UNIVERSITY OF SOUTHAMPTON

FACULTY OF SOCIAL AND HUMAN SCIENCES

Mathematics

Developing a Multi-Methodological Approach to Hospital Operating Theatre Scheduling

by

Marion Louise Penn

Thesis for the degree of Doctor of Philosophy

March 2014

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF SOCIAL AND HUMAN SCIENCES

Mathematics

Doctor of Philosophy

DEVELOPING A MULTI-METHODOLOGICAL APPROACH TO HOSPITAL

OPERATING THEATRE SCHEDULING

by Marion Louise Penn

Operating theatres and surgeons are among the most expensive resources in any hospital, so it is vital that they are used efficiently. Due to the complexity of the challenges involved in theatre scheduling we split the problem into levels and address the tactical and day-to-day scheduling problems.

Cognitive mapping is used to identify the important factors to consider in theatre scheduling and their interactions. This allows development and testing of our understanding with hospital staff, ensuring that the aspects of theatre scheduling they consider important are included in the quantitative modelling.

At the tactical level, our model assists hospitals in creating new theatre timetables, which take account of reducing the maximum number of beds required, surgeons' preferences, surgeons' availability, variations in types of theatre and their suitability for different types of surgery, limited equipment availability and varying the length of the cycle over which the timetable is repeated. The weightings given to each of these factors can be varied allowing exploration of possible timetables.

At the day-to-day scheduling level we focus on the advanced booking of individual patients for surgery. Using simulation a range of algorithms for booking patients are explored, with the algorithms derived from a mixture of scheduling literature and ideas from hospital staff. The most significant result is that more efficient schedules can be achieved by delaying scheduling as close to the time of surgery as possible, however, this must be balanced with the need to give patients adequate warning to make arrangements to attend hospital for their surgery.

The different stages of this project present different challenges and constraints, therefore requiring different methodologies. As a whole this thesis demonstrates that a range of methodologies can be applied to different stages of a problem to develop better solutions.

List of Contents

List of Figures

List of Tables

Authors Declaration

I, ………Marion Louise Penn…………………………………,

declare that the thesis entitled;

………Developing a Multi-Methodological Approach to Hospital Operating Theatre ……… Scheduling …………………………………………………………………

and the work presented in the thesis are both my own, and have been generated by me as the result of my own original research. I confirm that:

1. this work was done wholly or mainly while in candidature for a research degree at this University;
2. where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
3. where I have consulted the published work of others, this is always clearly attributed;
4. where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
5. I have acknowledged all main sources of help;
6. where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
7. none of this work has been published before submission.

Signed: …………………………………………………………………..

Date:………………………………………………………………….

Acknowledgements

Definitions

GP - General Practice

ICU - Intensive Care Unit

IP - Integer Programing

FCFS – First Come First Served

FFD - First-Fit Decreasing

FIFO - First In First Out

LP - Linear Programming

MIP - Mixed Integer Programing

MS - Management Science

NHS - National Health Service

OR - Operational Research

PACU - Post Anaesthesia Care Unit

SA - Simulated Annealing

SD - System Dynamics

SODA - Strategic Options Development and Analysis

SPT- Shortest Processing Time

VBA - Visual Basic for Applications

VR - Vitreoretinal (a type of eye surgery)

Chapter 1: Introduction

Across the world operating theatres are among the most expensive of hospital resources and are in high demand with surgical groups competing for theatre time in many cases. In the UK hospitals face increasingly demanding targets for reducing waiting times and avoiding cancellations, whilst at the same time having to address the patient choice agenda (DoH, 2012b), payment by results (DoH, 2012a) and generally the need to work within tight budgetary and resource constraints.   Therefore, it is important that hospitals make efficient use of their resources through detailed planning and efficient systems. The use of operating theatres plays a significant role in this, as discussed in the literature. Our literature review (Chapter 2) demonstrates that there have been a number of studies across the world exploring various aspect of the planning of operating theatre schedules. However, there has been limited implementation of the results of these studies, possibly because individually they only consider some aspects of theatre scheduling.

This introduction provides a brief background to the challenges involved in theatre scheduling; sets out the objectives for the project, including those of the separate stages; describes the contribution made by the thesis overall; and sets out the structure of the remainder of the thesis.

1.1 Background

This section introduces the terminology used throughout this thesis, the literature review (Chapter 2) describes the various aspects of theatre scheduling and challenges surrounding them in much greater detail.

Hospital operating theatres or operating rooms are specially designed spaces within hospitals where surgery is conducted. The types of surgery are grouped based on the specialty areas of the surgeons within the hospitals and surgeons only conduct operations within their areas of expertise.

Patients are also separated into priority groups based on how urgently they require treatment based on their clinical need. The highest priority patients are emergency patients, also referred to as non-elective patients, who have the most serious conditions

and require treatment as soon as possible and whose operations are not planned in advance. All other patients are classified as elective patients and their surgery is planned in advance. Elective patients are divided again based on urgency into urgent and routine. Urgent patients need to be treated within the next few weeks (the exact details will vary for different specialties). Routine patients will benefit from surgery, but they can wait without impacting on their expected clinical outcomes.

A further grouping within elective patients are referred to as day cases or surgical outpatients, this group includes those having less invasive surgery who will not usually require a stay in hospital after surgery.

The ways in which decisions are required in theatre scheduling are generally divided into three levels: strategic, tactical and day-to-day. At the strategic level decisions include the likes of: the number and type of theatres a hospital should have; the opening times of operating theatres; how the theatre time available should be divided between specialties and individual surgeons. The tactical level decisions tend to focus on the master timetable of which theatre slot is taken by which surgeon over a planning period. The day-to-day decisions include both the advanced scheduling of patients when they are booked into particular theatre slots and the ordering of operations each day. These levels and the reasons for using them are covered in detail in Chapter 2.

Within the literature on theatre scheduling, waiting time is taken to mean the time patients wait between the decision that they require surgery and the operation taking place.

1.2 Objectives

The objective of this study is to combine the use of different operational research techniques to develop methodologies for use in hospitals to improve the scheduling of operating theatres. This work will focus on tactical and day-to-day scheduling of elective cases. Our analysis will take into account such considerations as the availability of operating theatres, equipment, staff, and beds, the stochastic nature of the time taken for individual operations and length of stay, along with other considerations as proves feasible. The potential impact of non-elective cases will be considered as appropriate.

Due to the complexity of the challenges involved in hospital theatre scheduling, including the variation between different levels of decision making, it is necessary to apply a variety of techniques in order to address the problem as a whole. This approach is supported by the literature review, which demonstrates that other studies have treated the different levels of theatre scheduling separately. The case for a multimethodological approach is further set out in Chapter 3, while making the case for including qualitative modelling with otherwise quantitative research.

The initial objective for the qualitative modelling is to identify the significant factors that affect theatre scheduling and the importance of improving theatre schedules within the running of the hospital as a whole.

At the tactical level the objective is to use of linear programming to generate cyclic master theatre timetables assigning surgeons to theatre slots, taking into account a variety of factors identified in the literature review and from the qualitative study with a local hospital.

At the day-to-day scheduling level the objective is to generate and objectively compare a variety of scheduling algorithms that could be used when booking individual patients, considering their implications for waiting times and efficient use of operating theatres.

1.3 Contribution

The most significant difference between this study and previous work in the area is the close link to how UK hospitals currently work, which has been maintained in order to ensure that the findings can be used by such hospitals. This approach is demonstrated in the emphasis placed on understanding the current challenges around operating theatre use faced by hospitals in general, before developing quantitative models.

At the master theatre timetable generation stage a wider range of factors are considered than in any previous study and flexibility is incorporated to allow hospital managers to explore how changes like providing evening theatre sessions could impact on other aspects of the hospital such as bed usage. The consideration of surgeon's availability enables managers to run the model both with and without the restrictions imposed by the preferences and other activities of surgeons, allowing assessment of the impact those

restrictions are having on the timetable. Thus, indicating the desirability of discussing changes to those restrictions and providing quantifiable evidence to determine the potential effects of such changes. The consideration of restrictions based on equipment availability allows limitations on the availability of resources including particular staff as well as items of equipment to be considered at the master theatre timetabling stage.

In the day-to-day scheduling stage surgeons ideas about how to improve the scheduling algorithm are combined with ideas from the machine scheduling literature to allow exploration of a diverse range of theatre scheduling algorithms, providing new insights into how different algorithms deal with the challenges of online scheduling with a variety of due dates.

1.4 Structure

Following this introduction, Chapter 2 gives a detailed review of the literature applying operational research techniques to different aspects of operation theatre scheduling. This includes consideration of the operational research methods applied to theatre scheduling; the levels of theatre scheduling; the aspects of theatre scheduling the papers consider; the way in which uncertainty is incorporated and the level of implementation of the literature in this area.

Chapter 3 discusses the reasons for including the soft operational research technique of cognitive mapping, before describing its implementation in this case and giving the implications for following chapters.

Chapter 4 considers the tactical problem of developing master theatre timetables giving details of the methods used and the reasoning behind their selection. Then the mathematical formulation used is given and justified. This is followed by discussion of the implementation of the method and discussion of the results and the conclusions drawn from them.

Chapters 5 and 6 explore the day-to-day scheduling problem of booking individual patients into theatre slots. This begins with discussion of the literature on applying scheduling techniques in other application areas. Details of the modelling and data used

are set out along with the results and discussion of the implications for selecting appropriate scheduling algorithms.

Finally Chapter 7 summarises the work undertaken throughout the thesis and explores its implications for theatre scheduling, as well as making suggestions for further work in the field.

Chapter 2: Literature Review

This chapter reviews the academic literature on the application of operational research techniques to hospital operating theatre scheduling, considering what has been done and how. This is done with a view to understanding the work that has already been undertaken in the field and to identify the gaps in that work for exploration later in this thesis.

The importance of scheduling the use of operating theatres has long been recognised and as far back as 1978, when Magerlein and Martin (1978) published a review paper on the subject. More recently Cardoen et al. (2010a) and Guerriero and Guido (2011) have published reviews of the literature in this area. The former focusses on allowing the reader to identify manuscripts relevant to their research interests as well as identifying areas to be addressed in future. The second review paper focuses on identifying the most mathematically interesting optimisation and simulation models. Our review does not specifically aim to assist the reader in identifying studies most relevant to each area of theatre scheduling, as these are identified in the chapters where they are relevant. Our overall aim is to identify gaps in the work undertaken by others to enable the selection of new areas for research later in the thesis. We look at the range of modelling techniques used in each area of theatre scheduling, with the aim of understanding how they have been used in the past to inform our decisions on which techniques to implement. Therefore, while our aims are similar, we have a different emphasis and there are differences between the groups of papers reviewed.

Like Cardoen et al. (2010a) and Guerriero and Guido (2011), we do not include clinical considerations and our scope is limited to those papers that address issues related to the planning and scheduling of operating theatres. Cardoen et al. (2010a) limit the papers considered to those published in or after 2000, while we predominantly consider such papers, some earlier publications are included, particularly if they address an area not covered in more recent contributions to the field. Cardoen et al. (2010a) also consider a number of conference papers; these have largely been omitted from this review in favour of journal articles. The exclusion of older conference papers is particularly valid as it is likely that the authors will have also written papers on similar work, so this avoids duplication.

The papers included in this review have been identified by a combination of searching OR journals and web facilities like Google Scholar using terms relating to theatre scheduling, identifying relevant references from papers, searching the journals that published these papers and expanding the list of search terms based on the titles of papers found via references. This search has identified a wide variety of papers including some published in clinical journals.

This review begins with a brief examination of the methods that are used in the papers considering various aspects of theatre planning and scheduling. Followed by consideration of the different types of surgery and discussion in detail of the extent to which they are covered in the literature, breaking them down into the different stages of planning where relevant. In later sections we discus where there are variations in the ways the different types of surgery and levels of planning are treated.

This is followed by discussion of the studies which look at particular details of the planning process and those considering issues related to theatre scheduling. Then we explore the different objectives or performance measures used and other aspects of surgery considered by the various papers. We proceed by considering of the extent to which the stochastic aspects of the problem are addressed, before exploring the extent to which the work in the literature has been implemented. The review concludes with discussion of the implications of the literature review for further work in the area in general and the remainder of this thesis in particular.

2.1 Methods

The most common operational research techniques to be applied to theatre planning and scheduling are variations on mathematical programming, simulation and heuristics. This section discusses each of these methods in turn, giving a brief description of the method and examples of how it is used in the literature. This is followed by examples of papers using multiple methods and less commonly used methods.

2.1.1 Mathematical Programming

Linear programming (LP) is a method of optimisation first developed by Dantzig in the 1940s, for "finding the maximum or minimum of a linear function subject to linear restrictions" (Maros and Mitra, 1996). It is a widely used operational research

technique, so much so that approximately 40% Winston's (1994) text book on 'Operations Research' is "devoted to linear programming and related optimisation techniques".

LP is used by Dexter et al. (2002a, 2002b), firstly to determine the worst case scenario for the cost of preoperative care by using LP "to determine by how much changing the mix of surgeons can increase total variable costs while maintaining the same total hours of OR [operating room] time for elective cases". The results demonstrate that changes to the operating schedule have the potential to have a significant adverse effect on costs, so this should be considered in allocating operating room time. In the second paper (Dexter et al., 2002b), LP is used again to consider the financial implications of the operating time allocated to surgeons. This time the conclusion is that up to 22% of surgeons could have their operating time reduced inappropriately because of sampling error generated from error in the data or processes used to analyse it, so confidence intervals should be used when making such decisions.

Integer Programming (IP) "is an LP in which some or all of the variables are required to be nonnegative integers" (Winston, 1994); usually if the variables are a mixture of those required to be integers and those which can take real number values then we have a Mixed Integer Programing (MIP) problem. Winston (1994) states that "many real life problems can be formulated as IP's" and it would seem that this includes a number of problems associated with theatre planning and scheduling. While operating room hours can be treated as positive real numbers, as in the studies discussed in the previous paragraph, factors such as the day on which an operation is to be performed and by which surgeon, require integer variables.

Gallivan and Utley (2005) use a MIP to assign the procedures taking place in a treatment centre to a cyclic timetable, with the goal of smoothing the use of hospital beds. They conclude that the use of such methods to achieve intelligent scheduling can "have a major impact on the variation in expected bed demand during the planning cycle". Blake et al. (2002), Vissers et al. (2005) and Santibanez et al. (2007) all use MIPs to assign blocks of time in theatres to surgeons. Demonstrating that such methods can "reduce resource requirements needed to care for patients after surgery, while maintaining throughput of patients" and confirming that "there is potential to

significantly reduce post-surgical bed requirements while maintaining throughput of patients" (Blake et al., 2002).

Zang et al. (2008) consider both the allocation of operating room time to specialities and the assigning of blocks of theatre time to groups of surgeons. They use MIP modelling for both of these problems. Overall MIP has been used by a number of authors, including a significant number using it to assign blocks of theatre time to surgeons.

There is a specific type of MIP problem known as job shop scheduling, which is one of the most studied scheduling problems due to its industrial applications (Pham and Klinkert, 2008). Classically it involves the scheduling of n jobs to be processed on m resources. Variations include jobs with due dates, available dates and penalties for jobs done outside of these dates. At first glance this seems like a methodology that it would be straight forward to adapt to operating theatres, by considering the operations to be jobs and the theatres resources. However, the problem in theatre scheduling is more complex as the surgeons and other equipment need to be scheduled as well as the operating room. Pham and Klinkert (2008) are rare among researchers in this field in that they use job shop scheduling for the scheduling of surgical cases. Even with adaptations to the method, their study only considers the theatre time available and not the surgeons' time or other resource availability. Su et al. (2011) treat operation room scheduling as a flexible job-shop scheduling problem, which is one of the hardest combinatorial optimisation problems. They assign surgeons to cases, but do not take account of surgeon availability or other resource constraints. Therefore, it would seem that this well-studied formulation is less generally suitable for the type of problems we will be looking at.

Part of the complexity of operating theatre scheduling is that the duration of the patients operation and their length of stay in hospital can vary widely between patients, even those undergoing the same procedure. Gerchak et al. (1996), Denton et al. (2007) and Adan et al. (2009) all make use of the branch of mathematical programming called stochastic programming, in order to take account of such variations. Adan et al. (2009) work with stochastic length of stay, which they find compares favourably with the results from a similar model with deterministic length of stay. Gerchak et al. (1996) and Denton et al. (2007) both consider the scheduling of individual patients taking account

of other stochastic aspects of the problem, including patient arrivals to the system and operation duration.

Goal programming is another variation on LP, which applies when a problem can be formulated as an LP with several (possibly conflicting goals) and the goals can be written as an additive linear objective function (Winston, 1994). Arenas et al. (2002) select this method because other health care studies show "that the flexibility of choosing the priorities of the model is a special advantage because it permits the decision maker to make different choices to find the one that represents the best option under each circumstance." Thus, goal programming applies to decisions with multiple objectives and allows the consideration of different balances of those objectives.

Arenas et al. (2002) proceed by applying goal programming to theatre scheduling with the overall aim of reducing the time patients spend on surgical waiting lists. Their first formulation ensures that no patient waits more than six months while maintaining the level of what they call "extraordinary interventions (which refer to operations in the hospital taking extra time, or referrals to other centres)". After attempting to find a solution with a maximum waiting time of 4 months, their second goal program incorporates operating theatre availability as a goal. Thus, they identify the interventions required to achieve a maximum waiting time of 4 months.

Goal programming is also used by Ozkarahan (2000) and Blake and Carter (2002). The objectives of the former are to be as close as possible to allocating the correct amount of time to each surgical specialty, to achieve the desired level of utilisation of the operating theatres; to assign cases to the best operating room for the case; to meet surgeons preferences with regard to the day of operation; and to ensure that the number of cases requiring intensive care is not more than the number of intensive care beds available. The objectives of the latter are more financial, so as not to breach a revenue cap; to recoup the costs; to provide surgeons with their preferred level of income; and to reduce deviations from the providers preferred volume of cases. Both of these studies have what might traditionally be considered constraints forming part of the objective function. Using goal programming in this way allows the relaxation of some of the constraints on a problem that might otherwise be infeasible, so that the solution closest to meeting the desired criteria can be obtained.

Column generation is a method that can be used to efficiently solve LPs that have large numbers of variables, as it increases the efficiency of the algorithm used to solve them. It has only been applied to theatre scheduling problems in the last few years. Belien and Demeulemeester (2008) and van Oostrum et al. (2008) both apply column generation to the problem of obtaining master surgical schedules, the former integrating the nurse scheduling into the process. Fei et al. (2008, 2009b) apply column generation to the scheduling of individual patients.

From the above, there are a range of papers that use LP and variations on it to solve different theatre scheduling problems; the aspects of theatre scheduling being considered and the aims of the study determine which is most appropriate for the specific problem considered by each paper.

2.1.2 Heuristics

Reeves and Beasley (1995) define a heuristic as "a technique which seeks good (i.e. near optimal) solutions at a reasonable computational cost without being able to guarantee either feasibility or optimality, or even in many cases how close to optimality a particular feasible solution is."

The development of heuristics has arisen largely from the issue that to solve some problems using linear programming techniques the "computational effort required was an exponential function of the problem size" (Reeves and Beasley, 1995). In other words, for some problems solving via linear programming might take an unreasonable amount of time so heuristic solutions are developed to solve specific problems in a reasonable amount of time.

Reeves and Beasley (1995) also argue that any optimisation routine is not actually finding an optimal solution to a model of the real-world, whereas heuristics can work with the real-world problem to find good solutions. For example in LP the objective and constraints have to be formulated as linear equations, which may well require simplification of the problem.

Some of the heuristics that have been applied to operation theatre scheduling are based on MIP or other mathematical programming methods. For example Cardoen et al. (2009b) explore a mixture of MIP solution techniques including heuristic methods in their consideration of the sequencing of patients receiving surgery each day. Fei et al. (2009b) apply column generation based heuristics to the problem of generating master surgical schedules; these are developed further by Liu et al. (2011a) who improve the efficiency of the algorithm enabling the solving of larger problem instances.

There are a range of heuristic algorithms which do not use aspects of mathematical programming methods, but use other methods to search for optimal solutions. Simple heuristics involve making small changes to the solution and moving to the new solution if it is better than the first until no similar changes generate an improvement on the current solution. Denton et al. (2007) develop a stochastic optimisation model for daily scheduling; having developed their model as a stochastic MIP it is particularly difficult to solve so they apply a selection of interchange heuristics, which they then use numerical analysis based on historical data to evaluate.

Metaheuristics have also been developed, often based on processes from other sciences to effectively search for an optimal solution. One such method is Simulated Annealing (SA). Briefly, this involves starting with an initial solution and then searching in the neighbourhood of that solution for a better solution: if a better solution is found the algorithm moves to that solution and repeats the process; if a better solution is not found, then the algorithm moves to the new solution with a given probability, where the probability of a move reduces as time passes (Dowsland, 1995). This method is applied by Sier et al. (1997) to the problem of scheduling individual patients.

Genetic algorithms (Goldberg, 1989) are another example of a metaheuristic, which works in a similar way to the selective breeding of plants and animals, in that parts of existing solutions are combined to form new solutions in the search for better overall solutions. This method is used by Roland et al. (2009) to simultaneously schedule patients operations and the staff who will be needed for those operations.

Tabu search is further example of a metaheuristic, which works by keeping a 'tabu' list of areas of the search area which have already been explored to encourage a wider

search for better solutions. This type of method is implemented by Dekhici and Belkadi (2010) in their exploration of theatre scheduling.

Riise and Burke (2011) use a metaheuristic algorithm based on iterated local search for the problem of scheduling admissions for patients.

Hans et al. (2008) explore a range of heuristics for assigning patients to days and theatres for surgery, including methods of obtaining the initial solution. They then compare simulated annealing with constructive approaches, including regret-based random sampling and random exchange; in the latter two randomly selected surgeries are exchanged if this yields an improved solution. They find that combining sampling methods and random exchange performs similarly to simulated annealing, but with much less computation time required.

Belien and Demeulemeester (2007a) explore the capacity of MIP based heuristics and metaheuristics to address the uncertainty involved in minimising the number of beds required by changing the master surgical schedule. For the MIP based approaches the non-linear objective function is replaced with a linear approximation, where the original objective can be kept for the heuristics. They obtained the best solutions with the SA metaheuristic, but with long processing times, compared to the MIP methods.

In conclusion, a variety of heuristics have been applied to different aspects of operation theatre scheduling. Heuristics are efficient at solving complex problems and can be used on formulations of the problem that are closer to the real world problem in some circumstances. However, there is no guarantee of an optimal solution and heuristics can still take a long time to obtain a good solution for some problems.

