential enquiry into perioperative deaths, 1990.

[126] A. Nemirovski and A. Shapiro. Convex approximations of chance constrained programs. *SIAM Journal of Optimisation*, 17(4):969–996, 2006.

[127] NHS Institute for Innovation and Improvement. Evaluation of The Productive Operating Theatre programme. Technical report, NHS Institute for Innovation and Improvement.

[128] NHS Institute for Innovation and Improvement. Reducing cancelled operations, 2015.

[129] OECD. *OECD Health Data 2001: A Comparative Analysis of 30 Countries.* Paris, 2001.

[130] S. N. Ogulata and R. Erol. A hierarchical multiple criteria mathematical programming approach for scheduling general surgery operations in large hospitals. *Journal of Medical Systems*, 27(3):259–270, 2003.

[131] B. Pagnoncelli, D. Reich, and M. Campi. Risk-return trade-off with the scenario approach: a case study in portfolio selection. *Journal of Optimization Theory and Applications*, 155:707–722, 2012.

[132] A. Rais and A. Viana. Operations research in healthcare: a survey. *International Transactions in Operational Research*, 18:1–31, 2010.

[133] N. J. Rezanova and D. M. Ryan. The train driver recovery problem - a set partitioning based model and solution method. *Computers & Operations Research*, 37(5):845–856, 2010.

[134] S. Robinson. Simulation model verification and validation: Increasing the users confidence. In *Proceedings of the 1997 Winter Simulation Conference. The Society for Computer Simulation, San Diego, CA*, pages 53–59, 1997.

[135] D. Ryan. Solution of massive generalized set partitioning problems. *Journal of the Operational Research Society*, 43(5):459–467, 1992.

[136] D. Ryan. Private Communication, 2012.

[137] D. Ryan and J. C. Falkner. On the integer properties of scheduling set partitioning models. *European Journal of Operational Research*, 35(3):442–456, 1988.

[138] D. Ryan and B. A. Foster. An integer programming approach to scheduling. *Computer Scheduling of Public Transport: Urban Passenger Vehicle and Crew Scheduling, A. Wren (ed.), North Holland, Amsterdam*, pages 269–280, 1981.

[139] H. Saadouli, B. Jerbi, A. Dammak, L. Masmoudi, and A. Bouaziz. A stochastic optimization and simulation approach for scheduling operating rooms and recovery beds in an orthopedic surgery department. *Computers and Industrial Engineering*, 80:72–79, 2015.

[140] P. Santibañez, M. Begen, and D. Atkins. Surgical block scheduling in a system of hospitals: an application to resource and wait list management in a british columbia health authority. *Health Care Management Science*, 10(3):269–282, 2007.

[141] R. Sargent. Verification and validation of simulation models. *Journal of Simulation*, 7:12–24, 2013.

[142] SAS Institute Inc. SAS Software, Version 9.4, 2015, Cary, NC, http://www.sas.com/.

[143] A. Shapiro and T. Homem-de-Mello. A simulation-based approach to two-stage stochastic programming with recourse. *Mathematical Programming*, 81(3):301–325, 1998.

[144] A. Soroudi and T. Amraee. Decision making under uncertainty in energy systems: state of the art. *Renewable and Sustainable Energy Reviews*, 28:376–384, 2013.

[145] A. L. Soyster. Convex programming with set-inclusive constraints and applications to inexact linear programming. *Operations Research*, 21:1154–1157, 1973.

[146] D. P. Strum, L. G. Vargas, and J. H. May. Surgical subspecialty block utilisation and capacity planning. *Anesthesiology*, 90:1176–1185, 1999.

[147] B. Tam, M. Ehrgott, D. Ryan, and G. Zakeri. A comparison of stochastic programming and bi-objective optimisation approaches to robust airline crew scheduling. *OR Spectrum*, 33(1):49–75, 2011.

[148] A. Testi, E. Tanfani, and G. Torre. A three-phase approach for operating theatre schedules. *Health Care Management Science*, 10(2):163–172, 2007.

[149] A. Thiele, T. Terry, and M. Epelman. *Robust linear optimization with recourse*. Technical report, Lehigh University, Bethlehem, PA, USA, 2010.

[150] Trisoft. Theatreman - theatre management system, July 2009.

[151] D. C. Tyler, C. A. Pasquariello, , and C. Chen. Determining optimum operating room utilization. *Anesthesia and Analgesia*, 96(4):1114–1121), 2003.

[152] M. van Houdenhoven, J. M. van Oostrum, E. Hans, G.Wullink, and G. Kazemier. Improving operating room efficiency by applying bin-packing and portfolio techniques to surgical case scheduling. *Anesthesia and Analgesia*, 105(3):707–714, 2007.

[153] J. M. van Oostrum, E. Bredenhoff, and E. Hans. Suitability and managerial implications of a master surgical scheduling approach. *Annals of Operational Research*, 178(1):91–104, 2010.

[154] J. M. van Oostrum, M. van Houdenhoven, J. L. Hurink, E. Hans, G.Wullink, and G. Kazemier. A master surgical scheduling approach for cyclic scheduling in operating room departments. *OR Spectrum*, 30(2):355–374, 2008.

[155] P. T. Vanberkel, R. J. Boucherie, E. W. Hans, J. L. Hurink, W. A. M. van Lent, and W. H. van Harten. An exact approach for relating recovering surgical patient workload to the master surgical schedule. *Journal of the Operational Research Society*, 62(10):1851–1860, 2011.

[156] J. M. H. Vissers, I. J. B. F. Adan, and J. A. Bekkers. Patient mix optimization in tactical cardiothoracic surgery planning: a case study. *IMA Journal of Management Mathematics*, 16(3):281–304, 2005.

[157] Wales Audit Office. Operating Theatres and Day Surgery, Cardiff and Vale University Health Board, 2013.

[158] Welsh Assembly Government. Designed for Life: Creating world class health and social care for Wales in the 21st century, 2005.

[159] Welsh Government. NHS Wales: Delivery Framework 2013-14 and Future Plans, 2013.

[160] W. E. Wilhelm. A technical review of column generation in integer programming. *Optimization and Engineering*, 2(2):159–200, 2001.

[161] G. Wullink, M. V. Houdenhoven, E. W. Hans, J. M. van Oostrum, M. van der Lans, and G. Kazemier. Closing emergency operating rooms improves efficiency. *Journal of Medical Systems*, 31(6):543–546, 2007.

[162] B. Zhang, P. Murali, M. M. Dessouky, and D. Belson. A mixed integer programming approach for allocating operating room capacity. *Journal of the Operational Research Society*, 60:663–673, 2009.