2.1.3 Simulation

Pidd (1998) summarises simulation as follows;

"The analyst builds a model of the system of interest, writes computer programs which embody the model and uses a computer to imitate the system's behaviour when subject to a variety of operating policies. Thus, the most desirable policy may be selected."

It is also one of the more popular Operational Research techniques according to Robinson (2005) and Hollocks (2005) states that; "Discrete-event simulation first emerged in the late 1950s and it has grown in popularity steadily to be now recognised as the most frequently used of the classical Operational Research techniques across a range of industries – manufacturing, travel, finance, health and beyond."

So simulation allows the user to test scenarios on a model of the system of interest, thus allowing the exploration of effects of different policies or scenarios. Simulation has been used across a wide range of health applications (Anderson et al., 2003), including a number of studies on theatre scheduling. These go back as far as Barnoon and Wolfe's (1968) study using Monte Carlo simulation to evaluate various operating room schedules.

More recently simulation has been used by Dexter and Traub (2002) to investigate the effects of variations in case durations on the performance of scheduling heuristics; by Dexter et al. (2000a) to investigate the effects of scheduling strategies on labour costs; by Dexter et al. (1999a, 1999b) to explore scheduling methods to maximise the use of operating room time; and by Sciomachen et al. (2005) to investigate the effects on wards of various scheduling policies.

These studies are all similar in that they are evaluating different policies for the scheduling of individual patients; the differences are mainly in the performance measures considered.

Some studies use simulation to explore operating room utilisation. For example Tyler et al. (2003) ran simulations to determine the highest utilisation that can be achieved, whilst staying within the goals of operations starting within 15 min of scheduled time and having cases end no more than 15 min past the end of the day. With these restrictions they found that the highest utilisation that can be achieved is between 85% and 90%. Dexter et al. (1999c) use simulation to evaluate strategies to decrease variability in utilisation and Mazzei (1999) also use simulation to explore methods of increasing utilisation.

A number of studies use simulation to evaluate scheduling policies for surgical departments or to investigate how such departments should be set up. For example Marcon et al. (2003a) use simulation to determine the number of post anaesthesia care beds required. Bowers and Mould (2005) use simulation to determine the effects on bed usage and theatre utilisation of treating ambulatory care (those who go home on the day of surgery) in separate theatres. Marcon et al. (2003b) also use simulation this time to evaluate scheduling policies based on the risk that the expected schedule is not realised.

Simulation is also used by Epstein and Dexter (2002) to consider the potential effects of errors in the data on previous theatre use for the allocation of operating room time to surgeons. They find that such effects are sufficiently small that it is not necessary to undertake data cleansing on this aspect of theatre data.

Everett (2002) describe a simulation model, which they claim can be used to allocate resources day-to-day, monitor how the current system is performing or for strategic decisions about the long term redeployment of resources.

Simulation is often used in studies that consider more than one decision level within operation theatre scheduling. Of the thirteen papers covering 'mixed decision level' identified by Guerriero and Guido (2011), nine use either simulation alone or simulation in combination with another method.

In some of the studies discussed above, simulation is used to compare heuristics for scheduling patients. In other studies, it is used to evaluate other methods or combined with other methods as part of a system addressing several aspects of theatre scheduling; these studies are discussed in the following two subsections. Simulation has not been used to suggest good/optimal schedules in the way LP methods and heuristics have. This is not generally a feature of simulation modelling, although work is on-going to implement optimisation through simulation (Fu, 2002).

2.1.4 Other Methods

There are also papers that use other methods or numerical analysis of theatre data or compare data from different hospitals. For example Longo and Masella (2002) discuss a benchmarking study comparing the organisation of operating theatres in eight Italian

hospitals, Basson and Butler (2006) use data-envelopment analysis and Sobolev et al. (2008) using the statecharts paradigm to capture the behavioural aspects of the delivery of surgical care. Dexter (2000) and Dexter et al. (2000b) are examples of studies using statistical analysis of data rather than particular OR techniques.

2.1.5 Combining Methods

As well as studies that focus on applying one method, there are examples where several types of method are used. For example, Lamiri et al. (2009) consider the problem of scheduling elective surgery, when operating space is shared by elective and non-elective cases using a range of methods. They begin by approximating the stochastic elements of this problem with deterministic values for use in Monte Carlo simulation and MIP to investigate convergence. They go on to consider several heuristic and meta-heuristic methods, including constructive heuristics, improvement heuristics and simulated annealing. They observe that "the quality of heuristics' solutions degrade as the uncertainty's variability increase". Thus, when using heuristics they found that the stochastic problem was easier than the deterministic one, which is counter intuitive as the stochastic problem would seem to be more complicated.

The previous section mentioned studies that combine simulation with other methods to evaluate those methods. A further example of this is Ogulata and Erol (2003), who describe a hierarchical system of goal programs, when scheduling operations, to take account of acceptability of slots to patients and assigned surgeons, and then use simulation to evaluate the methods on a case study.

Persson and Persson (2010) use optimisation within their simulation model to determine the schedule each day for use in the simulation model.

Testi et al. (2007) divide the problem of scheduling operating theatres into three phases, which they call 'session planning', 'master surgical schedule' and 'scheduling individual patients'. They solve MIP problems for the first two stages and use simulation to solve the third phase. Thus, they are dividing the problem into stages and applying appropriate methods to each.

Tanfani and Testi (2010) take a more integrated approach using optimisation for one stage of the problem and feeding the results into a simulation for the next stage. This is a more integrated approach but still requires different methods for the various stages of the problem considered.

2.1.6 Discussion of Methods

The methods most commonly applied to theatre scheduling problems can be divided into three groups, mathematical programming techniques, heuristics and simulation. Within these groups there are a number of methods used as demonstrated above.

As well as studies that focus on applying one method, there are examples where several methods have been applied to different aspects of the problem. Thus, a range of operational research techniques have been applied to different aspects of operating theatre scheduling, with some studies combining methods to solve different aspects of the problem or using simulation to compare methods. In late chapters we refer back to this discussion of the methods used in theatre scheduling, to inform the selection of appropriate methods for each of the aspects of theatre scheduling that we model.

2.2 Types of Operating Theatre Scheduling

In this section we breakdown the problem of hospital operation theatres scheduling into stages, based on the types of patient considered and the types of decision being made. As well as defining the stages, the studies addressing each stage of the problem are discussed.

The highest level of classification of hospital patients is whether they are inpatients or outpatients. The former are those patients who stay overnight 'in' the hospital and the latter those who visit for an appointment during the day and return home, spending the night 'out' of hospital. The majority of patients receiving surgery are inpatients, although there are a number who are able to return home on the day of their operation, these are referred to as daycases in the UK although elsewhere they may be referred to as receiving "outpatient and same-day surgery" (Dexter et al., 2002a). Outpatients receiving surgery can be treated as inpatients with a length of stay of one day as is done by Adan and Vissers (2002). Outpatients who do not receive surgery are not involved in theatre scheduling and are therefore not considered further in this review.

The more significant distinction in terms of operating theatre scheduling is whether patients are elective or non-elective. Cardoen et al. (2010a) define these thus: "The former class represents patients for whom the surgery can be well planned in advance, whereas the latter class groups patients for whom surgery is unexpected and hence needs to be performed urgently." The remainder of this section will discuss how the literature covers these two types of patients.

2.2.1 Non-Elective Surgery

Cardoen et al. (2010a) divide the studies they have analysed into those that consider electives only, those that consider non-electives only and those that consider both. Of the 124 papers they list as references only one (Bhattacharayya, 2006) is identified as restricting consideration to non-electives; while a further 20 references consider both electives and non-electives.

We have identified a further four papers where non-electives alone are considered. Jones (2002) explore the extent to which it is possible to forecast the demand for emergency care (including both patients requiring surgery and those who do not), demonstrating that the number of beds required by emergency patients is more predictable than the number or such patients arriving each day. They also find that while temperature and influenza rates affect emergency demand, including them in a model based on seasonal demand does not improve the predictions. Perhaps this is because temperature changes and influenza rates are subject to seasonal patterns and had therefore to some extent been included in the seasonal model. It is desirable to be able to predict demand for emergency beds because it affects the number of beds available for elective patients.

Bhattacharayya et al. (2006) explore the value of having a dedicated operating theatre specifically for orthopaedic non-electives. Bowers and Mould (2002) also consider orthopaedic cases, this time in terms of the increased efficiency that may be derived from concentration of care in one hospital. Dexter and Epstein (2006) consider how to plan the staffing requirements for care required on weekends and during holidays, as only non-elective care is taking place at these times and staff are needed to have the theatres open, this is effectively considering non-elective scheduling.

A further two papers consider the use of separate non-elective theatres; Wullink et al. (2007) demonstrate that closing dedicated emergency rooms and treating non-electives in the same theatres as electives can increase efficiency; while Bowers and Mould (2004) discuss the improvements in throughput that can be achieved by scheduling patients willing to accept a higher probability of cancellation to have elective care in non-elective theatres.

Like Cardoen et al. (2010a) we also found a handful of other studies which consider both elective and non-elective operations. The majority of these, for example Lamiri and Xie (2007), Bowers and Mould (2004) and Everett (2002), use simulation models to consider changes to or how to plan the use of the surgical unit. Others, such as Sier et al. (1997) and Van Houdenhoven et al. (2007b), consider non-electives, by allowing for an average daily number of such cases.

Zonderland et al. (2010) provide a detailed consideration of the balance of allowing time for more urgent cases to reduce overtime, with the possibility of underutilisation. They use queuing theory and Markov decision processes to do this.

Patrick and Puterman (2007) explore the potential of increasing utilization if some of the non-elective cases can be carried over to the next day, rather than assuming that they must all be treated on the day of arrival. They conclude that if only 10% of non-electives have the flexibility to be carried over to the next day this can have a significant reduction in the growth of waiting times for electives.

It is clear that only a small proportion of studies consider non-elective surgery, Cardoen et al. (2010a) consider this surprising given "that the larger degree of uncertainty is the main reason why operating room scheduling [literature] urges other scheduling methodologies than the machine scheduling procedures developed for industrial systems". It may be that forecasting the unplanned non-elective surgeries has appealed less than the more complex problem of planning for the predictable elective cases. Also more efficiency can be achieved by adjusting elective surgery.

2.2.2 Elective Surgery

As implied above there are considerably more studies that consider the various aspects of planning and scheduling elective surgery. Indeed as there are so many studies that consider elective surgery, we divide them into those dealing with strategic, tactical and day-to-day decisions. The separation of theatre management decisions into these three levels occurs in several papers for example Hans et al. (2007), Santibanez et al. (2007), Wachtel and Dexter (2008), Cardoen et al. (2009b, 2010a) and Guerriero and Guido (2011). This section takes each of these levels in turn, defining the problems at each level and giving examples of the work that has been undertaken to address them.

2.2.2.1 Strategic decisions

Wachtel and Dexter (2008), define strategic decisions as the highest level of decisions that may require years of planning before implementation, such as the building of a new unit. Hans et al. (2007) and Cardoen et al. (2010a) include a broader range of decisions in the strategic level. They consider strategic decisions to be those assigning theatre capacity to specialties / surgical services / individual surgeons and regarding long term staffing. We take the latter definition of strategic level for this review, as well as including concepts from the higher level of long term strategic planning.

There are a number of studies addressing different aspects of changing the operating room time allocated to surgeons involving Franklin Dexter (Dexter et al. 2000 a, b, 2001a, 2003, O'Neill and Dexter, 2007, Wachtel and Dexter, 2008). These studies explore the factors that should be considered in allocating theatre time to surgeons, the potential effects of poor data on this process and the financial impacts of changes at this level.

Operating room utilization has also been a focus for papers at this level. It is considered in some of the studies by Dexter mentioned above as well as by Tyler et al. (2003), Mazzei (1999), Strum et al. (1999), Van Houdenhoven et al. (2007a) and Olivares et al. (2008). Tyler et al. (2003) examine the highest level of utilization that can be achieved, without delay or running late. As mentioned in Section 2.1.3, they find that 85-90% utilization can be achieved, with increased variability in case duration reducing the achievable utilization.

There are studies at a strategic level that consider the potential effects of policy changes that hospitals could make with regard to the use of their operating theatres. For example studies by Utley et al. (2003) and Gallivan et al. (2002a, b) explore the impact on capacity of hospitals changing from a system where patients are put on a waiting list and given an operating date at short notice to booking patients when the decision to operate is made potentially months in advance. The latter resulted in some discussion in the letters pages of the British medical journal with responses from Rogers et al. (2002) and Castille et al. (2002).

Another policy change that could occur at the strategic level is the degree of pooling of operating rooms and/or surgical lists, compared with using dedicated operating theatres for different surgical groups and assigning patients to surgeons at an early stage. Batun et al. (2010) demonstrate that pooling of operating theatres and parallel surgery could lead to significant cost reductions. Vasilakis et al. (2007) show that pooling of outpatient clinic patients can reduce waiting times, which implies that pooling of surgical patients could also reduce waiting times. The extent to which pooling is possible depends on clinical considerations, as some cases will require a specific surgeon and operating theatre to provide appropriate care.

Other strategic level studies look at evaluating the efficiency of the entire operating suite across multiple inputs and outputs using data-envelopment analysis(Basson and Butler, 2006); planning to reduce waiting lists (Arenas et al., 2002 and Mullen, 2003); strategies for the use of recovery beds (Augusto et al., 2010); the number of recovery beds needed (Marcon et al., 2003a). Thus, a variety of issues affecting the longer term planning and operating unit wide policies are explored in the literature covering strategic level decisions relating to operating theatre planning.

Peltokorpi's (2011) findings question the effectiveness of changes at strategic level, concluding that "Based on the results, it could be argued that proper operative practices are more important than correct strategic decisions in terms of improving OR [operating room] performance. This also offers a good opportunity for operating units in which implementing new operative practices is typically easier than changing strategic orientation." This implies that focussing on tactical and day-to-day decisions will be more effective as well as more straightforward than exploring strategic issues.

2.2.2.2 Tactical decisions

Once the amount of operating theatre time per group has been defined at the strategic stage the next stage is to assign that time to surgeons. Hans et al. (2007) describe two different methods for doing this, the open and closed block methods. "In the open block method, OR time is assigned following the arrival of specialties or specialists [with cases to schedule]" with patients then scheduled on a first come first served basis. The closed block method is more commonly used (Guerriero and Guido, 2011), it involves assigning blocks of theatre time to specialties or surgeons. Indeed Cardoen et al. (2010a) only mention the closed block method at the tactical level, defining the process of assigning blocks of theatre time as the development of a master surgical schedule. Pham and Klinkert (2008) argue that "Non-block scheduling systems have turned out to have lower utilization and more case cancellations", and they also indicate that these systems are not favoured by surgeons as their surgery times are more spread out without a closed block system.

Given that the tactical stage is primarily concerned with assigning blocks to surgeons or groups of surgeons, there is not really a tactical planning stage for an open block booking system. Thus, the papers at this level are all considering closed block systems.

Adan et al. (2009), Belien and Demeulemeester (2007a), Blake et al. (2002), Hans et al. (2007), Van Houdenhoven et al. (2008), Zang et al. (2008), van Oostrum et al. (2008) and Santibanez et al. (2007) all consider the generation of master surgical timetables (or schedules) assigning time to surgeons or surgical groups in a cyclic timetable for a closed block system. The problem considered by these studies is essentially the same; how to assign the theatre time available to individual surgeons or surgical groups. All of these studies use variations on mathematical programming to solve the problem. The differences between them are in the detail of the methods used and the aspects of surgery included in the objectives and constraints, which are discussed in the relevant sections of this review.

2.2.2.3 Day-to-day decisions

This level comprises the advance scheduling of individual elective patients into the available blocks and also the sequencing of patients on the day of surgery. It is addressed by studies from as far back as the studies of Barnoon and Wolfe (1968) and Ernst et al.

(1977). Since then a significant number of papers have considered various aspects of this stage of the scheduling process. This subsection discuss studies covering just advance scheduling, then just sequencing, followed by those that cover both. There are variations in the methods used at this level so the studies are grouped by method within the areas of scheduling; the methods themselves are discussed above in Section 2.1.

Simulation is used in the majority of studies that assess strategies for the advanced scheduling of patients. Dexter et al. (2000a) use simulation to explore strategies for scheduling cases that cannot be completed in the block time assigned to surgeons into 'overflow' time. They find that the lowest staffing costs are achieved if surgeon and patient preferences are not taken into account; however, some flexibility can be given to surgeon and patient with only a small impact on costs. Sciomachen et al. (2005) use simulation to apply scheduling rules to the whole of the advanced scheduling process, scheduling first by longest waiting time, then longest processing time and finally by shortest processing time. They also test scenarios around the use a master surgical schedule and introducing a recovery room. They conclude that a flexible master surgical schedule can reduce the number of overruns significantly and that the introduction of a recovery room would reduce overruns and allow more operations to be performed. The results from comparing scheduling rules indicate that the best rule to use depends on the objectives in terms of reducing overruns or total overtime. Dexter and Traub (2002) also use simulation to compare scheduling rules based on scheduling patients to the earliest or latest start times available. Van Houdenhoven et al. (2007) apply the bin-packing problem algorithms Best Fit Decreasing heuristic and Regret-Based Random Sampling again testing the models with simulation. Dexter et al. (1999b) also apply bin packing algorithms and test them with simulation. However, their focus is slightly different as they are looking at adding additional cases once the schedule has been planned, rather than general routine scheduling.

All of the above studies test and compare different scheduling policies that could be applied by hospitals for the advanced scheduling of patients. This demonstrates how effective simulation can be for comparing methods, but also its weakness; simulation only compares the methods considered and does not suggest alternate methods or when exception should be made to the rules.

Gerchak et al. (1996) apply stochastic dynamic programming to the advance scheduling problem; they focus on the need to allow unscheduled time for a variable number of emergency (non-elective) cases. This type of mathematical programming technique finds the best solution to the problem of when to schedule groups of cases, given the information available.

Guinet and Chaabane (2003) also formulate the problem as a mathematical programming model, although they conclude that it is an NP hard problem and give heuristic methods that find solutions quickly. Guinet and Chaabane (2003) assume that the cases to be booked over the next two weeks are known when the problem is solved. This demonstrates the limitation of optimisation methods compared with heuristics, since they require sufficient cases to require scheduling to be able to consider the possible solutions; this is in contrast to heuristics, which allow small numbers of cases to be scheduled at once by following the heuristic rules.

Thus, for the advanced day-to-day scheduling problem a significant number of studies use simulation to find good heuristic rules for scheduling, while a few others have used optimisation techniques to find optimal solutions each time further cases are scheduled. There do not appear to be papers comparing the results of such techniques with the types of heuristic tested in the simulations, which would be of interest in terms of comparing the resulting theatre utilisation.

As the cases to be scheduled are known when sequencing patients assigned for surgery on a given day, this problem is more suited to mathematical programming techniques. Jebali et al. (2006) and Pham and Klinket (2008) both apply MIP to the problem, with the latter using the type of formulation used for job shop problems. Cardoen et al. (2009b) compare the results of using exact or heuristic methods to solve the problem formulated as a MIP. They find that such procedures could produce much better schedules than a human planner, with greater success at finding feasible solutions. Cardoen et al. (2009a) find that a dynamic programming approach is also effective on real-life problems.

Denton et al. (2007) use stochastic modelling for the daily sequencing problem, taking account of the variable nature of duration of operations to produce a sequencing rule that reduces both the turnover time between operations and overtime costs.

Heuristic techniques have also been applied to the problem. For example, Sier et al. (1997) used simulated annealing to schedule the patients for surgery the next day. So a variety of methods have been applied to the sequencing problem. However, Marcon and Dexter (2007) compare the sequencing of individual surgeons theatre slots with that recommended by software, focussing on the effects on the staffing required for post-surgery care and found that "the uncoordinated decision-making of multiple surgeons working in different ORs [operating rooms] can result in a sufficiently uniform rate of admission of patients into the PACU [post-surgery unit]". Thus, the value of computer aided sequencing depends on the objectives of the process.

There are also a small number of papers which consider both the advanced scheduling and sequencing problems. Ogulata and Erol (2003), Roland et al. (2009) and Fei et al. (2010) all do this by splitting the problem up and solving the stages separately. While Lamiri et al. (2009) consider both stages at once, the advantage of which is that better solutions to the second stage are not ruled out by keeping to the framework set out in the first stage; however, the problem is significantly more complex.

There are also studies looking at other aspects of day-to-day scheduling. Related to advance scheduling of cases, Dexter and Macario (2004) consider a system in which operating room time has been allocated to surgical services and when unfilled time should be released to services that have filled their allocated time. In the on the day level of scheduling, Dexter (2000) considers the circumstances under which moving the last case of the day in one theatre to a different theatre will reduce staffing costs. Also at the on the day level, Dexter et al. (2007) investigate the potential of real time status displays and computer recommendations to assist theatre staff in making effective decisions when changes to the schedule are required on the day of surgery.

Taking a wider view of day-to-day scheduling Dexter et al. (2000b) discuss the potential pit falls of a system for combing the scheduling of outpatient appointments and surgery; particularly if new patients requiring outpatient appointments are always assigned to the

surgeon with the shortest wait for surgery, which can result in cyclic variations in waiting times.

Related to the sequencing problem, Basson et al. (2006) explore methods of predicting which patients will fail to attend for surgery. They recommend that those with a high probability of non-attendance should be scheduled at the end of the day to minimise disruption.

Both Hans et al. (2007) and Cardoen et al. (2009b) mention a further level of online planning and monitoring, where the schedule is adapted to take account of changes such as the need to account for the inclusion non-elective cases. As non-electives are discussed above this in Section 2.2.1 this does not require further consideration here.

2.2.2.4 Research at more than one level

A small number of papers address two or more of the levels of elective scheduling, using the results from one level to schedule the next. For example, Hans et al. (2007) start with an IP model to allocate operating theatre capacity to surgical specialties, with the aim of ensuring that each specialty will have sufficient theatre time to complete their cases. Then they go on to develop a master surgical schedule to allocate blocks of time to each specialty so that each is assigned the correct amount of theatre time based on the first stage.

Testi et al. (2007) address all three levels; dividing theatre time between surgical groups, producing a master schedule and then using simulation to explore the day-to-day scheduling problem, as discussed in Section 2.1.5.

2.2.3 Discussion of Types of Operation

The studies exploring operation theatre scheduling do divide broadly into three levels, strategic, tactical and day-to-day. Even where studies cover more than one of these levels, they address each separately. Suggesting that this is a complex problem that requires division into smaller problems in order to make it manageable and that this division is logical based on the structure of the problem. Following this example from the literature we consider the tactical (Chapter 4) and day-to-day (Chapters 5 and 6)

aspects of theatre scheduling separately and apply different modelling techniques to each.

2.3 Studies Related to Theatre Scheduling

In addition to the papers mentioned above which directly address various aspects of theatre scheduling there are also a number of studies addressing topics related to it. This subsection will briefly discuss the types of topics that are related to theatre scheduling and give examples of papers for each.

In order to schedule patients to make full use of the theatre time available, it is necessary to know how long each patient's procedure will take. However, due to variability in the complexity of cases this varies even among patients undergoing the same procedure. Studies focussing on predicting operation duration include those by Strum et al. (2000) and Combes et al. (2008), both using data mining techniques, and Dexter et al. (2002d) who use pooling of cases to assist the prediction of cases with little historical data. Pandit and Carey (2006) compare the predictions made by surgeons, anaesthetists and nurses with those made by computer systems and find that "surgeons are not more optimistic than anaesthetists or nurses in their estimates of the time needed for operations. Furthermore, these subjective estimates are all in agreement with more objective data from the theatre computer." This suggests that it would be valid to use surgeon's predictions of operation times when scheduling surgery.

Also, considering surgery durations, Dexter and Macario (1999) consider the decrease in case durations that would be required to complete additional cases without going into overtime. They conclude that this is unlikely to create sufficient additional time, so cases are best added by optimising the overall operating schedule.

Following surgery the majority of patients require a hospital bed, so if no bed is available then their surgery cannot take place. This has already been mentioned in Section 2.2.1 in discussion of Jones (2002) and the number of beds required for emergency care, and is considered in the following section as one of the factors affecting theatre scheduling. Examples of studies taking broader view of bed capacities for the whole hospital are by Harper and Shahani (2002) and Gorunescu et al. (2002).

Vanberkel et al. (2011a) consider how to calculate the effects of the master surgical schedule on other areas of the hospital, so there is also research exploring how theatre scheduling fits into the rest of the hospital.

Just as predicting operation durations is necessary for planning theatre schedules, predicting length of stay in hospital is necessary for predicting bed usage. This is addressed on a general level by Lee et al. (2001), Atienza et al. (2008), Gustafson (1968), Cleary et al. (1991), Lee et al. (2005), Riihimaki et al. (2010) and Singh and Ladusingh (2010). Ridley et al. (1998) address the issue of predicting length of stay specifically for intensive care (an area where the number of beds can be particularly tight).

In addition to theatres and beds, appropriate staff members are required for operations to take place. Dexter and Epstein (2006) consider the staffing required for dealing with emergencies at weekends and on bank holidays as discussed in Section 2.2.1. Belien and Demeulemeester (2007b) address a more general staffing problem using column generation.

There are also a number of studies considering the potential impact on efficiency of changes to the way operating rooms are used (Bowers and Mould, 2005, Cantlay et al., 2006), anaesthesia is conducted (Bellamy, 2002, Broadway et al., 2001, Bromhead and Jones, 2002) and how patients recover after surgery (Augusto et al., 2010). Such changes may increase efficiency, but can only take place if the clinical judgement is that they will not cause a decrease in the quality of care provided to patients.

The discussion above covers just a sample of the studies covering topics related to operating theatre scheduling. These studies demonstrate that the scheduling of operating theatres is a complex process and links into the running of the hospital as a whole. There are also a large number of studies on clinical aspects of surgery, which are beyond the scope of this review.

2.4 Aspects of Surgery

Given that surgery is in itself a complex process it is unsurprising that the process of scheduling when it will take place is also complex with a number of factors to be

considered. As some aspects are treated as constraints by some studies and objectives by others, we have not separated out constraints and objectives, but consider each aspect in turn including discussion of how it is used by the various studies.

The most obvious factors and also the most significant for theatre scheduling are the theatre time available and the availability of surgeons; there are a significant number of other aspects of surgery to consider. Cardoen et al. (2010a) identify the following aspects of surgery used as performance measures (objectives);

- Waiting time, either of patients or surgeons
- Utilization, either undertime, overtime or general utilization of operating rooms or wards, including separate types of wards such as intensive care unit (ICU) and post anaesthesia care unit (PACU)
- Levelling of operating room time, ward usage, PACU, holding area or patient volume
- Makespan
- Numbers of patients deferred/refused
- Various financial considerations
- Preferences of surgeons or patients

They also identify the following aspects of surgery that are treated as hard constraints;

- Holding area
- Ward
- ICU
- PACU
- Equipment
- Surgical staff
- Budget
- Regular operating room time
- Operating room overtime/undertime
- Precedence constraints/time lags
- Release/due dates
- Demand constraints

Comparing these two lists reveals that there are a number of factors that can be treated either as objectives or constraints of a model to support operating theatre scheduling. The remainder of this section will consider each of these factors in turn discussing whether they are used as objectives or constraints of the model at each of the levels of scheduling and giving examples of the studies that use them. As many of the studies have been discussed above in previous sections, they are for the most part only mentioned briefly in this section. This is followed by detailed discussion of the objectives and constraints of a small number of studies, to illustrate how many factors are considered in each individual study.

2.4.1 Waiting Time

The length of time that they have to wait for surgery is particularly important to patients as no one wants to suffer with a condition while waiting for surgery to cure it. In recent years, UK governments have made reducing waiting times a major target for the NHS, which highlights the importance of this area. Papers which do not explicitly consider waiting lists are perhaps assuming that increasing efficiency with which theatres are used will decrease waiting lists.

There are studies that focus on the waiting lists, such as Mullen (2003), who consider how to prioritise waiting lists. Arenas et al. (2002) and Everett (2002) both consider waiting times explicitly in their studies.

At the strategic level, Bhattacharayya et al. (2006) consider the value of a dedicated theatre for orthopaedic trauma in terms of the reduction in waiting times of running such a system.

At a tactical level, Santibanez et al. (2007) consider waiting list management to be an important part of scenario planning and Testi et al. (2007) consider waiting lists in the tactical phase of their study.

At day-to-day scheduling level Dexter et al. (2000b) consider the potential impact on waiting times of different strategies. Minimising waiting time for patients forms part of the objective function for Guinet and Chaabane (2003), Ogulata and Erol (2003) and Denton et al. (2007).

In considering sequencing of patients for each day of surgery, Denton et al. (2007) consider the waiting time of each patient from the scheduled time of the operation, which is slightly different from the other papers incorporating waiting times where the wait from referral or decision to admit is counted.

Thus, a number of studies have waiting time as part of the objective or as a performance measure. At the day-to-day scheduling level, a number of studies have waiting time as a constraint on the date on which a patient's treatment can be scheduled. For example Pham and Klinkert (2008) have a constraint on the maximum waiting time and Fei et al. (2009a) have a deadline for each surgical case. Thus, consideration of waiting time is particularly relevant at the day-to-day scheduling level, when the waiting times of individual patients are considered, so it is an important factor in Chapters 5 and 6.

2.4.2 Numbers of Patients Deferred/Refused

The number of patients deferred or refused treatment under different scheduling policies is considered at the strategic level by Gallivan et al. (2002a) and at the day-to-day scheduling level by Arenas et al. (2002), Sciomachen et al. (2005) and Testi et al. (2007). Other studies assume that all the patients will be treated and the schedule will run as expected. As is discussed in Section 6.6 including cancellation and rebooking of patients would make it hard to judge how effective a booking algorithm is before the cancellations occur, so we do not include cancellations in our day-to- day scheduling model in Chapter 6.

2.4.3 Financial

The significant contribution of operating theatres to hospital budgets is highlighted by a number of papers (Lamiri et al., 2008, Hans et al., 2007 and van Oostrum, 2008), the majority of which mentioning that they consider how this costly resource can be best used. The studies with financial performance measures are generally by research groups working in the USA; Abouleish et al. (2003), Batun et al. (2010), Dexter et al. (2001a, 2002a, 2002b, 2002c, 2003, 2005), Dexter and Ledolter (2003), O'Neill and Dexter (2007) and Strum (1999) all include financial performance measures while considering strategic level aspects of operation theatre scheduling. Zang et al. (2008) from California and Blake and Carter (2002) from Canada have financial objectives in their

tactical level studies. Given the strong emphasis given to financial considerations in the USA, it is not surprising that researchers there have a stronger emphasis on finances than in other areas.

Our discussions with local hospital staff (see Chapters 3 and 4) concurred with the observation in the literature that finance is not a direct consideration in the European literature on hospital theatre scheduling and it is therefore not considered directly in our models.

2.4.4 Preferences of Surgeons or Patients

Ideally, the preferences of surgeons and patients would be considered when scheduling theatres; however, only a few studies take account of this.

At a tactical level Blake and Carter (2002) consider surgeons' ability to 'generate preferred level of income'.

Belien et al. (2009) consider surgeons' preferences for repetitive schedules and for different case mixes at a tactical level, thus allowing a limited influence to surgeon preferences. Also, at a tactical level, Testi et al. (2007) take account of preference of surgical groups for particular days during pre-processing. They define a surgeon preference objective, but this appears to relate to the case mix on the waiting list rather than actual preferences of surgeons. Ozkarahan (2000) considers surgeons' preferred operating rooms.

At the level of sequencing patients on the day of surgery, Cardoen et al. (2009b) consider surgeons preferences for the surgery they perform at the beginning of the session.

Perhaps the consideration of preferences is so limited because of the large number of other considerations involved in theatre scheduling and a drive towards efficiency. In Chapter 4 we explicitly consider the preferences of surgeons for theatre slots in our tactical level modelling.

2.4.5 Holding Area

Hospitals usually have a holding area where patients wait prior to surgery. It would seem that the capacity of this area is rarely considered limiting enough to include in a study of theatre scheduling. Only Pham and Klinkert (2008) consider it at a tactical level and Marcon and Dexter (2007) at day-to-day scheduling level.

Holding areas did not arise as a limitation in our discussions with hospital staff (see Chapter 3) so they are not included in our modelling.

2.4.6 Ward

As is discussed above in Sections 2.2.1 and 2.3, the majority of patients require a stay in a hospital ward following their operations, so the availability of hospital beds directly affects the ability to perform operations. This aspect is generally not considered at a strategic level, although Adan and Vissers (2002). Dexter et al. (2002a. 2002c, 2003) include overall annual ward time at a strategic level, rather than considering the beds used each day as at other levels of operation theatre scheduling.

At a tactical level, Vissers et al. (2005) consider bed availability. Levelling of demand for beds is the main objective for Gallivan and Utley (2005), Hans et al. (2007) and Belien and Demeulemeester (2007a) in their studies on the development of cyclic surgical schedules. Santibanez et al. (2007) also aim to minimise the number of beds required and van Oostrum et al. (2008) aim to level bed usage. These last two studies are also mentioned below as they also consider post anaesthesia care unit and intensive care unit usage, respectively. Also, at a tactical level, Blake and Carter (2002) treat the number of beds available as a constraint on the model.

At the day-to-day scheduling level, Bowers and Mould (2002) consider use of ward capacity as a performance measure in their simulation study, while Testi et al. (2007) and Guinet and Chaabane (2003) consider ward capacity as a constraint. Bekker and Koeleman (2011) provide an in-depth consideration of the potential for theatre scheduling to create variability in bed demand and develop models for which the objective is to reduce such variability.

Given that the way individual patients are scheduled on a day-to-day basis has greater potential to make best use of ward capacity than the scheduling of blocks in the master schedule, it seems strange that ward capacity is considered more at the tactical than day-to-day level. This implies that space available on wards should be considered in both tactical and day-to-day scheduling.

2.4.7 Intensive Care Unit

An intensive care unit (ICU) is effectively a small ward where small numbers of patients with high levels of need are cared for. As for wards, if an intensive care bed is not available for a patient expected to require one after surgery, then the surgery will be cancelled. Given the small number of beds in most ICU units this can be an important consideration, particularly for surgical units conducting more complex cases likely to require ICU beds following surgery.

Adan and Vissers (2002) are rare among strategic level studies in considering this aspect of theatre scheduling. As for wards, Dexter et al. (2002a, 2002c, 2003) include annual ICU usage rather than the number of ICU beds used each day.

Van Oostrum et al. (2007), Hans et al. (2007) and Van Houdenhoven et al. (2008) consider ICU capacity as a major objective in their tactical level studies. Also, at a tactical level, Vissers et al. (2005) consider ICU capacity, as their study focuses on cardiothoracic surgery, among the more complex types of surgery, so the ICU is of particular importance.

At day-to-day scheduling level Jebali et al. (2006) consider the number of ICU beds available when scheduling patients.

As the ICU is effectively a specialist ward, it is possible that it is included implicitly in other studies that consider bed usage.

2.4.8 Post Anaesthesia Care Unit

It is common practice for patients to move to a post anaesthesia care unit (PACU) after surgery before being moved to beds in wards, so the number of spaces in the PACU has a similar effect on the ability to perform surgery as the number of ward beds. At the

strategic level, Marcon et al. (2003a) use simulation to determine the number of beds required in such a unit. Marcon and Dexter (2007) explore the impact of surgeons' scheduling practice on PACU usage.

At the tactical level, Santibanez et al. (2007) consider the impact of the schedule on recovery beds, including the staff and equipment required for recovery.

At the day-to-day level, Marcon and Dexter (2007) consider the impact of surgeons' sequencing of cases on the PACU. Also, at the day-to-day level, Fei et al. (2008, 2010) consider the availability of recovery beds when scheduling patients.

At the level of sequencing patients for surgery each day, Cardoen et al. (2009b) consider the affect the sequence will have on the closing time of the PACU at the end of the day, to avoid staff overtime and other costs associated with the unit remaining open longer than planned. Cardoen et al. (2009a, 2009b) both consider levelling of the number of bed spaces required in the PACU over the course of each day as an objective.

2.4.9 Equipment

It is possible particularly for large and/or expensive equipment, that only a very limited number of sets are available for use at the hospital. It is also possible that due to the need to sterilise equipment between patients such equipment might only be available for use once each day. This restriction does not appear to be taken into account by studies at strategic or tactical level, which makes sense as it is only at the day-to-day planning stage that individual patients and their needs are considered. Guinet and Chaabane (2003) and Jebali et al. (2006) consider whether required equipment will be available in each operating theatre.

At the day-to-day scheduling level, the required equipment is considered by Sier et al. (1997). At the level of sequencing patients for surgery each day, Cardoen et al. (2009a, 2009b) both consider the available equipment as constraints on the problem.

This limited consideration of the availability of equipment in the literature, is a gap in the research that has been done to date and is considered further in Chapter 4.

2.4.10 Surgical Staff

The most important staff for operation theatre scheduling are the surgeons. Papers vary as to whether they schedule individual surgeons or surgical groups, depending on the level of detail required. Virtually all studies addressing aspects of operation theatre scheduling consider surgeons in some way.

A number of other types of staff are involved in surgery, from anaesthetists and other theatre staff to the porters who bring patients from the wards to the theatres. At the strategic level, McIntosh et al. (2006) and Pandit et al. (2007) explore how best to plan the staffing of theatres with the objective of reducing labour costs. Dexter et al. (2003) and Dexter and Epstein (2006) also consider reducing staffing costs. McIntosh et al. (2006) explore how far in advance staffing should be planned to increase productivity. Marcon et al. (2003a) investigate the number of porters available, because they are focused on the number of PACU beds required and the speed at which porters remove patients to other parts of the hospital can affect this.

At a tactical level, Belien and Demeulemeester (2008) consider the objective of integrating nurse and surgical scheduling. Also, Belien et al. (2006) demonstrate how the demand for nursing staff can be visualised based on the master theatre schedule, to aid planning.

At the day-to-day scheduling level, Marcon and Dexter (2007) and Roland et al. (2009) consider scheduling patients whilst considering staffing requirements/constraints. Also, Dexter et al. (2000a) and Dexter (2000) consider the effects of different strategies on labour costs.

This demonstrates that while surgeons time is clearly an important factor in theatre scheduling, the scheduling of other staff groups can also be significant.

2.4.11 Operating Room Time

Considering the amount of operating room time available is central to any study concerning operation theatre scheduling. It is considered by all studies on the topic either via constraints on the time used or in the form a performance measure around theatre utilisation.

At the strategic level, van Houdenhoven et al. (2007a) and Tyler et al. (2003) consider as a performance measure, the highest level of operating theatre utilization that can be achieved with limits on the acceptability of sessions running overtime. 100% utilization cannot reached due to the variability in operating times, although if a higher risk of running overtime is accepted and the case mix is uncomplicated, then utilisation close to this figure can be achieved (van Houdenhoven et al., 2007a). As discussed in Section 2.1.3 more realistic goals for utilisation are between 85-90% depending on the case mix (Tyler et al., 2003).

Also, at the strategic level, Mazzei (1999), Dexter et al. (1999a) and Dexter and Macario (2002) consider how to allocate theatre time to surgeons or surgical groups with the aim of maximizing theatre utilization.

As papers at the tactical level are generally considering the allocation of theatre time to surgeons or surgical groups, as such it is necessary that they consider the theatre time available as a constraint (e.g. Blake and Carter, 2002, Gallivan and Utley, 2005, van Oostrum et al., 2008 and Fei et al., 2009).

At the day-to-day scheduling level, operating theatre time is either considered as a constraint on the operations that can be scheduled each day or as a performance measure. For example, Guinet and Chaabane (2003) have constraints relating to the regular and overtime theatre hours, for days where overtime is allowed. Gerchak et al. (1996) consider the trade-off between maximising utilization of theatres and reducing overtime and delays. Basson et al. (2006), Dexter et al. (1999b) and Van Houdenhoven (2007a) all aim to maximise utilisation. Murat and Nepal (2010) consider the effect of different policies for the order of operations each day on the overtime performance of operating rooms.

This shows that there are different ways of considering operating room time depending on the level of scheduling being undertaken. This is reflected in the difference between treating available time as a constraint in Chapter 4 when we are scheduling at tactical level and as both a constraints and performance measures for day-to-day scheduling in Chapters 5 and 6.

2.4.12 Makespan

Cardoen et al. (2010a) define makespan as 'the time between the entrance of the first patient and the finish of the last'. So it is in effect a measure of how operating room time is used. It is considered at the day-to-day scheduling level by Marcon and Dexter (2007) and Pham and Klinkert (2008). It is effectively a variation on theatre utilisation so is not considered separately in the rest of this study.

2.4.13 Precedence Constraints

At a detailed level of scheduling, there are likely to be differences in patients levels of clinical need, with some studies dealing with this implicitly by using different due dates for patients (see Section 2.4.14), while others address it directly. For example, Ogulata and Erol (2003) have three different priority levels for patients, allowing them to give precedence to some patients. Min and Yin (2010) prioritise patients based on a weighted sum of numerical values representing clinical criteria, such as disease progression, pain or dysfunction, they then schedule based on this priority order. They assume that the priority of a patient does not change until they are removed from the waiting list, and include a cost penalty relating to the waiting time in the objective function, thus also providing some degree of priority to those who have waited longest.

For the problem of sequencing patients on the day of surgery there are further priorities to consider. Cardoen et al. (2009a, 2009b) give precedence to scheduling young children at the start of the day, followed by prioritised patients "for instance, patients who already had a cancelled surgery once or surgeries that the surgeon preferably performs at the beginning of the slot". Precedence is also given to children, in the order youngest first by Sier et al. (1997).

The question of how to prioritise patients has its own field of literature, which is summarised by MacCormick et al.'s (2003) paper. They conclude that the debate over the ethical basis for prioritisation is on-going and that the impacts of assumptions made during prioritisation have yet to be explored. The question of how prioritisation is achieved is a clinical decision so we will not consider it in further detail here. However, in Chapters 5 and 6 we do take account of the prioritisation created by surgeons specifying how urgently each patient requires treatment.

2.4.14 Release/Due Dates

Patients become available for operation over time as the decisions to operate are made; rather than all being available at the start of a planning period. They may also have due dates for their operations based on their clinical need or targets for limiting waiting times. Where they are considered release dates and due dates are treated as constraints on the model.

At the strategic level Marcon et al. (2003a, 2003b) use release dates for simulation of patients operations.

At tactical and day-to-day scheduling levels, Jebali et al. (2006), Fei et al. (2008, 2009a, 2009b, 2010) all use due dates to ensure that patients operations are scheduled on time. Guinet and Chaabane (2003) and Lamiri et al. (2008, 2009) use both release dates and due dates for the scheduling of individual patients.

Cardoen et al. (2009a, b) give consideration the distance patients are travelling for their surgery, and penalise a sequence in which these patients have operations starting before 11am. This is in effect a release time treated as a soft constraint by making it part of the objective.

Due dates in particular are in effect a way of restricting the waiting time of patients, so these examples also relate to Section 2.4.1.

As due dates are particularly important when scheduling individual patients they are an important factor in the day-to-day scheduling considered in Chapters 5 and 6.

2.4.15 Demand for Surgery/Theatre Time

The demand for surgery is an important consideration in advanced planning of theatre scheduling. At the strategic level Adan and Vissers (2002) consider demand by having 'target patient through-put', and Dexter et al. (2002b, 2005) consider surgeons' demand for theatre time. Generally at the strategic level, consideration is given to the demand for different types of surgery when allocating theatre time to surgeons/surgical groups and Testi et al. (2007) have this as an objective for the strategic phase.

At the tactical level, if is necessary to know how much demand for theatre time each surgeon/surgical group has in order to assign theatre slots to them. Therefore, we expect to see demand constraints at this level, as is the case in Blake and Donald (2002), Santibanez et al. (2007), Belien and Demeulemeester (2007a), Zang et al. (2008), Belien et al. (2010) and other tactical level studies.

At the day-to-day level of theatre scheduling, demand is considered implicitly as the individual patients requiring surgery are considered.

We follow these trends in the consideration of demand by using known demand for theatre slots per surgeon in Chapter 4 and dealing with demand implicitly as individuals arrive for booking in Chapters 5 and 6.

2.4.16 Examples of Bringing Factors Together

Thus far we have only discussed separately the types of objectives/performance measures and given examples of the constraints. The studies discussed in this section are examples of the way in which objectives and constraints combine in different papers.

Marcon et al. (2003a) is an example of the use of simulation at strategic level, the objective of the study is to determine the minimum number of PACU beds required. To do this they take account of the efficient use of theatre time, the staff available for the theatre including surgeons an anaesthesiologists, the number of porters available and patients' length of stay in the PACU unit. Thus, they combine the more significant factors for theatre scheduling (surgeons and theatre time) with the porters available and the use of the PACU.

Utley et al. (2003) have the objective of determining the number of beds required if patients are booked further (potentially months) in advance. They also take into account other ward admissions including emergencies and the demand for beds. In focussing on bed usage, they do not consider other factors including theatre time or staff availability.

Santibanez et al. (2007) explore the tactical problem of allocating blocks to surgeons, with the objective of minimise the peaks in use of post-surgery resources, such as beds,

whilst maximising throughput. They include as constraints, the theatre time available, the time required by surgical groups, the throughput of patients over the planning horizon and the desire to have a schedule that repeats weekly. Since they are working at the level of surgical groups rather than individual surgeons they do not consider staff such as individual surgeons or equipment constraints.

Thus, studies bring together some of the aspects of theatre scheduling, depending on the focus of their studies, but as these examples show not all of the aspects that could be considered relevant are considered by any individual study.

2.4.17 Discussion of Aspects of Surgery

This section demonstrates that a wide range of aspects of operation theatre scheduling are covered in the operational research literature on the area. There is variation in the aspects considered both between and within the different levels of scheduling. Also there is variation in how the aspects are considered; as performance measures for comparing different scheduling methods, objectives to be achieved as much as possible or constraints within which scheduling must take place``.

The aspects considered by any particular study depend on the focus and objectives of the authors, as it is logical to only include those aspects which are relevant to the problem considered to avoid over complications.

Looking at the broader picture, it is interesting to note that the effect on wards is considered more frequently at the tactical level than the day-to-day level. One might expect that in the scheduling of individual patients, there would be further scope to consider the beds that will be required to reduce the risk of cancellation due to lack of beds.

The number of factors considered in scheduling operating theatres demonstrates the complexity of working in this area as well as the extent of the potential for affecting the rest of the hospital.

The selection of objectives is important and different selections can give rise to distinct solutions as found by Riise and Burke (2011).

The aspects of theatre scheduling covered by the literature has directly affected the consideration of which aspects to consider in our work as is discussed in more detail in Section 4.1.3.

2.5 Uncertainty

Any attempt to improve the scheduling of hospital operating theatres is further complicated by uncertainty over the patients who will arrive over the planning horizon, the duration of operations and the lengths of stay of patients following operations. The stochastic nature of these factors means that a scheduling technique which performs well on average may perform badly in practice.

There are some studies that focus specifically on tactics to address these aspects of the uncertainty surrounding theatre scheduling. For example, Dexter et al. (2001b) consider how much delay to schedule between patients to improve the likelihood of each surgical case starting on time given the uncertainty over case durations. In their exploration of the potential effects of advanced booking, Gallivan et al. (2002 a, b) conclude that the variability of emergency admissions, length of stay and cancellations means that any attempt to make a firm commitment by booking patients months in advance may well reduce capacity.

Simulation models are particularly effective at considering this variation as such models can be run many times to explore the best and worst case scenarios as well as to calculate the expected average performance. Thus, the studies discussed in Section 2.1.3 incorporate the uncertainty surrounding operation theatre scheduling, via the method used. This aspect of simulation is demonstrated well by Marcon et al. (2003) who use tables to illustrate the maximum, average and minimum numbers of beds and porters needed over a number of runs of each of different simulation scenarios.

Simulation is used by Batun et al. (2010) to demonstrate the benefits of pooling operations between surgeons (rather than assigning surgeons before booking) can reduce the costs caused by uncertainty within the system.

Due to the unpredictable nature of demand for non-elective care, uncertainty is particularly important for any study considering it. As discussed in Section 2.2.1,

studies either forecast such demand for care by looking at the factors that cause variability or incorporate it via simulation.

Deterministic methods such as IP and MIP, which Section 2.1.1 discusses, are used by a significant number of studies on operation theatre scheduling, but do not take account of any of the stochastic aspects of theatre scheduling. Zhang et al. (2008a) recognise this, so having obtained a weekly operating room allocation from their MIP they run a simulation model to test the performance of the resulting allocation taking account of the stochastic nature of the "surgery time, demand, arrival time and no-show rate". Similarly, Lamiri et al. (2009) use simulation to test a variety of methods including heuristic and meta-heuristic methods.

Other studies using a range of LP related and heuristic methods do not go as far as testing with simulation, but evaluate the results by using instances of the problem. Roland et al. (2010) "perform 10 experiments in order to estimate the performance of the genetic algorithm" and Jebali et al. (2006) use 25 problem instances to evaluate their models. Fei et al. (2009b), Cardoen et al. (2009a) and van Oostrum et al. (2008) all use test data generated from real life data from specific hospitals to test how their models would perform under uncertainty.

There are studies which explicitly consider the stochastic aspects of theatre scheduling while using methods other than simulation. Gallivan and Utley (2005) consider booking of individual cases into a cyclic timetable; working at this level allows the clever construction of a linear program that uses the probability of different lengths of stay.

Also considering the variability of length of stay, Belien et al. (2009) incorporate the variance into their MIP. Belien and Demeulemeester (2007a) give a formulation for taking account of such variability, but as it is non-linear they consider various methods of approximation and using heuristics to solve the problem.

Stochastic programming is a variation on LP which takes account of the stochastic elements of the problem under consideration. The first application of this to operation theatre scheduling appears to be by Gerchak et al. in 1996, where only the stochastic nature of the number of requests for surgery each day is considered. More recently

Denton et al. (2007) apply stochastic programming to consider uncertainty around operation durations; Adan et al. (2009) apply it to the tactical scheduling problem considering uncertainty around length of stay; and Min and Yin (2010) apply it to the day-to-day scheduling problem whilst considering patient priority.

So aside from the studies using simulation there are relatively few papers that take account of the stochastic elements of operating theatre scheduling in their main methodology and those that do focus on the stochastic nature of just one factor. There are a number of other studies using experimentation to test how their methods perform under uncertainty. This shows that uncertainty is an important aspect of theatre scheduling; in later chapters it is considered as much as possible, and where it is not considered the reasons are clearly set out.

2.6 Implementation of Research

Given the potential for improvements in operation theatre scheduling demonstrated by the models discussed above their findings have the potential to make significant improvements in hospitals, one might expect to find them in use across a range of hospitals.

In their review of surgical scheduling, Magerlein and Martin (1978) note that in general the schemes for supporting scheduling that they discuss have not been implemented; suggesting that early studies in the area were not generally implemented in hospitals. It appears that a lack of implementation of modelling has been an issue in healthcare as whole not just in relation to operating theatres as Lagergren (1998) cites a 1981 survey which found that only 16 out of 200 health related project recommendations had been acted upon. Eldabi et al. (2007) review the use of simulation in healthcare, they observe that the impact of simulation is weak and Jun et al. (1999) expect that "future modellers will continue to face difficulties implementing their results". Taking a further generalisation Wiers (1997) discusses concern that despite the large volume of studies considering various aspects of scheduling with operational research and artificial intelligence techniques implementation of these techniques in scheduling is rare.

Lagergren (1998) suggest that the situation in health care in general is improving as about a third of the studies they reviewed had been implemented to some extent. So what is the situation with regard to operation theatre scheduling?

Gemmel and Dierdonck (1999) conducted a thorough literature review and a telephone survey of Belgian hospitals to assess the extent to which the methods used for the scheduling of admissions for surgery fits with the theoretical models available on the topic. The conclude that "most hospitals have not worked out an admission scheduling policy indicating which resources are critical in the scheduling process and how information on the availability of these resources is captured". Therefore, hospitals are often not aware of the issues affecting their scheduling process, let alone employing operational research methods to overcome them.

Also, Cardoen et al. (2010b) discuss the results of a survey of hospitals in Flanders, in which they asked about the systems used for scheduling operating theatres. They found that "despite the proliferation of computerised planning and scheduling procedures proposed by the scientific community, the implementation of such systems still seems to be low". This suggests that there has been a consistent lack of implementation of research in this area.

This lack of evidence of implementation is also apparent in the studies we have discussed. While a significant number of the studies use data from hospitals, and as discussed above in Section 2.5 some test their methods with data from hospitals, only a handful of papers discuss the implementation of their methods.

The earliest paper that we have found mentioning implementation is from 1977 by Ernst et al., who state that their method had at that time been in use for more than 2 years. More recently Kusters and Groot (1996) state that in two out of the three test environments their system was accepted and used as the basis for admissions decisions. Blake and Donald's (2002) integer programming model for the tactical level scheduling problem had been in use for several years when they wrote the paper. Testi et al. (2007) state that "the approach has been accepted and implemented by the surgical units of the department involved in our study". There are also some papers where implementation is

implied, for example, Arenas et al. (2002) comment that the decision makers were satisfied with the methodology, but do not explicitly say that it was used.

Even in the papers that do indicate that the models have been implemented it is only in the hospitals/units that the researchers worked with on the study, although there must be other hospitals with similar problems. Also, the methods that have been implemented are from different levels of theatre scheduling, suggesting that there is not a particular problem with implementation at any of the levels.

It is possible that for some of the papers which do not mention implementation, the models have been implemented since publication of the paper. However, there are papers that give the reasons for non-implementation. For example, Sier et al. (1997) discuss the need to develop a 'front end' for the system and to integrate it with hospital systems before implementation can take place. Also, if implementation was taking place post publication, then we would expect the surveys by Gemmel and Dierdonck (1999) and Cardoen et al. (2010b) to have found evidence of this.

There are a few more recent papers which suggest that implementation has occurred. Vanberkel et al. (2011b) discuss some of the ways in which their model aided the development of a master surgical schedule, implying that it was used as part of the process. Isken et al. (2011) mention use of their model in several hospitals and have made it available as open source software.

Cardoen et al. (2010a) conclude that it is "somehow contradictory to see that in a domain as practical as operating room planning and scheduling, so little seems to be effectively applied". The remainder of this section explores possible reasons for the lack of implementation and strategies to overcome it in future.

Discussing the lack of implementation of scheduling techniques in general, Wiers (1997) comments that most papers "focus on the system's architecture and implementation issues are apparently regarded as trivial. The success of scheduling techniques in practice can only improve when researchers are aware of the implementation pitfalls through learning from each other's experiences". This suggests that greater sharing of

ideas and experiences is required to improve the chances of implementation, which does not seem to be occurring in the literature on scheduling hospital operating theatres.

In discussing possible reasons for the lack of implementation, Eldabi et al. (2007) comment that "while it is possible to assess the simulation benefits on defence and manufacturing systems, such benefits seem less tangible when it comes to healthcare simulation". So in comparison with other application areas, it is harder to demonstrate the benefits of modelling in healthcare.

A common theme in the papers discussing the lack of implementation is the lack of awareness of operational research techniques among hospital decision makers, with Cardoen et al. (2010 a, b) suggesting more training for managers in the area, for example thorough games to illustrate the effects that applying the principles of operational research can have on scheduling of patients.

Harper and Pitt (2004) propose a 'project life cycle for successful implementation' of healthcare modelling, in which several stages involve working with hospital staff to: understand the problem form their point of view; build credibility of the modellers; and model and acknowledge the politics within the organisation. They proceed to list examples of projects where this life cycle has been successfully implemented in healthcare. This demonstrates the importance of working with hospital staff alongside undertaking of quantitative analysis.

There is also a need to allow flexibility in the model, both to take account of changes within a hospital and differences between hospitals, if the model is to be more widely applicable (Gemmel and Dierdonck, 1999).

Due to the lack of implementation of research into hospital operating theatre scheduling, this section includes consideration of implementation within the wider areas of healthcare and scheduling. On these broader levels, there is also evidence of a lack of implementation. However, there are indications that implementation is improving and this section identifies the need to work closely with hospital staff and incorporate flexibility into modelling to create a more implementable model. Jun et al. (1999)

mention that even if implementation does not occur, modelling can have benefits such as developing a greater understanding of the system and identifying unexpected problems.

The identification of this gap in the literature directly influenced our decision to work closely with local hospitals to understand their needs, particularly though the cognitive mapping exercise described in Chapter 3.

2.7 Discussion

This section both summarises this review of the operational research literature on operating theatre scheduling in hospitals and discusses the implications for future work.

At the start of this review (Section 2.1), we identify simulation, linear programming and heuristics as the three main classes of methods applied to theatre scheduling problems. Simulation models are effective for evaluating policies or testing scenarios, particularly for comparing the effectiveness of different scheduling policies. However, they do not directly suggest alternative policies and they do not necessarily give an indication of how much potential there is to improve on current methods. Linear programming methods (including variations on LP) do give optimal solutions for the criteria specified, if it is possible to specify the problem with linear equations and inequalities. However, for some problems, the computation time may be excessive. Heuristics have been shown to be effective for finding faster solutions and it is not necessary to specify the problem in linear form, but the solution they find is likely to be suboptimal and it may not be possible to determine how far it is from optimal. Therefore, the decision over which of these methods to use depends on the type of problem being considered and how it is possible to formulate the problem, as well as the speed with which solutions are required.

In Section 2.2 we discuss how the problems relating to operation theatre scheduling can be broken down into those considering elective or non-elective surgery, with the former category further dividing into strategic, tactical and day-to-day scheduling problems. The majority of the papers discussed consider problem(s) at one or another of these levels and where more than one level is considered the different levels are explored separately. This suggests that it would be sensible for future research to make use of

these levels and if more than one is to be considered that this should be done sequentially. This is done in Chapters 4, 5 and 6.

There are a range of factors with the potential to influence theatre schedules, as the number of factors considered as objectives or constraints in the literature demonstrates (Section 2.4). This has formed part of the consideration of the factors to consider in our models, particularly in Section 4.1.3.

Section 2.6 shows that there is a general lack of implementation of the research done in this area within hospitals and suggests working closely with them to understand the challenges they face. Suggesting that the decisions over which factors to include in future work, should come from such collaboration. Chapter 3, Sections 4.1.2 and 6.1 give further details of the collaboration with local partner hospitals include in our modelling.

The stochastic aspects of theatre scheduling (Section 2.5) should also be considered where appropriate to ensure that the modelling reflects the real situation sufficiently to be useful in practice. These are considered where possible in later chapters and where they are not included the reasons are explicitly discussed.

Overall we perceive that the strategic level of operating theatre scheduling is subject to greater political considerations, as there are more subjective factors involved. Also, changes at this level occur less frequently, so it would be harder to achieve implementation at this level. Therefore, given our goal of implementation, in Chapter 4 we focus on addressing the tactical scheduling problem and in Chapter 5 we focus on the day-to-day scheduling problem. Discussions of the selection of methods for each of these levels are given in the appropriate chapters.

Chapter 3: Qualitative Modelling – Understanding the Problem

The literature review gives us a clear understanding of previous academic work on operating theatre scheduling, with a lack of evidence of implementation of that work in hospitals. In order to increase the potential of our work to be implemented in hospitals it is necessary to gain an understanding of the challenges hospitals face surrounding theatre scheduling; the intention being both to gain some insight into the lack of implementation and the needs of hospitals.

In recent years there has been increasing use of multimethodology in studies across a range of applications, from taxation systems (Brown et al., 2006) to assessing fitness to drive (Hindle and Franco, 2009) and combining a variety of methods such as mixtures of qualitative methods for example, problem structuring with ethnography (Horlick-Jones and Rosenhead, 2006); or mixing qualitative and quantitative methods (Kotiadis and Mingers, 2006). These studies demonstrate that the application of multimethodological research can successfully enhance the scope and effectiveness of research in a range of areas. This chapter provides an example of how such a study can be implemented in a healthcare setting.

The inclusion of qualitative modelling in this study allows us to gain an understanding of the challenges hospitals face surrounding theatre scheduling; the intention being both to gain some insight into the lack of implementation and the needs of the particular hospital studied.

This chapter firstly explores the reasons for using a multimethodological approach, particularly including qualitative modelling (soft OR) in a research area that has thus far been dominated by quantitative methods (hard OR). We then go on to explain the qualitative method chosen in this case, how it was implemented and the effects of its implementation.

3.1 Why Qualitative Modelling

Qualitative modelling also referred to as soft OR, allows the incorporation of factors that are difficult to quantify and thus cannot be included in the type of quantitative methods generally applied to operation theatre scheduling (Cardoen et al, 2010a). Often qualitative modelling includes the production of diagrammatic representations of the system and how different factors influence one another.

A number of healthcare related studies use the technique known as System Dynamics (SD), which combines qualitative and quantitative modelling. For example, SD has been used to model the implications of future demands for social services, where the qualitative modelling aspect of SD was particularly helpful in developing an understanding of the problem and being able to test that understanding with social services employees (Desai et al., 2008). Rohleder et al. (2007) also found the qualitative side of SD particularly useful, concluding that the creation of a simple causal loop diagram "may provide valuable insight". Lane et al. (2003) conclude that using system dynamics can be particularly successful with respect to involving clients in the process. This suggests that qualitative methods are valuable for gaining understanding of the problem and keeping clients involved.

There is plenty of evidence in the OR literature for the use of qualitative modelling, both as a tool in itself and in combination with quantitative modelling:
- "it has simply been argued that describing a system is, in itself, a useful thing to do and may lead to better understanding of the problem in question." (Coyle, 2000).
- "The need to build models interactively with participants (managers, owners and associated actors) of problem situations." (Wolstenholme, 1993). If participants are actively involved they will have greater understanding and acceptance of the results of the models.
- "SD combines qualitative and quantitative aspects, and aims to enhance understanding of a system and the relationships between different system components" (Brailsford et al., 2004). Combining qualitative and quantitative modelling can enhance the value of the work overall.

The combining of qualitative and quantitative modelling is not limited to those using SD. Mingers and Brocklesby's (1997) "Framework for Mixing Methodologys" sets out

how the attributes of a variety of methods should be considered when combining methodologies, and is clear that qualitative and quantitative methods and/or aspects of these methods can be combined. Mingers (2001) considers combining methods and concludes that "soft methods and techniques should be used in combination, both with themselves and with more traditional quantitative modelling, to yield a richer form of multimethodology".

In their survey of practitioners' use of multimethodology, Munro and Mingers (2002) find that the majority of those combining methodologies use either only soft or only hard techniques, while "Relatively few combine both hard and soft approaches in a single intervention". Their study identified 21 such examples and they observe that: "although most uses of multimethodology are based on a single paradigm, there is a small but significant movement within OR/MS that is both multimethodological and multiparadigmatic".

The term paradigm is defined by Mingers and Brocklesby (1997) as "a very general set of philosophical assumptions that define the nature of possible research and intervention". A number of authors considering multimethodology discuss potential issues with the mixing of methods from different paradigms: Mingers and Brocklesby (1997) and Kotiadis and Mingers (2006) discuss "paradigm incommensurability" in some depth, while others like Eden at al. (2009) mention it with limited detail. In practice, all of these authors and many others have used multimethodology with no apparent problems arising from mixing methods from different paradigms. Pidd (2004) sums this up by saying "The bumblebee flies, but we just do not understand how", since mixing methodologies works in practice but not in theory. Others, like Zhu (2011), have made a more considered attempt to explore the theoretical basis for mixing methodologies from different paradigms. It seems that a consensus on the topic has yet to be reached, with Harwood's (2011) discussion on the topic attracting swift responses from Mingers (2011) and Jackson (2011).

Kotiadis and Mingers (2006) identify the difficulties of shifting between paradigms for the practitioner as a possible reason for the low incidence of multimethodological studies combining qualitative and quantitative methods, but state that "it is possible for a person to become multimethodology literate given sufficient determination". They go on to

give examples where "both paradigms have been adopted at different stages in the project". This use of different methods at different stages makes the mental adaptation between paradigms easier as it creates some separation between them. This supports the use of soft methods to develop an increased understanding of the problem before applying hard OR methods, which is how the methods are combined in our study on operation theatre scheduling. Such combining of soft and hard OR in series is recognised by Pollack (2009) as a common format for multimethodological projects.

There are other examples of combining hard and soft OR methods in the literature. For example, Robinson (2001) combines simulation with facilitation, Brown et al. (2006) apply hard and soft OR to the tax system and Ackermann et al. (1997) combine simulation and decision analysis to assist with litigation relating to the channel tunnel. Also, Sobolev et al. (2008) provide an example of mixing methodologies in a healthcare setting by incorporating aspects of statecharts in their simulation of patient flow in surgical care.

Perhaps the most relevant example of mixing methodologies is that of Sachdeva et al. (2006), who looked specifically at the effects of mixed methodologies on implementation of operational research in healthcare. Their study combines cognitive mapping and simulation and illustrates that mixed methodologies can be effectively applied in a healthcare setting. Furthermore, their conclusions include:

•   "there is a greater likelihood of acceptance of results emerging from a positivist paradigm, with a unique role for outcomes research for healthcare, which has been enhanced with soft OR" (Sachdeva et al., 2006).

•   "after obtaining a holistic understanding of the system using hard and soft OR, stakeholders were willing to implement results from each independently" (Sachdeva et al., 2006)

As discussed in the introduction, one of the gaps observed in the literature on operating theatre scheduling is a lack of implementation (Cardoen et al., 2010a). Since the use of mixed methodologies enhances the chances of implementation, this suggests a strong argument for the application of mixed methodologies.

Overall, the literature suggests that mixing qualitative and quantitative methodologies has been undertaken successfully in a variety of contexts including healthcare and may have a particular advantage in terms of increased implementation in healthcare. The latter point is particularly important given the lack of evidence of implementation of academic work on operation theatre scheduling.

3.2 Why Cognitive Mapping

In SD, the type of qualitative modelling used is known as influence diagrams. These focus on the interactions, particularly feedback loops, involving different aspects of the system. They form the basis of the quantitative model that can be simulated on a computer in the qualitative modelling.

Based on the methods used in the literature, it is likely that we will be using methods such as Mathematical Programming and Discrete Event Simulation rather than the population level simulation of SD in this project. Mathematical Programming is particularly relevant at the tactical level, as the literature demonstrates (Adan et al. (2009), Belien and Demeulemeester (2007a), Blake et al. (2002), Hans et al. (2007), Van Houdenhoven et al. (2008), Zang et al. (2008), van Oostrum et al. (2008) and Santibanez et al. (2007)) that it has proved possible to formulate and optimize aspects of the problem. Also, in this case, it is desirable to include an awareness of concepts and the links between them as much as the feedback loops of a system, so while previous work with SD illustrates the value of combining quantitative and qualitative modelling, we do not consider SD the best tool for this problem.

Cognitive mapping involves forming of influence diagrams superficially similar to those used in SD (Pidd, 2003). Eden and Ackermann (2001) define a cognitive map as "a model designed to represent the way in which a person defines an issue. It is not a general model of someone's thinking, neither is it intended to be a simulation model of decision making. It is a network of ideas linked by arrows ... . The arrows indicate the way in which one idea may lead to, or have implications for, another."

The objective, given in the introduction to this chapter, from the researcher's point of view is to gain an understanding of the challenges surrounding operating theatre scheduling in hospitals. A model that illustrates the connections between ideas will

allow us to explore these challenges in a structured way including consideration of their interconnections. The production of such a model with hospital staff will enable us to test the correctness of our understanding, as they can see it represented on the map. Therefore, we will use cognitive mapping for the qualitative modelling. This is similar to the approach taken by Franco and Lord (2011), who use cognitive mapping in the first phase of their multimethodological study to "elicit, share and examine stakeholders' views of the situation so that an improved understanding of the issues ... could be achieved among stakeholders". Indeed such use of soft methods at the start of a project is included in Mingers and Brocklesby's (1997) framework, where they refer to it as 'front-ending'.

Other visual mapping methods that could be used are mind mapping and rich pictures; these do not have a cause-and-effect relationship (Daellenbach and McNickle, 2005) and so do not give as much detail on the interrelationships between concepts.

3.3 Implementation

3.3.1 Overall Approach to Cognitive Map Development

Specific methodologies have been developed for the development and analysis of cognitive maps, notably that known as SODA (Strategic Options Development and Analysis). Eden and Ackermann (2001) describe the relation between cognitive mapping and SODA as "SODA is the *approach* to working with clients, out of which has grown the particular *technique* of cognitive mapping.". There are two versions of SODA. The original methodology (SODA I) is based on individual interviews to create individual cognitive maps which are merged and used to explore the problem, while the latter version (SODA II) dispenses with the individual interviews and builds the map directly with the team (Pidd, 2003).

For this study, fifteen individual unstructured interviews were conducted to gain an understanding of the physical restrictions on Operating Theatre Scheduling for including in the hard OR as well as the wider context being explored by the cognitive mapping. Due to restrictions on the availability of hospital staff, there was not sufficient time for individual cognitive maps to be constructed during the interviews that were undertaken. The author also attended meetings and discussions at the hospital on topics related to theatre management; including attending a half day Theatre Workshop lead by the

Theatre division's Clinical Lead and a ten monthly theatre management group meetings. To see the theatres in action a half day of day case orthopaedic surgery was also observed. Information was also drawn from the understanding gained from the review of academic literature.

Given the restrictions on the availability of hospital staff, we adapted SODA I by including individual interviews, but without developing individual maps; this information was then combined with ideas from other sources to create a single map. This map was then presented to hospital staff in a workshop, giving them the opportunity to improve it and work with it to explore future actions.

Figure 1 is the resulting cognitive map summarising the factors affecting theatre management, including potential changes and their likely effects. We do not include financial considerations explicitly as this would over complicate the map because cost affects almost every aspect of theatre management. Also, in the NHS, financial implications are not given the same importance as they would be in a private hospital, since costs are closely linked to the amount of theatre hours available (due to the high staffing costs) and targets like reducing cancellations do not have financial implications. Clinical considerations are not included, except implicitly, as the focus of this work is on the management of theatres. These limitations have restricted the size of the map making it more accessible to hospital staff members who were unfamiliar with cognitive mapping and had limited time available to learn about it and explore the map.

As the concepts have been written concisely on the map Appendix A uses the numbering on the map to give further explanations for many of them.

The meeting with hospital staff to improve the map resulted in the addition of nodes 35, 36, 41, 42, 65, 66 and 67 and a couple of additional arcs. This shows that the original map had captured the majority of relevant concepts and that the meeting allowed refinements to take place.
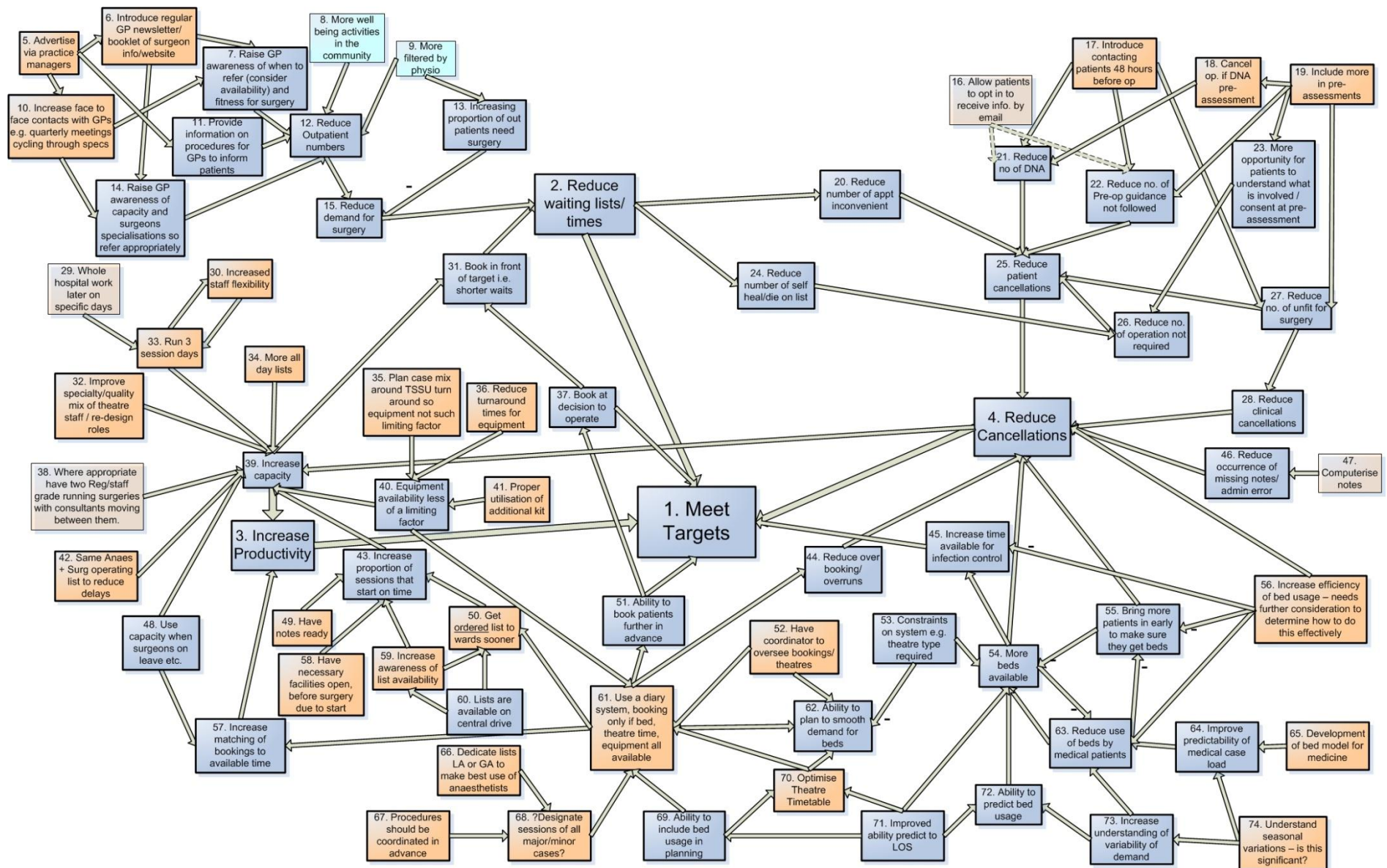
Figure 1: Cognitive map of issues surrounding theatre scheduling

3.3.2 Detail of Map Construction

The following paraphrases the process of drawing a cognitive map given by Pidd (2003):

1. Goals should be identified early since these are the aims and provide the context for the rest of the map.
2. Place other concepts leading to the goals – identify concepts that may be 'strategic issues'
    -   The concepts should be expressed as bipolar concepts; that is as pairs of psychological (not necessarily logical) opposite ideas.
    -   Arrows are placed between pairs of concepts giving the direction of causality, with negative arrows implying that the concept at the tail of the arrow has a negative effect on the concept at the head of the arrow.
3. If possible concepts should be action oriented.
4. The map should ideally be drawn so that it flows upwards with goals at the top.

Due to the complexity of the links between concepts, spreading out the targets (Reduce waiting lists, Reduce Cancellations and Increase productivity) allows much better spacing of the concepts on the map, making it easier to interpret than if we followed the ideal of placing targets at the top of the map.

Also, due to the complexity of the map, only one side of the bipolar construct is included in each case. We have selected the positive side of the construct in most cases to keep the process positive psychologically. This is supported by the statement by Pidd (2003) that "If the concept is clear without the second pole, then one is not needed".

The majority of the links are positive (an increase in the construct at the base of the arrow is expected to result in an increase in the construct at the head of the arrow) so positive signs have not been included; where necessary, negative signs have been included beside the arrows.

The constructs where there is potential for changes to be made, via changes to hospital policy and actions, are represented by tan coloured boxes and the effects of these changes are coloured in blue, which fits with the idea of identifying 'strategic concepts'.

The key targets are in larger boxes and font size so that they stand out. The pale blue concept boxes to the upper left of the map relate to issues out of the control of the hospital, the paler tan coloured boxes are possible long term changes and the dashed arrows indicate connections about which there is considerable uncertainty over both their existence and strength.

3.4 Analysis and Discussion

Daellenbach and McNickle (2005) describe the process of analysing cognitive maps as;

• Starting with a detailed examination of the map by the clients to ensure it is accurate and there are no other concepts to add to it.

• Checking for feedback loops, identification of core constructs (those with many arrows to/from them)

• Looking for emerging themes (highly interlinked groups of constructs with few links to the rest of the map).

Eden and Ackermann (2001) also emphasise the consideration of clusters of nodes on the map. Such analysis of our map is described in the paragraphs that follow.

At the workshop mentioned previously, the map was examined in detail, with hospital staff using large printouts to which additions and corrections could easily be made by hand. The staff attending this workshop were a mixture of hospital theatre managers and other theatre staff; the meeting was advertised to all staff whose work related to theatres and those who were available attended. The workshop lasted for an hour and a half and was led by the author. It enabled us to explain the content of the cognitive map to staff as well as involving them in checking its accuracy and adding extra concepts. The purpose of the workshop was both to test our understanding of the problem and to help participants develop ideas and consider the interactions between concepts. As the analysis of the map had been begun before the workshop, this could be continued in the workshop to provide feedback to participants.

It is interesting to note the low number of feedback loops in the diagram, which may be due to the small size of the map. Figure 2 reproduces a particularly significant feedback loop with regard to the availability of hospital beds, if more patients are brought in early, then there will be fewer beds available, which will reduce use of beds by medical patients, encouraging surgeons to continue to bring patients in early. Conversely, if

fewer patients are brought in early, more beds will be available and the use of beds by medical patents may increase. Bringing patients in early increases their length of stay in hospital and thus increases the cost of providing their treatment as well as being an inefficient use of beds. This suggests that the use of beds by medical patients can cause surgeons to act defensively and thus use beds less efficiently, and this is an area worthy of further attention.
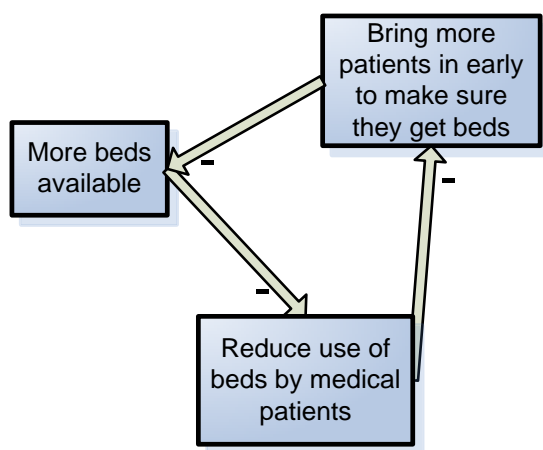
Figure 2: Feedback loop regarding bed availability

The potential change of 'Use a diary system, booking only if bed, theatre time and equipment are available' links into a significant number of areas and affects all of the targets (some more directly than others), making this a core construct. This demonstrates the significance of the potential effects of the tool being developed by this research. Reducing cancellations is core target, reflecting the significance of this in meeting other targets and the disruption that cancellations cause to individual patients. Demand for healthcare, including operations is increasing, so action that will allow more operations to take place is seen as important and therefore increasing capacity is a core construct. Increasing capacity also has the advantage of allowing greater flexibility.

In the upper left corner of the map, the cluster of nodes relating to communication with GPs is highly interconnected with only one connection to the rest of the map and is thus an emerging theme. It is an area that has received less attention recently, which is being considered for further work within the hospital theatre management team.

The other clusters are less well defined, but are predominantly groups of concepts surrounding the main targets.

Overall the map illustrates the complexity of the issues surrounding operating theatre management and scheduling. The significant number of nodes relating to improving the use of beds and the effects of such on cancellations demonstrates the importance of considering the effects on bed usage within a theatre scheduling system. Since the map shows that theatre scheduling can have a significant effect on the efficiency with which theatres are used, and influences the number of cancellations both directly and via the effect on bed usage, it illustrates the importance of considering operating theatre scheduling. The map also demonstrates that there are a range of other factors affecting cancellations (upper right corner) and there are other considerations which the scheduling method cannot be expected to consider, such as the emerging theme surrounding contact with GPs (upper left corner).

Some of the concepts included on the map have direct implications for the creation of methods to support operating theatre scheduling. For example, the middle of the left hand side of the map has concepts relating to the possibility of running more theatre sessions per day on some days of the week, suggesting this should to be considered in any further work to support theatre scheduling. To the right of that is the concept of running more all day lists (that is the same surgeon in a theatre all day). The extent to which this is desirable should also be considered.

3.5 Conclusions

In addition to developing our understanding of the issues surrounding theatre scheduling the cognitive map illustrates the importance of some of the decisions being taken within the hospital. In addition to our research aims, one of the directors who requested a copy of the cognitive map, and found it particularly helpful in explaining links between concepts, which he found obvious, to others.

The process in itself was considered valuable within the hospital, as it increased their understanding of the interactions between the changes they were considering and their targets. Additionally, the cognitive mapping allowed the author to meet the hospital staff and introduced them to the project as a whole and its potential significance to the hospital.

This mapping process demonstrated the need for research to improve theatre scheduling in hospitals, thus providing motivation for the work discussed in the chapters that follow. The detailed understanding of the running of theatres and challenges faced by staff proved invaluable in determining the factors to consider in designing master theatre timetables, as is discussed in Sections 4.1.2 and 4.1.3. In particular the inclusion of consideration of flexibility over the number of theatre slots per day, the desirability of having the same surgeon having consecutive slots in one theatre on the same day, the importance of bed availability and the need to smooth bed usage have all been taken directly from this process in to the modelling in Chapter 4.

For the day-to-day scheduling case study (Chapter 6) we worked with a surgeon at a different hospital and the understanding of the issues surrounding theatre scheduling gained from the cognitive mapping assisted greatly with understanding his concerns. The importance of using theatre time effectively, reducing waiting times and avoiding cancellations identified by the cognitive mapping transferred over, as did many of the other aspects from the map. We also continued the practice of working closely with hospital staff and discussing the modelling with them throughout Chapter 6.

Thus, the cognitive mapping process has developed the understanding of the issues surrounding theatre scheduling of both the author and the hospital theatre managers. A benefit of this approach is to ensure that the researchers are aware of the important factors to include in further modelling work and of the wider context into which such modelling fits. This is proving invaluable in our on-going modelling work with both the original partner hospital and an additional local hospital.

Chapter 4: Developing Master Surgical Timetables

This chapter describes the development of the system to facilitate the generation of master surgical timetables. Starting by defining the problem, followed by the process of selecting an appropriate method, we describe the mathematical formulation of the problem and how the formulation is implemented. Then results from running the model are presented, and consideration is given as to how the formulation could be improved. The chapter ends with sensitivity analysis and discussion of the conclusions that can be drawn from this process.

4.1 Defining the Problem

The different levels of theatre scheduling are discussed in detail in the literature review, Chapter 2, along with the reasons for giving consideration to the master surgical timetable. This section recaps some of the findings from the literature review adding information from contact with the collaborating hospital to define the problem in detail.

4.1.1 The Literature

A number of papers have addressed the development of master surgical timetables (or schedules). Belien and Demeulemeester (2007a) define the master surgical schedule thus: "a cyclic timetable that defines the number of operating rooms available, the hours that the rooms will be open, and the surgical groups or surgeons who are to be given priority for the operating room times." Similar definitions are also given by Blake et al. (2002), Blake and Donald (2002), Santibanez et al. (2007), Testi et al. (2007), Oostrum et al. (2008), Belien et al. (2006) and Hans et al. (2007, 2008). Thus, there is considerable agreement in the literature about the definition of the overall problem.

All of the papers on the topic consider the amount of operating time available and the amount required by each specialty (surgical group) or surgeon. As discussed in Chapter 2, Cardoen et al.'s (2010a) review of the literature on operating room scheduling includes a table of the other aspects considered including wards, ICU (intensive care unit), equipment, surgical staff, budget and operating room overtime/undertime. The various studies relating to operation scheduling have considered different combinations of these, usually as constraints on the timetables generated. There are other factors that

have been considered in some studies, for example, the only study we have found that gives explicit consideration to surgeons' preferences for particular days or times of day is Santibanez et al. (2007). Thus, there are a significant number of factors involved in theatre scheduling and we will need to decide which to include in to our approach to the problem.

There is recognition in the literature of the stochastic nature of operation scheduling (see Section 2.5), as the operation duration and the patients' length of stay in hospital will vary between individual patients. Cardoen et al. (2010a) identify a substantial number of studies that take account of the stochastic nature of at least one aspect of theatre scheduling, although they are considering all levels of theatre scheduling not just master surgical timetable development. Many of these studies consider the stochastic aspects of the problem by using simulation techniques, examples of which can be found in Arenas et al. (2002), Ballard and Kuhl (2002), Bowers and Mould (2005) and Testi et al. (2007).

A limited number of studies incorporate stochastic aspects of the problem, while using methods other than simulation. Belien et al. (2007) explore various methods of incorporating the stochastic nature of patients' length of stay in hospital following surgery. However, they only consider the specialties allocated and the operating room time available, so the study is limited. Gallivan and Utley (2005) also consider the stochastic nature of length of stay, for scheduling individual procedures rather than blocks assigned to surgeons over a cyclic timetable.

As discussed in Section 2.6, the majority of studies regarding the development master surgical timetables do not mention implementation in real hospitals, or only mention having obtained data from hospitals, so it is unclear if they have been implemented. A notable exception to this is the work of Blake and Donald (2002), which has been in use at Mount Sinai hospital since 1996. The model in question assigns theatre time to surgical divisions, when changes to the current allocation are required. The model is relatively straightforward, which implies that perhaps some of the models that have been developed may not be implemented because they are too complex for hospital staff to understand. The complexity of the model needs to be balanced with the need to model the system accurately and take account of sufficient factors within it.

4.1.2 Interviews with Hospital Staff

As discussed in Chapter 3, interviews with staff at a collaborating hospital were conducted as part of this study. In addition to developing the cognitive map, these interviews have greatly assisted the author in understanding the constraints on the master surgical timetable. Of the constraints identified in the literature review, the interviews revealed that operating room time, surgeons (and other staff) availability and the equipment required are all significant.

Notably the interviews have raised the issue of the operating room type required for different operations, which is not mentioned in Cardoen et al. (2010a). Indeed the literature generally considers the theatres to be identical and this is mentioned explicitly by Fei et al. (2008), Hans et al. (2008), Lamiri et al. (2009) and Oostrum et al. (2008). The exceptions to this are Santibanez et al. (2007), Testi et al. (2007) and Zhang (2008a). The former "require compatibility between the operating room and the specialty". Testi et al. (2007) and Zhang (2008a) consider theatres with particular characteristics, but they do not allow any theatre to belong to more than one type, which can be the case in reality.

Surveys of operating theatres in Great Britain and Ireland by Humphreys et al. (1995) and Smyth et al. (2005) indicate that the majority of hospitals have different theatre types for different types of surgery. In 1995 "only 32% did not have a designated theatre for any specialist surgery" (Humphreys et al., 1995) and in 2005 the extent of theatre designation has increased so that 80% of hospitals have designated theatres for orthopaedic surgery and 50% have "designated theatres for a variety of other surgical subspecialties". Thus, not only do the vast majority of British hospitals have different types of theatre for specific types of operation, but also the extent to which this is the case is increasing. This implies that any model that is to be useful to hospitals in carrying out their theatre scheduling must take account of theatre types.

The list of constraints from Cardoen et al. (2010a) given above includes wards and ICU, the latter being a specialised type of ward. The cognitive mapping exercise highlights the constraints on the availability of beds, particularly in ICUs, as important in reducing the number of cancellations, and therefore in enabling hospitals to meet the targets set by

the government. The interviews also reveal that the type of ward is important as patients are separated based on their needs and the infections they have been screened for.

The hospital has 11 theatres of 6 different types and they operate a cyclic timetable over two weeks with morning and afternoon slots in all theatres. For some theatres, the afternoon slot is extended into an afternoon/evening slot.

Discussions with surgeons and other surgical staff have revealed that they find it significantly reduces pressure on them if the same surgeon has an all-day slot, i.e. the same surgeon is assigned the morning and afternoon slots for a particular day. On the other hand, having the same surgeon assigned a morning slot in one theatre followed by an afternoon slot in another theatre is undesirable, as if the morning slot runs late then the afternoons surgery for two theatres will start late causing considerable disruption. It is also preferable to have the timetable repeating weekly if possible, so that it is easier to remember it, which is one of the objectives in Blake et al. (2002) and Belien et al. (2009) for tactical scheduling.

The hospital recently redesigned their master theatre timetable to take into account changes in staffing. This was a substantial project for the hospital managers, which took place over a 4 month period and culminated in a small number of managers spending a weekend rearranging a paper timetable to achieve the final version. At the start of this project, the intention was to ask surgeons for their preferences and incorporate these into the final schedule. This was not possible due to the complexity of finding a feasible timetable by hand, but it does raise the desirability of taking account of the preferences of those involved in developing timetables. The literature review (Chapter 2) discusses studies that include preferences in the objectives. Of these studies four are considering preferences while generating master surgical timetables. Specifically, Ozkarahan (2000) and Belien et al. (2009) consider preferred operating rooms, Blake and Carter (2002) consider ensuring that "physicians are able to generate a preferred level of income", and Testi et al. (2007) define surgeon preference based on length of stay, with the aim of scheduling short stay patients at the start of the week so that wards can close at weekends, which is more a preference relating to wards than to surgeons. Thus, some surgeons' preferences are considered in the literature, but they are not asked to provide a full list of their preferences by theatre, day of the week and time of day.

The difficulties experienced by the hospital, with which I was working, in obtaining a feasible timetable, highlight how constrained the timetable is. This is largely due to the desirability of using theatres as much as possible and the constraints on the availability of surgeons and is considered further in Section 4.3.3.

The staff at the hospital with which I was working were also unable to take account of the potential effects on bed usage of the new timetable, as computerised support was not used. This highlights the need of a model that can take into account the relevant factors to assist hospital surgical divisions in devising new timetables.

4.1.3 Identifying the Important Factors

Based on the interviews with hospital staff and the areas mentioned in the literature review, in order to develop useful master surgical schedules the following factors should be taken into account:

- The amount of theatre time available.
- The types of the theatres, as some procedures will require equipment that is not available in all theatres.
- The availability of beds in wards.
- The amount of theatre time to be assigned for each specialty or surgeon.
- The availability of surgeons and if possible their preferences.
- The availability of other resources, such as other staff and equipment.
- The desirability of all day slots and repeating slots weekly.
- The desirability that the same surgeon does not have a morning slot in one theatre followed by an afternoon slot in a different theatre on the same day.
- The stochastic nature of the problem.

The factors of theatre scheduling considered in the literature listed in Section 2.4 also include the waiting times, numbers differed/refused, precedence constraints, release dates and due dates. These are all aspects of theatre scheduling that relate to individual patients and are therefore not relevant at the tactical level of scheduling, when we are not considering individual patients.

Holding areas for before surgery starts and post anaesthesia care are also not included in our model because they did not arise as a limitation theatre scheduling in our discussions with hospital staff. While space in intensive care units is not explicitly included in this list they are a type of ward and can be included as such if wards are considered.

Staff other than surgeons can be considered by modelling them using the ability to limit the availability of equipment. They're scheduling is not modelled explicitly as within the partner hospital this is done on a rolling bases independent from the creation of the master theatre timetable.

Some of the factors considered will form hard constraints on the model and some are objectives to be aimed for in the timetables produced. As discussed in Sections 4.1.1 and 4.1.2, there are other studies that have looked at various combinations of these factors, but we have been unable to find an example incorporating all of them.

4.2 Method Selection

4.2.1 Type of Method

The variety of methods that have been applied to operation theatre scheduling are discussed in the literature review (Chapter 2), so this section will give a brief overview of the methods considered and the reasons for the selections made.

The types of method applied to operation theatre scheduling fall into three broad groups, simulation, optimisation using variations on linear programming and heuristics. All of these methods and the ways in which they have been applied to the different aspects of operation theatre scheduling are discussed in the literature review (Chapter 2).

Running a simulation model only distinguishes between the polices that are tested. In the case of the master surgical timetable, there may be a large number of feasible timetables to consider and running separate simulations for each would take an unreasonably long time. Also, a significant part of the problem for some instances will be finding feasible timetables to compare and simulation is not able to help with this. Therefore, while simulation is used by a number of authors to address aspects of theatre scheduling, it is not the method best suited to assist with the development of master surgical timetables.

The literature review reveals that the majority of studies have tackled the problem of developing master surgical timetables using IP or MIP techniques, so it seems reasonable to assume that such techniques can find optimal solutions in reasonable time, and therefore it is not necessary to develop heuristics to solve the problem. As discussed in Section 4.1, there are a significant number of factors to be considered in planning the master surgical timetable, and none of the studies reviewed addresses all of them. Thus, it may be that when all of these factors are included it is not possible to obtain a solution in reasonable time and heuristics may be required.

Based on the discussion above, we will first address the problem using IP/MIP techniques. However, if this does not give solutions in reasonable time we will consider the use of heuristics.

4.2.2 Type of Optimisation

The available techniques for solving IP/MIP problems include branch and bound, applying cutting planes, branch and cut and price and cut. For some problems the choice of technique affects the ability to produce solutions in a reasonable amount of time. The software available to us on existing licences, and therefore easiest to access, uses branch and bound. Therefore, branch and bound methods will be implemented first. If this approach does not produce acceptable results in reasonable time, then we will first look for cutting planes to improve the solution, if the results are still not acceptable then we will move to branch and cut and then if necessary branch and price.

4.2.3 Addressing Multiple Objectives

As discussed in Section 4.1.3, our problem has multiple objectives – it is desirable to smooth bed usage, give surgeons slots they prefer, have all day slots and repeat assignments weekly. This can be addressed using multi-objective mixed integer programming (Winston, 1994), where the objective function is the sum of weighted objectives. The ability to weight the importance of the various objectives is particularly desirable in this case as different hospitals may have different priorities and the objective function can therefore be adapted accordingly.

4.2.4 Cyclic or Dynamic?

There are fluctuations in demand at different times of year, so ideally the timetable would change dynamically to take account of demand variations and any other changes in the hospital. Currently cyclic timetables are used in this country, including at the partner hospital, so as one of the goals is to propose a model for implementation we will aim to stay close to the current system and produce a cyclic timetable.

Providing a computerised tool that allows hospitals to quickly find new timetables, rather than the lengthy paper exercise discussed in Section 4.1.2, creates the possibility of creating a new timetable several times during the year to reflect the changes that occur. Thus, with cyclic master surgical timetable the advantages of considering changes over the year can still be obtained, without a truly dynamic timetable.

4.3 Integer Programming Model Formulation

This section begins by describing the required elements of the mathematical formulation, before going on to set out the formulation in detail. The reasons for the various components of the model and how they are formulated are discussed in detail where this is appropriate.

4.3.1 Selecting Objectives

The most important objective is to find a feasible solution, which may not be straight forward given how heavily constrained the problem can be. Using MIP does not guarantee to find a feasible solution, but if such a solution exists, a computer can search faster than working by hand. Feasibility will be discussed in more detail later in Section 4.4.4.

For now, we will assume that a feasible solution exists, and the following are the objectives for the MIP:

- To schedule theatre slots such that the expected bed usage is smoothed. This is important since avoiding high bed usage will reduce the likelihood of cancellations being necessary due to lack of beds.
- To assign surgeons to slots for which they have high preference scores.
- To avoid assigning surgeons to slots for which they have low preference scores.

- To schedule two consecutive slots on the same day were one surgeon is in the same theatre as often as possible.
- To avoid scheduling consecutive slots on any day where the same surgeon is in two different theatres.
- To repeat the scheduling of the same surgeon in the same theatre each week as much as possible.

The relative importance given to each of these constraints may vary between hospitals. Indeed a hospital may wish to explore the effects on the suggested timetable of varying the weightings given. Therefore, each of these factors is given a weighting in the objective, which can be adjusted as required.

4.3.2 Selecting Variables

The aim of the master timetable is to assign theatre slots to surgeons, so the first variable is an array indicating if each surgeon is assigned to each slot, for all slots, theatres and days of the cycle.

The objective function needs to be linear, so further variables are required enable the counting of the number of times the following occur for use in the objective function:
- Slots that have been assigned to surgeons with low preference scores for those slots.
- If the same surgeon is in the same theatre for consecutive slots on the same day.
- If any surgeon is in the different theatres for consecutive slots on the same day.
- If the same surgeon is in the same theatre at the same time in consecutive weeks.

These will all be binary variables, set to be 1 if the condition specified holds and 0 otherwise.

Similarly, variables are required to keep track of the number of beds required for each day of the cycle.

4.3.3 Selecting Constraints

Some of the constraints on this problem are more straight forward than others. The more obvious constraints are:
- Slots can only be used if they are available and can only be used once.

- Each surgeon can be in at most one theatre at any time.
- Surgeons can only be assigned to slots when they are available to work in those slots.
- There may be a limit on the number of slots a surgeon can do in one day.
- The schedule must not require more of any equipment/ resource than is available.

Constraints are also required to set the binary variables discussed above in Section 4.3.2, based on the values of the array assigning surgeons to theatre slots.

If all of the theatres were of the same type or each type of operation could only be done in one type of theatre, then the constraint to meet demand would also be straightforward. However, as some types of operation could be done in more than one type of theatre more complex constraints are required. This is discussed in detail with the mathematical formulation of the constraints below in Section 4.3.6.3.

Further constraints are required to set the values of the variables used to count the expected number of beds required each day. These are discussed in Section 4.3.6.4.

4.3.4 Notation

The following sets out the notation used for the mathematical formulation of the problem;

C = The cycle length in days

d = The days of the cycle $1 \leq d \leq C$

T = The number of theatres

t = The individual theatres $1 \leq t \leq T$

H = The number of sets of theatre types

h = The individual sets of theatre types $1 \leq h \leq H$

I = The number of surgeons

i = The individual surgeons $1 \leq i \leq I$

S = The maximum number of slots in a theatre each day

s = The individual theatre slots $1 \leq s \leq S$

E = The number of equipment (recourse) types to consider

e = The individual equipment types $1 \leq e \leq E$

J = The maximum number of days a patient spends in a bed (this is set to be a large enough value that the number of patients with longer stays is negligible and can be ignored).

j = The different numbers of days patients spend in beds $0 \leq j \leq J$

K = The number of wards to consider

k = The individual wards $1 \leq k \leq K$

w = The desired frequency of repeating with in the cycle, set to 0 if no repeats are considered desirable, otherwise it must divide the cycle length exactly (usually weekly repeats will be desirable).

$$G_{h,t} = \begin{cases} 1 \text{ If } t \text{ is of typea typewithin set } h. \\ 0 \text{ Otherwise} \end{cases} \qquad \forall h,t$$

For example the set of types that can be used for daycase surgery includes most of the theatre types, but the set of types that can be used for major orthopaedic surgery is limited to a particular type of main theatres with special air filtration systems.

$R_{h,i}$ = the number of slots of within the set of types h required by surgeon i over the whole cycle (these values come from the strategic theatre scheduling decisions). $\forall h,i$

$$A_{t,d,s} = \begin{cases} 1 \text{ If slot } s \text{ on day } d \text{ in theatre } t \text{ is available.} \\ 0 \text{ Otherwise} \end{cases} \qquad \forall t,d,s$$

r = The value below which a preference score is considered low, this must be a non-negative number.

$P_{i,t,d,s}$ = The score given by surgeon i, for slot s on day d in theatre t, these must be non-negative numbers. $\forall i,t,d,s$

$$Q_{i,t,d,s} = \begin{cases} 1 \text{ If surgeon } i \text{ is available to operatein slot } s \text{ on day } d \text{ in} \\ \quad \text{theatre } t. \\ 0 \text{ Otherwise} \end{cases} \qquad \forall i,t,d,s$$

f_e = The number of sets of equipment of type e available. $\forall e$

$$F_{i,t,e} = \begin{cases} 1 \text{ If surgeon } i \text{ requires equipment } e \text{ when they are in in theatre } t. \\ 0 \text{ Otherwise} \end{cases} \qquad \forall i,t,e$$

$B_{i,t,j,k}$ = The number of patients who are still in beds used in ward k, j days after having an operation by surgeon i in theatre t (allowing for the long term cyclic implementation of the theatre schedule, see Section 4.3.6.4, for explanation of how this is used). $\forall i,t,j,k$

$D_{d,k}$ = the max number of beds available on day d in ward k. $\qquad\qquad \forall d,k$

M_i = the max number slots surgeon i can do in a day. $\qquad\qquad \forall i$

α, β, χ, δ, ε, γ are multipliers for use in the objective function.

4.3.5 Variables

The variables described in Section 4.3.2 can be defined as follows;

$$X_{i,t,d,s} = \begin{cases} 1 \text{ If surgeon } i \text{ is assigned to operate in slot } s \text{ on day } d \text{ in theatre } t. \\ \\ 0 \text{ Otherwise} \end{cases} \qquad \forall i,t,d,s$$

$$Y_{i,t,d,s} = \begin{cases} 1 \text{ If surgeon } i \text{ is assigned to operate in slot } s \text{ on day } d \text{ in theatre } t \\ \quad \text{and their preference score for such is less than } r. \\ \\ 0 \text{ Otherwise} \end{cases} \qquad \forall i,t,d,s$$

$$U_{i,t,d,s} = \begin{cases} 1 \text{ If surgeon } i \text{ is assigned to operate in slot } s \text{ on day } d \text{ in theatre } t \\ \text{and they are also assigned to operate in slot } s+1 \text{ on day } d \text{ in} \\ \text{theatre } t. \\ \\ 0 \text{ Otherwise} \end{cases} \qquad \begin{array}{l} \forall i,t,d \\ \forall s \le S-1 \end{array}$$

$$V_{i,t,d,s} = \begin{cases} 1 \text{ If surgeon } i \text{ is assigned to operate in slot } s \text{ on day } d \text{ in theatre } t \\ \quad \text{and they are also assigned to operate in slot } s+1 \text{ on day } d \text{ in a} \\ \quad \text{theatre other than } t. \\ \\ 0 \text{ Otherwise} \end{cases} \qquad \begin{array}{l} \forall i,t,d \\ \forall s \le S-1 \end{array}$$

$$W_{i,t,d,s} = \begin{cases} 1 \text{ If surgeon } i \text{ is assigned to operate in slot } s \text{ on day } d \text{ in theatre } t \\ \text{and they are also assigned to operate in slot } s \text{ on day } d-w \text{ in} \\ \text{theatre } t. \\ \\ 0 \text{ Otherwise} \end{cases} \qquad \forall i,t,d>w,s$$

$\mu_{d,k}$ = Is the expected number of beds required in ward k on day d. $\qquad\qquad \forall d,k$

Z_k = The minimum difference between the expected number of beds
required and the beds available on each of the days of the cycle k. $\qquad\qquad \forall k$

4.3.6 Mathematical Formulation of Constraints

4.3.6.1 Constraints linking other variables to X

The constraints on the type of value *X, Y, U, V* and *W* can take and setting *Y, U, V* and *W* from the value of *X* are;

X, Y, U. V and *W* are all binary.

To assign *Y*;

$$rX_{i,t,d,s} - P_{i,t,d,s}X_{i,t,d,s} \leq rY_{i,t,d,s} \qquad \forall i,t,d,s \qquad (1)$$

To assign *U* for *s = S*;

$$U_{i,t,d,S} = 0 \qquad \forall i,t,d \qquad (2)$$

To assign *U* for *s < S*;

$$X_{i,t,d,s} + X_{i,t,d,s+1} \geq 2U_{i,t,d,s} \qquad \forall i,t,d,s < S \qquad (3)$$

To assign *V* for *s = S*;

$$V_{i,t,d,S} = 0 \qquad \forall i,t,d \qquad (4)$$

To assign *V* for *s < S*;

$$X_{i,t,d,s} + \sum_{v=1,v \neq t}^{T} X_{i,v,d,s+1} \leq 1 + V_{i,t,d,s} \qquad \forall i,t,d,s < S \qquad (5)$$

To assign *W* for *d > w*;

$$X_{i,t,d,s} + X_{i,t,d-w,s} \geq 2W_{i,t,d,s} \qquad \forall i,t,d > w,s \qquad (6)$$

To assign *W* for *d ≤ w* ;

$$X_{i,t,d,s} + X_{i,t,C+d-w,s} \geq 2W_{i,t,d,s} \qquad \forall i,t,d \leq w,s \qquad (7)$$

The inclusion of both constraints (6) and (7) means that each repeat is counted once for each time it repeats within the cycle. For a 2 weekly cycle with weekly repeats desirable this counts each repeat twice, for a 3 weekly cycle with weekly repeats it counts each repeat three times, if it occurs in all 3 weeks. In cycles of 3 or more weeks this allows counting of repeats that occur in some but not all pairs of weeks of the cycle.

4.3.6.2 Other straight forward constraints

The straight forward constraints, based on 4.3.3 are given below. Where a summation is used then unless otherwise specified it is over all possible values of the case values given.

Slots can only be used if they are available;

$$\sum_i X_{i,t,d,s} \le A_{t,d,s} \qquad \forall t,d,s \tag{8}$$

Each surgeon is in at most one theatre at any time;

$$\sum_t X_{i,t,d,s} \le 1 \qquad \forall i,d,s \tag{9}$$

Surgeons' availability constraint;

$$X_{i,t,d,s} \le Q_{i,t,d,s} \qquad \forall i,t,d,s \tag{10}$$

Limit on number of slots each surgeon can do per day;

$$\sum_{s,t} X_{i,t,d,s} \le M_i \qquad \forall i,d \tag{11}$$

Equipment constraint;

$$\sum_{i,t} X_{i,t,d,s} F_{i,t,e} \le f_e \qquad \forall s,d,e \tag{12}$$

4.3.6.3 Demand constraints

Since the theatres could theoretically belong to any mix of sets of types and the demand is by set of types of theatre, it is not trivial to ensure that each surgeon's theatres are meeting their demand for different sets of types of theatre, without adding an additional variable for the set of types that each theatre is being used as to a problem with a large number of variables. Additionally if surgeons have 2 slots in a theatre which can deal with cases of the types they would do in type A or type B theatres, then they have more flexibility if they can schedule either type of case into either slot, rather than having to treat one slot as type A and one as type B, so additional variables would not be a good match for what is likely to occur in practice. Hence, there is a need for the following three constraints;

At least meet each surgeon's demand for each set of types of theatre;

$$\sum_t G_{h,t} \sum_{d,s} X_{i,t,d,s} \ge R_{h,i} \qquad \forall h,i \tag{13}$$

Each surgeon's overall demand is met exactly;

$$\sum_{t,d,s} X_{i,t,d,s} = \sum_h R_{h,i} \qquad \forall i \tag{14}$$

Each surgeon does not use each theatre more times than their total demand for its sets of types;

$$\sum_{d,s} X_{i,t,d,s} \leq \sum_{h} R_{h,i} G_{h,t} \qquad \forall i,t \qquad (15)$$

As each surgeon must have at least their demand for each set of types of theatre met, we require constraint (13).

Suppose that we just count the number of theatres a surgeon has of each set of types against their demand for that set of types (just constraint (13)). This could allow surgeons to be assigned more slots than they need, which would not fit with the objective of using theatre time as efficiently as possible so an additional constraint is needed to say that the number of theatre slots assigned to each surgeon is equal to the total of their demand for theatre slots; hence constraint (14).

Suppose we have sets of theatre types A, B and C, with theatre 1 in set A and theatre 2 in sets B and C. If a surgeon requires two slots in set of types B and one in set of types C then assigning them 1 slot in theatre 1 and 2 slots in theatre 2 will meet the constraints so far, as the two slots in theatre 2 mean there are two potential slots of set of types B or C available (meeting constraint (13)) and the total number of slots assigned is 3 (meeting constraint (14)), but there is no way that 1 slot in theatre 1 and 2 in theatre 2 can be assigned to be two slots of set of types B and one of set of types C as theatre 1 can only be of type A. Hence the need for constraint (15) that a surgeon is not assigned more slots in a theatre than their total demand for its sets of types. Thus, in this example, theatre 1 could not be assigned and this difficulty is resolved.

The above argument demonstrates the need for the constraints given, but it does not exclude the possibility that there are other possible situations where the demand is not met as intended, while these constraints hold, i.e. another constraint may be required. The following argument uses induction to show that these three constraints are sufficient, by showing that the constraints ensure that it would always be possible to assign a set of types to each theatre and thus the demand for each set of types is met.

Case 1: If all of the theatres are of all sets of types then $\forall h,t \ G_{h,t} = 1$, and effectively $H = 1$ as we only need to consider one set of types, so they can be ignored in the constraints, which become;

$$\sum_{t,d,s} X_{i,t,d,s} \geq R_{1,i} \qquad \forall i \tag{13}$$

$$\sum_{t,d,s} X_{i,t,d,s} = R_{1,i} \qquad \forall i \tag{14}$$

$$\sum_{d,s} X_{i,t,d,s} \leq R_{1,i} \qquad \forall i,t \tag{15}$$

In fact in this case constraints (13) and (15) are redundant as constraint (14) is a tighter version of either of them. In this case as all theatres are of all set of types we just need the number of theatres each surgeon requires to be able to assign the relevant types to each theatre and meet each surgeons demand. If the problem is feasible then the demand can be met and only constraint (14) is needed.

Case n-1: The demand has been met with the current mix of sets of types for each theatre.

Case n: The theatres remain of sets of types as in case $n-1$ except theatre K which is of all sets of types as in case n-1 except k (when it was assigned set of types k for some part of the solution in case $n-1$), then;

Constraint (13) remains as in case n−1 except for theatre K and set of types k, as now $G_{h,t} = 0$ when $h = k$ and $t = K$ so one term is removed from the sum which is greater than or equal to the demand for type k. This ensures that where in the solution to case $n-1$ theatre K was being assigned set of types k then in the solution to case n theatre K cannot be treated as meeting some of the demand for set of types k. And the demand for the other types is met as in the case $n-1$.

Constraint (14) continues to say that the total demand for theatres for each surgeon remains equal to the total number of theatres they are assigned.

Constraint (15) becomes tighter as $R_{h,i}G_{h,t} = 0$ when $h = k$ and $t = K$, where it was 1 before, so for each surgeon the limit on then number of times theatre K can be used reduced by their demand for set of types k, ensuring that the demand for set of types k is met by other theatres where a slot in theatre K had been assigned as set of types k that demand must now be met by a different theatre.

So where in a solution to case $n-1$ the demand for type k was met by theatre K then in case n that demand cannot be met by type k so the number of times theatre K can be used is reduced by constraint (15) and constraint (13) ensures that there must be available capacity in other theatres of type k to continue to meet the demand. In other words where a slot theatre K was being assigned type k, it must no longer be assigned as type k and a different theatre must be assigned instead.

So by induction we can move from case 1 to any combination of the possible sets of types of theatre and the demand constraints are sufficient to ensure that demand is met (N.B. the problem may be infeasible for some combinations of sets of theatre types, so going from case $n-1$ to case n may make the problem infeasible).

Along with the consideration of types of theatres this set of constraints to address the possible combinations of ways of meeting the requirements for theatres of different types is a new addition to the literature on tactical theatre scheduling.

4.3.6.4 Constraints relating to use of beds

The constraints relating to the availability of beds are:

$$\mu_{d,k} \geq 0, Z_k \geq 0 \qquad\qquad \forall d,k \qquad\qquad (16)$$

To assign $\mu_{d,k}$;

$$\mu_{d,k} = \sum_{i,s,t}\sum_{m=1}^{C}\left(X_{i,t,m,s} \sum_{n \in N} B_{i,t,(nC+d-m),k}\right) \qquad \forall d,k \qquad\qquad (17)$$

where N is the set of values of n such that;

$$\frac{m-d}{C} \leq n \leq \frac{J+m-d}{C}$$

To assign Z_k;

$$Z_k \leq D_{d,k} - \mu_{d,k} \qquad\qquad \forall d,k \qquad\qquad (18)$$

This method of smoothing bed usage is based on that used by Gallivan and Utley (2005) in their work on booking patients into a treatment centre. In their study, individual procedures are scheduled into a cyclic timetable and the probability that a patient is still in a bed a number of days after their operation is used to calculate the expected number of patients in beds on each day of the cycle.

For the master timetable, we are scheduling slots to surgeons, rather than individual procedures. As the case mix of each surgeon and the number of cases they treat in a slot can vary, it is not possible to work at the level of individual patients. Thus, our variation of the model used by Gallivan and Utley (2005) includes changes to adapt it to the new context, including moving from the probability of a patient being in a bed, to the expected number of beds used on each day of the cycle resulting from a surgeon's use of a theatre slot.

Using historical data, the expected number of patients in beds in ward k for each of the j days after each surgeon i has had a theatre slot in each theatre t is calculated, for inputting into the model as $B_{i,t,j,k}$. Constraint (17) uses this information to calculate the sum of the expected number of patients from each assigned theatre slot to each day of the cycle. This is similar to Gallivan and Utley's (2005) method for counting the expected contribution of each timetabled procedure to the number of patients in beds on each day of the cycle.

Constraint (17) is constructed as follows;

- As defined in Section 4.3.4 $B_{i,t,j,k}$ is the number of patients in beds in ward k, j days after surgeon i has a slot in theatre t, and j starts at 0 on the day of surgery.

- To calculate the number of patients in beds in ward k on day d of the cycle resulting from surgeon i having a slot in theatre t on day m of that cycle if $m \leq d$, then we require $j = d - m$, so that j is the number of days between m and d, giving $B_{i,t,(d-m),k}$. If $m > d$ then day m occurs after d in the cycle so its occurrence in the current cycle does not make any contribution to the number of patients in beds on day d of that cycle.

- To calculate the number of patients in beds in ward k on day d of the cycle resulting from surgeon i having a slot in theatre t on day m of the previous cycle,

we require $j = C + d - m$ giving $B_{i,t,(C+d-m),k}$. This applies to any day m of the previous cycle.

- Similarly for the contribution from day d, n cycles previously; we calculate the number of patients in beds in ward k on day d of the cycle resulting from surgeon i having a slot in theatre t on day m of that cycle, then $j = nC + d - m$ giving $B_{i,t,(nC+d-m),k}$.

- As patients can stay in hospital for longer than the cycle length, to obtain the total contribution to the number of patients in beds in ward k on day d of the cycle resulting from surgeon i having a slot in theatre t on day m, it is necessary to sum the contributions from all of the previous cycles.

- The maximum length of stay to be considered is J so only values of n such that $nC + d - m \leq J$ should be considered. Similarly operations from future days do not contribute so $nC + d - m \geq 0$. Rearranging these equations gives the set N of values for n as set out above.

- Therefore, the sum $\sum_{n \in N} B_{i,t,(nC+d-m),k}$ calculates the number of patients in beds in ward k on day d of the cycle resulting from surgeon i having a slot in theatre t on day m of the cycle.

- The contribution to number of patients in beds in ward k on day d of surgeon i having a slot in theatre t on day m is only $\sum_{n \in N} B_{i,t,(nC+d-m),k}$ if surgeon i is assigned slot in theatre t on day m. The summation $\sum_{i,s,t} \sum_{m=1}^{C} \left(X_{i,t,m,s} \sum_{n \in N} B_{i,t,(nC+d-m),k} \right)$ ensures that the contribution is only counted if the slot has been assigned as $X_{i,t,m,s}$ is 1 in this case and 0 otherwise. Summing over i, s, t and m ensures that the contribution from every scheduled theatre slot is included.

- Therefore, constraint (17) sets the value of $\mu_{d,k}$ to be the expected number of beds required in ward k on day d.

Constraint (18) then sets the value of Z_k to be the minimum difference between the number of beds available on day d in ward k and $\mu_{d,k}$.

4.3.7 Mathematical Formulation of the Objective

Maximise

$$\sum_{k} \alpha_k Z_k + \beta \sum_{i,t,d,s} P_{i,t,d,s} X_{i,t,d,s} - \chi \sum_{i,t,d,s} Y_{i,t,d,s} + \delta \sum_{i,t,d,s} U_{i,t,d,s} - \varepsilon \sum_{i,t,d,s} V_{i,t,d,s} + \gamma \sum_{i,t,d,s} W_{i,t,d,s}$$

α_k, β, χ, δ, ε, γ are weighting values and can be adjusted by the user to reflect the values

of their hospital and/or to explore how the balance effects the suggested timetable. This

is to be solved subject to constraints 1 to 18.

$\sum_{k} \alpha_k Z_k$ is the sum of the weighted minimum difference between the number of beds

available and the number required for each ward k, so by maximising this we make the

tightest difference between beds used and beds available as large as possible, and it is a

measure of how well the bed usage is smoothed in the timetable. These are weighted

separately to allow ward size and the importance of particular wards (such as ICU) to be

taken into account.

$\sum_{i,t,d,s} P_{i,t,d,s} X_{i,t,d,s}$ is the sum of the preference scores for the slots that have been assigned,

so it is a measure of how well the preferences of surgeons are met by the timetable.

$\sum_{i,t,d,s} Y_{i,t,d,s}$ is the number of times any surgeons are assigned to slots for which they have

low preference scores. This allows additional penalisation of assigning a surgeon to a

slot they consider undesirable. Without this constraint, a timetable where some surgeons

have slots for which they have very high preference scores and others have slots for

which they have very low preference scores could give a high score for the sum of the

preference scores and appear to be meeting the preferences of surgeons when it is not

doing so fairly.

$\sum_{i,t,d,s} U_{i,t,d,s}$ is the number of times that any surgeon is assigned to consecutive slots in the

same theatre on the same day, and measures how often this occurs in the timetable.

$\sum_{i,t,d,s} V_{i,t,d,s}$ is the number of times that any surgeon is assigned to consecutive slots in different theatres on the same day, and measures how often this occurs in the timetable.

$\sum_{i,t,d,s} W_{i,t,d,s}$ is the number of times that any surgeon is assigned to repeat the same slot in the same theatre at the interval specified for such repeats (usually weekly), and measures how often this occurs in the timetable.

Thus, we have defined the problem and formulated it as a multi-objective MIP, the next stage is to implement this with both real and randomly generated data and evaluate its performance.

4.4 Implementation

This section explains how the formulation developed in the previous section is implemented. Firstly the data requirements are discussed, along with how the data was prepared. This is followed by consideration of the software to be used and how the data is entered into the software. The validation and verification of the model is then discussed. Lastly, the reasons for not considering the stochastic elements of the problem are considered.

4.4.1 Data

The hospital staff working on the timetable would know the values to enter for the majority of the data that is required. This information for our example case is obtained from the interviews with staff. This includes the cycle length required, the repeat frequency that is desirable (weekly), the number of theatres, the types of those theatres, the maximum number of theatre slots available on any day in the schedule, the availability of the theatres, the number of surgeons, the availability of equipment/resources that needs to be considered at this level of scheduling, the maximum number of days that patients spend in beds to be used for modelling purposes, the number of wards and the maximum number of beds available in each ward.

Other information such as the surgeons' availability and preferences requires consultation with individual surgeons and in our tests a mixture of random sample data and data devised to test specific aspects of the model is used for these values.

We use the existing hospital timetable to give the number of slots in each type of theatre required by each of the surgeons. This information would usually come from the decisions made within the hospital in their strategic planning.

The values of $B_{i,t,j,k}$ the expected number of patients still in beds used in ward k, j days after being operated on by surgeon i in a slot in theatre t, are less straight forward to assign, but are obtained by collating the hospital records of cases carried out. They are calculated based on the historical number of patients in the beds of each ward on the days following operations by each surgeon in theatre t, if no such data exists then the data from a theatre of the same type as t or the type most similar to t.

4.4.2 MIP Software

Of the software available for solving MIP problems, FICO[TM] Xpress Optimization Suite is used for this case. This software was selected because the author already had access to a licence for the full version with the ability to handle problems of this size. Had it not proved capable of dealing with the problem effectively then we would have investigated using alternative software.

4.4.3 User Interface Spreadsheet

Hospital staff members are not expected to be familiar with optimisations programs like Xpress. Also entering the data directly into the arrays Xpress works with would be difficult to do accurately, since there is such a large quantity of data to load. However, hospital staff members in general are familiar with Excel and as setting out the data entry for the arrays in a spreadsheet makes it much easier to see what data goes where, Excel is used for the date entry for the model. Excel is also used as the front end of Blake and Donald's (2002) model, for similar reasons.

Due to the volume of data, the data entry is split over several spreadsheets. The main spreadsheet contains all of the general data that is the same for all surgeons as well as controls for loading the data and setting parameters for the optimisation. The worksheets for data entry and running the model contained in this spreadsheet is as follows.

- Controls: This worksheet contains the command buttons for implementing the VBA code as discussed in Section 4.4.4.1 below.
- General: This spreadsheet is for data relating to the cycle length, number of slots per day and the range of surgeons preference scores that the surgeons sheets will accept.
- Theatres and Types: This sheet is for data relating to the possible types of the theatres as well as the theatres and their sets of types.
- Theatre availability: Contains a table for the data on which theatres are available for which slots over the cycle.
- Resources: This sheet is for the data on which equipment or other resources are to be considered in the model and the numbers of each that are available.
- Surgeons: This sheet contains space for the data on the surgeons, their specialities, the maximum number of slots that they can work in a day and where their individual data entry sheets are stored.
- Beds: For data relating to the wards to be considered and the numbers of beds available in them.
- Weightings: This sheet is for the weightings to be applied to the objective function.
- Solver Setup: Is for setting the max runtime of the model, so that it will return the best solution found after a set amount of time, if the optimal has not been found at that stage.
- Quick update: This allows the user to update a selection of the data, from this workbook, which is faster than reloading everything including the data in the surgeons' workbooks.

Figure 3 illustrates an example of the sheet 'Theatres and Types' and demonstrates the type of layout used throughout the spreadsheet.

Theatre Descriptions

Number of Types of Theatres [6] (Calculated from types entered)

Number of Theatres [11] (Calculated from the theatres entered)

Types of Theatres

Please enter the names of the theatre types under theatre type

Theatre Type	Type Number
TC	1
TC laminar flow	2
Main	3
Main Laminar flow	4
Gynae	5
Andover	6

Theatres / Theatre Types

Theatre Name	Theatre Number	TC (1)	TC laminar (2)	Main (3)	Main Lami (4)	Gynae (5)
A	1	1	0	0		
B	2	1	0	0		
C	3	1	1	0		
1	4			1		
2	5			1		
3	6				1	
4	7				1	
5	8			1		
Andover1	9					1
Heathcote	10			0		
Andover2	11					

Figure 3: Example of the Theatres and Types worksheet

In addition to facilitating the input of data the main spreadsheet also contains sheets for displaying the results of the optimisation; this is done via the following worksheets;

- Suggested Timetable: Gives the timetable produced by the solver as a grid showing which surgeon is assigned to each theatre for each slot.
- Timetable Analysis: Gives details of how the timetable performs in relation to each of the objectives and a table of the expected bed requirements if the timetable is implemented.
- Beds Chart: Gives a graphical representation of the number of beds required compared with the number available over the course of the cycle of the timetable.

The spreadsheet also contains some workbooks used to store data to facilitate the quick update process and the raw results from the optimisation.

Figure 4 illustrates an example of the worksheet that gives the analysis of the suggested timetable.

Figure 4: Example of the Timetable Analysis worksheet

There are separate workbooks for the loading of each surgeon's data, where their availability, preference scores, theatre and equipment requirements and expected bed usage can all be entered. These contain the following worksheets;

- General: Reads across the data on the surgeon from the 'Surgeons' worksheet of the main spreadsheet, and has a data entry table for the number of slots that the surgeon requires in each type of theatre.

- Preferences: Contains a table for the entry of the surgeon's preferences for theatre slots on each day of the cycle, by slot and theatre.

- Availability: Contains a table for the entry of the surgeon's availability for theatre slots on each day of the cycle, by slot and theatre.

- Resources: Contains a table for entry of the resources that are required by theatre type for the surgeon.

- Beds: Contains a table for the entry of the numbers of beds expected to be required on different days after the surgeon has had a theatre slot in each theatre.

4.4.4 Inputting and Examining the Data

4.4.4.1 Loading data

Visual basic for applications (VBA) coding is used to load all of the data from the spreadsheets, format it and save it as a text document ready for use as input by Xpress. Similar code is also used to read data from an output file produced by Xpress, to bring the results back into Excel and display them to the user in the output worksheets described above.

89

4.4.4.2 Validating data

To reduce the scope for error in entering the data, Excel's validation function has been used on the majority of the data entry cells, to specify the type of data that can be entered. For example, on the surgeons' availability worksheets only binary values can be entered for the availability, the data is also assessed in the VBA code to identify potential issues so that they can be corrected before being entered into the model.

4.4.4.3 Basic feasibility testing

Given the complexity of the problem and that hospitals are expected to be using their theatres at close to capacity (close to infeasibility), it is possible that the problem entered will be infeasible.

The problem of diagnosing the cause of infeasibility in MIP and IP has been studied by a number of researchers; Chinneck and Greenberg have both produced a number of papers on the subject for example Chinneck (1997, 2001), Greenberg (1988), Greenberg and Murphy (1991) and Guieu and Chinneck (1999), all discuss aspects of the topic. The methods they describe for diagnosing infeasibility or subsets of constraints that are infeasible all require considerable knowledge of MIP/IP and the methods used to solve such problems. As the intention is that the model will be used by hospital staff with limited if any knowledge of these areas, it would be unreasonable to expect them to be able to interpret the results of the type of analysis conducted in these studies. Therefore, a more basic approach to resolving infeasibility is required.

Basic feasibility testing has been built into the VBA code that uploads the data from the spreadsheet and formats it for the solver. This will raise an appropriate error message to the user if the following arise;

- The values entered for any of the data are of the wrong type, e.g. non-binary numbers where binary is expected, preference scores higher than the maximum allowed etc.
- The number of sessions that a surgeon is available to operate in is less than the number of slots they are to be assigned overall.
- The number of sessions that a surgeon is available to operate in the theatres of each type is less than the number of slots they are to be assigned for that theatre type.

- The overall theatre availability is exceeded by the total demand for theatre slots.
- The theatre availability is exceeded by the total demand for theatre slots, by set of theatre types.

These are quick checks to identify the more obvious ways in which a problem could be infeasible, thus avoiding the user wasting time trying to run the solver on a problem that has a data entry error or is infeasible for straightforward reasons.

There are numerous more complex ways in which the problem could be infeasible, for example a subset of surgeons, may have similar, limited availability and therefore require the same set of theatre types to be used more often than is possible over a small period of time.

To enable users to find feasible solutions, we recommend that they firstly ensure that the current timetable would be feasible for the data that they enter (with any additional surgery to be added available for scheduling, either in currently empty slots or in the slots of that are not required in the new timetable). This ensures that a feasible solution will be found. Analysing the resulting timetable, and adding in the additional constraints that are required, should enable the user to identify compromises to their original problem that will yield a feasible timetable.

4.5 Verification and Validation of the Model

"One of the most important and difficult tasks facing a model developer is the verification and validation of the ... model." Banks et al. (1999)

This section describes the methods of verification and validation that have been applied in order to ensure that our model works as intended and is a sufficiently accurate to suggest timetables taking into account the objectives set out in Section 4.3.1.

4.5.1 Verifying the model

In the following description of model verification Anderson et al. (2003) are referring to the verification of a simulation model, although the goal is the same for verifying any model;

"… the process of determining that the computer procedure that performs the simulation calculations is logically correct. Verification is largely a debugging task to make sure that no errors are in the computer procedure that implements the simulation." (Anderson et al., 2003)

In summary, verification is checking that the model performs correctly and as intended by the modeller.

In the case of our theatre timetabling model verification involves testing that the constraints and objectives are constructed to perform as expected. This is done using a very simple example of the problem, with a three day cycle, two theatres and two surgeons. Having run the model on the basic example, small changes are made to the data and the model is rerun to check that the resulting changes to the timetable are as expected. For example checking that increasing the preference score a surgeon has for a particular slot above all of the other preference scores (calculated to ensure it will outweigh the other objectives and there is a feasible timetable with that surgeon assigned to that slot) results in a timetable where the surgeon is assigned that particular slot.

This process is conducted rigorously for all of the constraints and the model does behave as intended.

The VBA codes to upload the data and downloading the results to the spreadsheet were also verified by making small changes to the data and checking that these were correctly reflected in the data loaded and the way the results were set out.

4.5.2 Validating the model
Again in the following Anderson et al. (2003) are referring to the validation of a simulation model, although the goal is the same for validating any model;

"… the process of ensuring that the simulation model provides an accurate representation of a real system. Validation requires an agreement among analysts and managers that the logic and the assumptions used in the design of the simulation model accurately reflect how the real system operates." Anderson et al. (2003)

For the master surgical timetable the partner hospital has an existing timetable, so the conditions can be set so that only the existing timetable is feasible and the values of the objectives can be checked against the real values. This is achieved by limiting the availability of surgeons to only their slots in the existing timetable, with their demand by theatre type set to the number of slots of each type that they have in the current timetable, thus forcing the selection of the current timetable. As expected this produces the current timetable, with surgeons assigned their current slots, and the objective function value as expected. Thus, the model is validated against the real-life system.

4.6 Stochastic Considerations

In Sections 2.5 and 4.1 the stochastic elements of the problem are raised as important factors in theatre scheduling, but this is not addressed in the formulation set out above.

For the master surgical timetable, the most significant area of variability in the results is around the number of beds required. This arises from both variability in the length of time patients stay in beds after their operations and in the number of patients each surgeon treats in a theatre slot, which in turn is results from the stochastic nature of demand. Combining the effects of these two factors over all theatre slots would require excessive amounts of calculation and thus considerably increase computational time, making the use of the model less attractive to hospital managers

Given that the timetable will be repeated throughout the year in a cyclic fashion, over the year the bed usage will average out, but not considering the stochastic nature of the problem means that no consideration is made of the variation around the average. However, as no consideration of the effects of the surgical timetable on bed usage is currently made in producing the timetable, employing the average usage will be a considerable improvement on the current situation.

The most significant argument for not considering the stochastic nature of the problem at this level is that once a master surgical timetable is selected and implemented the variations in length of stay can be considered at the day-to-day scheduling level, when the combinations of individual patients to book into slots are considered. At this level the expected number of patients for each slot will be known and the stochastic elements of the problem can be much more effectively considered. Ideally, the master timetable

would be implemented using the recommended booking strategy from Chapter 6 and then rerun with the data collected on bed usage from the theatre slots using both results to investigate the potential for further improvements.

4.7 Improving the Formulation

This section discusses how the formulation of the problem was investigated and adapted by the addition of cutting planes to reduce computation time.

4.7.1 Exploring the Linear Relaxation

In order to gain further insight into the running of the model the values of variables in the solution to the LP relaxation of the MIP can be considered. In this case many of the values for binary variables in the solution to the LP relaxation of the problem are 0.5. Therefore, additional constraints to bring the solution to the LP relaxation closer to having binary variable values where relevant will assist with finding a solution to the original MIP problem more efficiently.

The variables $U_{i,t,d,s}$ and $W_{i,t,d,s}$ have positive coefficients in the objective function, so the optimisation process will be looking to make their values as big as possible. The upper values of these variables are limited by constraints (3), (6) and (7) and these make use of the fact that these are binary variables to set the values of $U_{i,t,d,s}$ and $W_{i,t,d,s}$ correctly from the value of $X_{i,t,d,s}$. In the linear relaxation of the problem non-binary values of $U_{i,t,d,s}$ and $W_{i,t,d,s}$ are permitted so these constraints are not having the intended effects.

If $X_{i,t,d,s}$ is 0 then both $U_{i,t,d,s}$ and $W_{i,t,d,s}$ should also be zero, as if a surgeon is not operating in a slot in a particular theatre then they can't be operating in that slot in that theatre and the one after it in the same theatre or in the same slot and theatre with the desired repeat frequency in the timetable, but there are examples in the solutions to LP relaxations where this is not the case.

It is reasonable to expect that adding constraints that reduce the feasible region for the LP relaxation, but not the feasible region of the original problem would speed up the solution process as the solution to the LP relaxation should be closer to the solution of the MIP problem. Thus, the following constraints can be added to the problem.

94

Tighter constraints from 3;

$$X_{t,t,d,s} \geq U_{t,t,d,s} \qquad \forall t,t,d,s<S \tag{3a}$$

$$X_{t,t,d,s+1} \geq U_{t,t,d,s} \qquad \forall t,t,d,s<S \tag{3b}$$

Tighter constraints from 6;

$$X_{t,t,d,s} \geq W_{t,t,d,s} \qquad \forall t,t,s,d>w \tag{6a}$$

$$X_{t,t,d+w,s} \geq W_{t,t,d,s} \qquad \forall t,t,s,d>w \tag{6b}$$

Tighter constraints from 7;

$$X_{t,t,d,s} \geq W_{t,t,d,s} \qquad \forall t,t,s,d\leq w \tag{7a}$$

$$X_{t,t,d+w,s} \geq W_{t,t,d,s} \qquad \forall t,t,s,d\leq w \tag{7b}$$

This does in fact increase the speed at which the solver optimises the MIP, with the solution gap (between the best bound and best solution) after 100 seconds going from 9.44% without the extra constraints, down to 4.04% with them. A similar improvement is seen after 200 seconds with the solution gap going down from 6.27% to 3.99% when the constraints are added (all of these results were obtained on the same computer with the same settings).

Therefore, the constraints set out in this section are included in the formulation used to obtain the results set out in the remainder of this chapter.

4.8 Results

This section gives the results obtained from the model, compared with the current timetable, to demonstrate the improvements that can be achieved. The example used is for a hospital with 11 operating theatres, which can be split into 6 sets of types, 54 surgeons, with up to 3 slots per day, 9 equipment types to consider and we are looking for a 2 weekly repeated cycle. The information on bed requirements by ward was not available so we treat all of the beds available as belonging to one ward and use a maximum length of stay of 30 days. This results in a problem with 124755 variables, 124740 of which are binary variables, 15 non-negativity constraints and 138366 other constraints.

As mentioned in Section 4.4.4, the model has been run with the surgeons availability set to ensure that the current timetable is produced, this means that the data on the current

timetable has been collected in the same manner as the data on the suggested timetable, allowing them to be easily compared.

The suggested timetable discussed in this section is based on the assumption that all of the surgeons are available all of the time. In reality this is not the case, but as full information on surgeon's availability and preferences is not available (due to the time and effort as well as raising of expectations that would be involved in collecting this data when the hospital does not intend to change the timetable in the near future) any other version would have required further assumptions and potentially taken us further from reality. It is in any case interesting to see how much the usage of beds could be smoothed if the surgeons where all available for all slots.

In looking at the results, it should be remembered that this kind of improvement could not be expected for a real problem due to the restrictions on the availability of surgeons. It is hoped that in future when the partner hospital wishes to change the theatre timetable, it will be worth while collecting data on surgeons availability and preferences and using the model to assist in the development of the new timetable.

Table **1** and Figure 5 give the comparison of the original timetable with the timetable suggested by the MIP when run in Xpress-IVE.

Table 1: Comparison of objectives for current and suggested timetables

Indicator	Original Timetable	MIP Timetable
Max beds used	90	83
No. Surgeons changing theatres	0	0
No. of all day slots	30	55
Repeat weekly	166	178

Comparing current and suggested timetables bed usage.
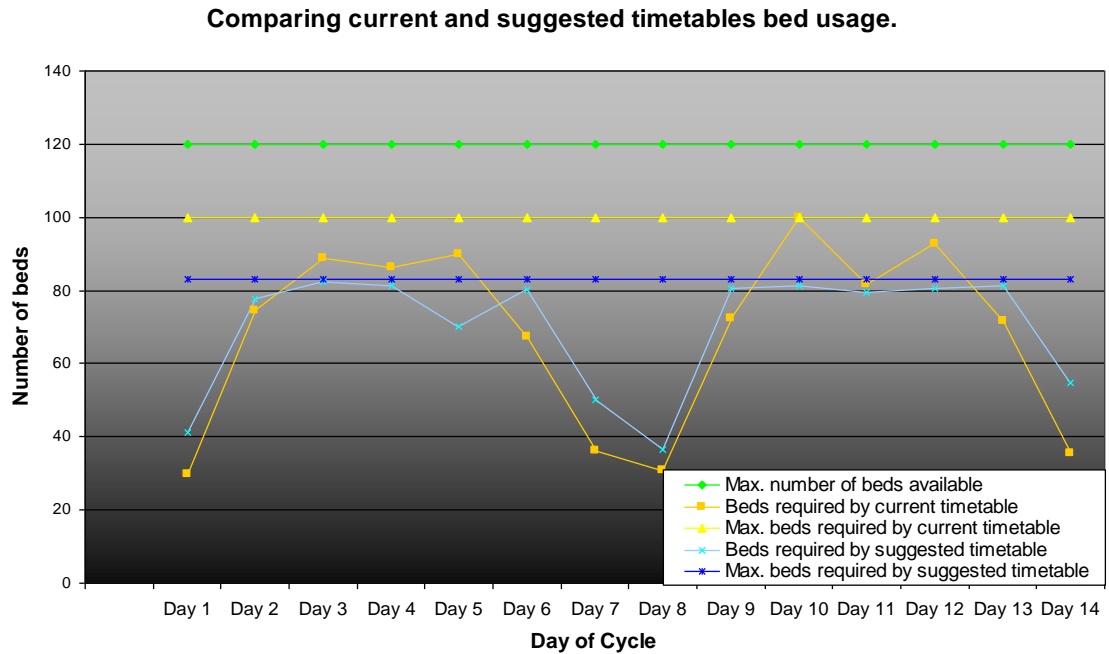
Figure 5: Graph to illustrate the bed smoothing achieved by the model compared with the existing timetable.

These results demonstrate that considerable improvements in the smoothing of bed usage, numbers of all day slots and numbers of slots repeated weekly can be achieved. The extent to which such improvements can in fact be achieved will depend on how limited surgeons' availability is. The bed usage shown in the graph are the expected averages for each day of the 14 day scheduling cycle, so increasing the gap between these and the number of beds available increases the flexibility for the variations around the averages.

4.9 Sensitivity Analysis

This section discusses what is involved in sensitivity analysis before going on to show how this applies to the linear programming formulation for the master surgical timetable.

In linear programming, sensitivity analysis involves calculating the extent to which the problem can change before the optimal solution changes. This is particularly useful if there is uncertainty around the values of the data for the objective or constraints. For linear programming, standard calculations can be used to perform sensitivity analysis (Winston, 1994).

Sensitivity analysis for problems involving integers is more complex as the calculations based on the linear programming sensitivity analysis only apply if the solution to the LP relaxation of the IP is integer.

For this problem there are some inputs for which it is straightforward to predict that the solution will be particularly sensitive to changes in their values. For example if a surgeons availability changes for a theatre slot to which they are assigned, in the optimal timetable then that timetable will no longer be feasible. However, it is much harder to predict if a surgeon becoming available for a slot for which they were previously unavailable would change the optimal solution, as it depends if that surgeon operating in that particular slot would provide a better solution.

The following describes our consideration of the effects of the changes felt to be most significant to the problem of finding good master surgical timetables.

4.9.1 Sensitivity to Objective Function Weightings

For this particular problem it is desirable that changes to the weights in the objective function result in changes to the optimal solution. This is so that users can quickly obtain solutions for different objectives so that they can compare the effects on the timetable produced. Therefore, this section explores the sensitivity of the solution to changes in the objective function coefficients.

Table 2 gives the results of running the model with identical problems, except that the objective function weightings are adjusted slightly each time.

As only one ward is considered in the example we use a single value of α for these tests. The relative sensitivity to the weightings for different wards would depend on the relative sizes of those wards.

Changes to the value of β are not considered as preference scores have not been entered for this data set.

Table 2: Exploring the effects on the results of small changes in the weightings used in the objective function, all of the other data entered was identical each time.

Weightings used					Results				
α	χ	δ	ε	γ	Difference between beds available and max beds used	All day slots	Changing theatres	Weekly repeats	Low pref. scores
5	5	5	5	10	9.69	4	0	162	1
6	5	5	5	10	22.09	5	0	159	1
5	6	5	5	10	9.62	4	0	162	1
5	5	6	5	10	9.43	4	0	158	1
5	5	5	6	10	9.69	4	0	162	1
5	5	5	5	5	11.02	4	0	156	0
5	20	5	5	10	10.75	14	0	155	1

This table illustrates that just small changes to the weights are required to produce different timetable suggestions, so the solution values are sensitive to the weightings used. Note that changes to the value of ε are not resulting in changes to the results as none of the timetables are including any surgeons operating in different theatres in consecutive slots on the same day.

4.9.2 Flexibility in the Surgeons' availability

Initially, we did not have access to information on the availability of surgeons, so the model was run as if they are all available all of the time. This is not only far from the case in real life; it also took so long to run the model that the computer ran out of memory, before reaching a solution. Therefore, in order to explore the effects on solution time and how good a solution (based on the gap between the best bound and the best solution after 10 minutes) could be obtained trials with random data have been run.

To do this random data samples, all with a two week timetable, in a hospital with 11 theatres, of 6 sets of type of theatre, 54 surgeons, up to 3 slots per day and 9 equipment types to consider were generated. Each of these problems will have in access of 100,000 variables and 100,000 constraints. In order to ensure that the problems would be feasible, a random assignment of surgeons to slots is generated and then the availability

of surgeons is set to ensure that they are available for the slots assigned on that timetable and a percentage of the rest of the time. To explore the effects of different limitations on surgeons' availability the model was run for five instances each time with the percentage of the time (other than the feasibility check timetable) that surgeons were available set to each of 100%, 90%, 75%, 50% and 25%.

Figure 6 and Figure 7 are scatter graphs showing how the solution gap and time vary with the availability of surgeons. They both demonstrate that as surgeon's availability decreases the best available solutions under the new constraints are found faster. This is to be expected because as surgeons' availability becomes more limited there are fewer possible solutions to explore in the attempt to find an optimal solution.
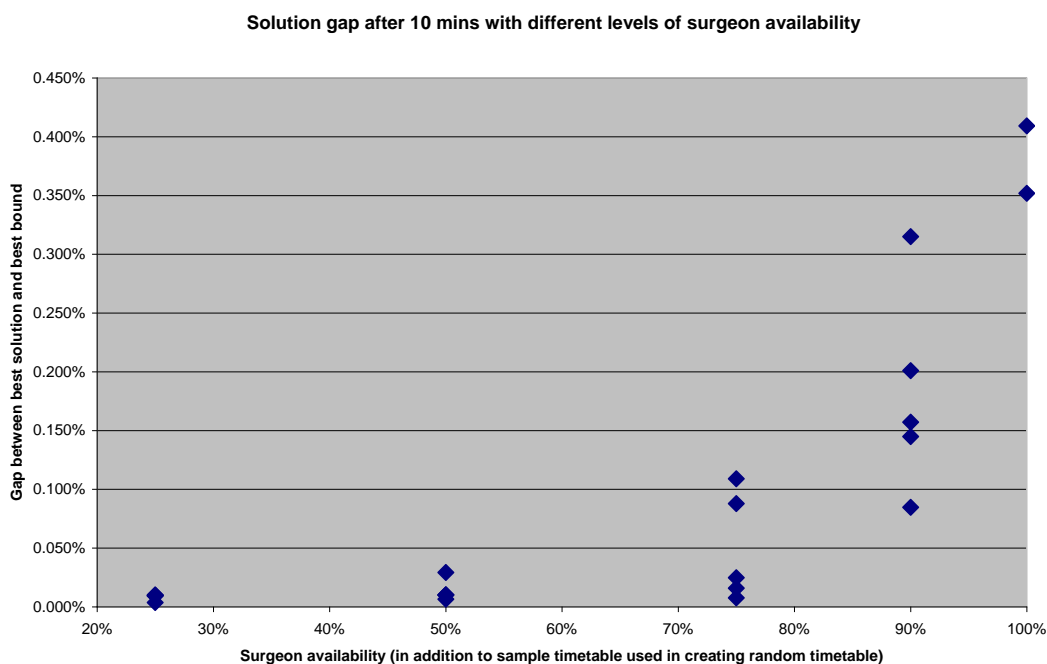
Figure 6: How the solution gap varies with the availability of surgeons

Some of the values are too close to each other to be distinguished in the graph: there are 5 points marked for each level of surgeon availability.

Number of seconds to reach optimal solution (if reached within 600 secs) with different levels of surgeon availability
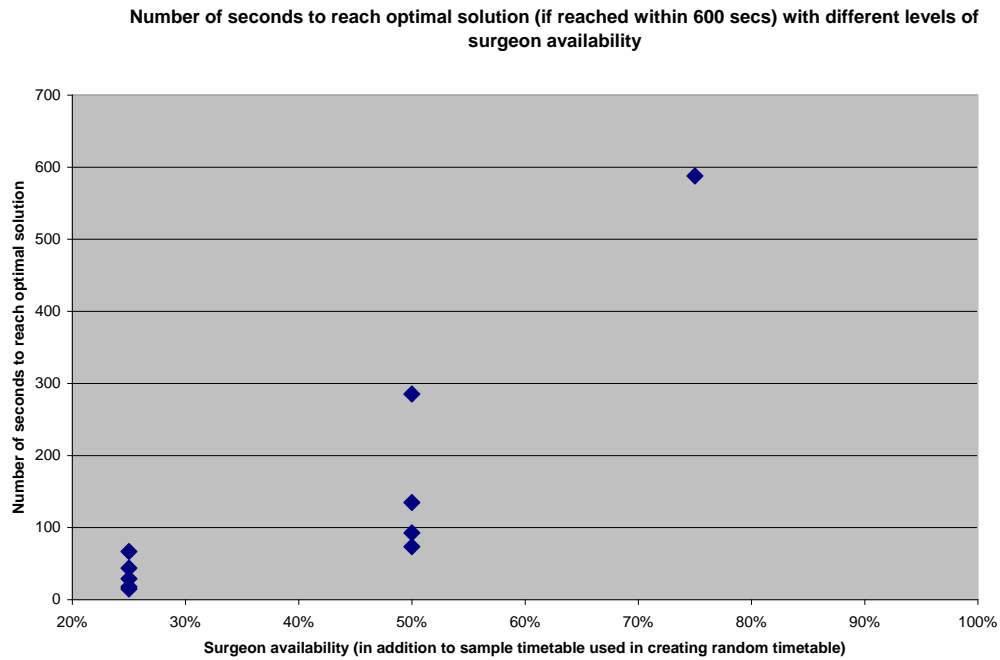
Figure 7: How the solution time varies with the availability of surgeons, where optimal solutions were found.

In reality the availability of surgeons is quite constrained as they have outpatient clinics and other tasks that have to be conducted at specific times. For example, some surgeons work at other hospitals at certain points in the week. Therefore, it is useful that better solutions are found faster with more limited surgeons' availability as this means that solution times for real life problems will be reasonable.

4.10 General Flexibility and Limitations

The previous section has shown that the model can be adapted to incorporate variations in the availability of surgeons. This could be particularly useful in demonstrating how much difference changes in particular surgeon's availability makes to possible timetables, which would facilitate discussion on changing availability.

The model will not directly suggest where changes to surgeons' availability would help to improve the timetable. Although comparing the timetable produced with no limitations on surgeons' availability with that where the surgeon's availability constraints are included would suggest changes to try.

In fact, if adjustments to the surgeons' outpatient clinics where also being considered the clinic spaces could be considered as theatres in the system and the outpatient and surgical slots could be allocated simultaneously.

As discussed in Section 4.4.4.3 it is possible that changes to the availability could result in the problem becoming infeasible, hence the recommendation to ensure that the current timetable with any additional sessions fitted into it is feasible initially. Additional constraints on availability can then be incorporated gradually allowing the user to identify any that make the problem infeasible. This addresses the limitations on the extent to which the model can identify the constraints that make problems infeasible.

The ability to vary the weightings of the objectives (see Section 4.9.1) allows users to adapt the model to their priorities and to generate and compare different timetables based on variations around these priorities.

The data entry process is structured so that the numbers of surgeons, theatres, slots per day, wards and beds can be adapted to reflect the structure of any size of hospital. The other features like theatre types and equipment considered are also fully adjustable. The use of an Excel interface ensures that the data entry process is accessible to hospital staff so that the full extent of its flexibility can be taken into account.

The ability to include limited availability of equipment in the model allows not only equipment, but also other staff limitations to be taken into account. It is possible that there are other limitations on the timetable that another hospital would want to include that cannot be taken into account by the model. For example, we have not considered the possibility that some equipment can only be used once on any day, this is mentioned in the literature review in Section 2.4.9. However, this level of detail is better addressed at the day-to-day scheduling level when it would be possible to consider if a particular patient's treatment would require particular equipment.

The most significant limitation of the model is that it does not take account of variations in the case mix of each surgeon from week to week and the variations in patients' lengths of stay. Users should be particularly aware of this when considering the expected bed usage, as the values given are averages and the actual usage will vary.

Even so this inclusion of bed usage is a significant step from previous models which have not considered it when constructing master theatre timetables.

Any limitations on the accuracy of the data for predicting bed usage, particularly when new surgeons are starting and the data would need to be estimated, should also be considered when interpreting the results.

4.11 Conclusions

This chapter illustrates that it is possible to solve the MIP for the problem of finding master surgical timetables in a reasonable amount of time using a standard MIP solver. Therefore, no further work is planned to apply column generation or heuristics.

The stochastic nature of some aspects of the problem has not been incorporated at this level as it is felt that this can be addressed more effectively in the day-to-day scheduling process. All of the other factors identified in Section 4.1.3 have been included in the formulation, which as far as we are aware has not been done before. The specific elements of this model that are new contributions compared with what is done elsewhere in the literature are;

- The consideration of the availability of beds; particularly that this is done at ward level and in the tactical rather than day-to-day scheduling.
- The inclusion of the expected contribution to bed usage of surgeons having particular slots is new, as is allowing the number of beds available to vary over the course of the repeat cycle for the timetable.
- In the literature different theatre types are rarely considered, so our inclusion of not only different theatre types, but also the way in which sets of theatres can meet different demand is novel.
- Consideration of the availability of surgeons is limited in the literature and the ability to consider their preferences is new.
- The ability to allow for the availability of other resources, such as staff and equipment, by using the equipment availability function is new.
- The inclusion of weekly (or other) repeats in a cycle of longer than one week is unusual, as is allowing for the need to avoid the same surgeon having consecutive slots in different theatres.

It did not prove possible to implement this model within a local hospital within the timescale of this project, due to the pressures on hospital staff. However, it is hoped that such implementation will take place in due course.

It is also possible to consider scheduling surgeons' outpatient clinics at the same time as their slots in theatre, if the clinic spaces are treated as theatres in the model; this would increase the availability of surgeons and should allow further improvements in the objective function.

It would also be interesting to investigate the improvements that could be achieved with a truly dynamic timetable.

Chapter 5: Booking individual Patients

This chapter describes the exploration of scheduling rules for booking individual patients; starting with the definition of this aspect of theatre scheduling, going onto explore the implications of both the healthcare scheduling literature and more general scheduling literature. The chapter concludes with discussion of methodology for exploring scheduling algorithms including providing a general model, the focus of Chapter 6 is then to populate and test the model with exemplar data.

5.1 Defining the Problem

The advanced booking of surgery for elective patients, involves scheduling individual patients into the available theatre slots. Patients are not available for booking until the decision has been made to treat them and they require treatment within a certain time period, based on either clinical urgency or waiting time targets such as those in the UK. The underlying problem is similar to an online machine scheduling problem with due dates. There is variability in time remaining until due date as well as inter arrival and service times.

Patients may be urgent or routine based on clinical assessment of their need for surgery. As the names suggest urgent patients need to be given priority and have target treatment dates closer to the dates on which the decisions are made to treat them. For routine patients the target treatment dates are the NHS target of referral to treatment in 18 weeks, less the time already elapsed since referral when the decision that surgery is required is made. This creates two types of priority base on clinical need and the administrative priority of meeting the 18 week target. The importance placed on not breaching the 18 week target means that a patient who is close to breaching it may be treated as being as urgent as someone whose surgery is clinically urgent. This consideration is made clear in the theatre scheduling literature and is based on strategic level decisions regarding the urgency of patients and acceptable waiting times.

There are a number of other factors to consider as well, so that the problem of booking patients into theatre slots involves:
- predicting surgery time;
- maximising utilisation of theatre (and surgeons) time;

- meeting waiting time targets;

- allowing slots for more urgent patients, than those currently under consideration;

- being fair in terms of the time waited by similar patients;

- avoiding overtime – including allowing for emergencies (this could also be seen as minimising cancelations depending on the mix of use of overtime and cancellation of surgery that occurs when a slot is over booked);

- patient's preferences (for example to avoid pre-booked holidays);

- considering the length of time for which beds will be needed after surgery;

- availability of the equipment required;

- suitability of the mix of cases for each theatre slot (allowing for surgeons' preferences).

In developing the algorithm used to book patients into theatre slots, other factors to consider are:

- how far into the future to have theatre slots open for accepting bookings;

- whether to book patients as soon it has been decided that they need surgery or nearer to their due dates;

- whether to book patients in batches or not;

- if batches are used, how big should they be.

As discussed in the literature review (Chapter 2), targets to book patients as soon as the decision that an operation was needed were introduced in 2002. This resulted in some discussion in the British Medical Journal with Gallivan et al. (2002a, 2002b) describing how this would inevitably reduce efficiency, while Rogers et al. (2002) disagreed. Since then this target has been dropped, however, it would be in the interests of patients to let them know the time of their operations well in advance and this feeds into the considerations listed above.

Currently there is variation in the methods used, not just between different hospitals and specialties, but even within teams of booking clerks working for surgeons in the same specialties. The scheduling is generally done by hand, with some hospitals still using paper diaries and cards with patient details. Thus, there is considerable scope for improvement.

5.1.1 Online verses offline scheduling

The distinction between online and offline scheduling is important in this chapter. Online scheduling involves scheduling jobs (in our case patients) as they arrive into the system overtime, without knowledge of what jobs will arrive later. Offline scheduling involves having information on all of the jobs to be scheduled at the point when scheduling takes place. Our problem as defined above is an online scheduling problem as the patients will arrive overtime and the scheduler will only have information on each patient when they arrive. In discussing the literature we will include studies that use both online and offline scheduling in order to avoid missing potentially useful methods discussed in the offline scheduling literature.

5.2 The literature

The literature review (Chapter 2) has already considered the coverage of day-to-day scheduling in the literature on theatre scheduling. This section briefly recaps this before going on to discuss relevant literature from other areas of scheduling.

5.2.1 Theatre Scheduling Literature

Simulation is used in the majority of studies that assess strategies for the advanced scheduling of patients. Dexter et al. (2000) use simulation to explore strategies for scheduling cases into 'overflow' time. Sciomachen et al. (2005) use simulation to apply scheduling rules to the whole of the advanced scheduling process, scheduling first by longest waiting time, then longest processing time and finally by shortest processing time. They also test scenarios around the use of a master surgical schedule and introducing a recovery room. The results from comparing scheduling rules indicate that the best rule to use depends on the objectives in terms of reducing overruns or total overtime.

Dexter and Traub (2002) also use simulation to compare scheduling rules based on scheduling patients to the earliest or latest start times available. Van Houdenhoven et al. (2007) apply the bin-packing problem algorithms, Best Fit Descending heuristic and Regret-Based Random Sampling again testing the models with simulation. Dexter et al. (1999) also apply bin packing algorithms and test them with simulation, but their focus is slightly different as they are looking at adding additional cases once the schedule has been planned, rather than general on-going scheduling.

All of these examples are testing and comparing different scheduling policies that could be applied by hospitals for the advanced online scheduling of patients. This demonstrates how effective simulation can be for comparing methods, but also its weakness; simulation only compares the methods considered, it does not suggest alternate methods or when exceptions should be made to the rules.

Gerchak et al. (1996) apply stochastic dynamic programming to the advance scheduling problem; they focus on the need to allow unscheduled time for a variable number of emergency cases. This type of mathematical programming technique finds the best solution to the problem of where to schedule cases, given the information available, but is only suitable for offline scheduling.

Guinet and Chaabane (2003) also formulate the problem as a mathematical program formulation, although they conclude that it is NP hard and give heuristic methods which find fast solutions. They assume that the cases to be booked over the next two weeks are known when the problem is solved, but patients are generally booked several weeks in advance with the expectation that their surgery dates will not be rescheduled. This demonstrates the limitation of optimisation methods compared with heuristics is that they are only suitable for offline scheduling; in contrast heuristics allow online scheduling by following the heuristic rules.

In summary, for the advanced day-to-day scheduling problem, a significant number of studies use simulation to find good heuristic rules for scheduling, while a few others use optimisation techniques to find optimal solutions each time further cases are scheduled.

5.2.2 More General Scheduling Literature

As discussed in the introduction to this chapter the problem of scheduling individual patients for surgery is a specific example of the wider field of machine scheduling, if we consider the operating theatres to be the machines and the patients jobs to be processed on those machines. Thus, considering the methods used in the literature on general scheduling may provide useful insights.

The literature relevant to this problem includes a number of studies on appointment scheduling, as well as more general scheduling, so these two areas are considered separately below.

5.2.2.1 Appointment scheduling literature

Scheduling patients for surgery is similar to making appointments for any type of service, so literature on appointment scheduling should be considered when looking for inspiration to improve the process of booking patients. The majority of papers relating to appointments scheduling refer to other aspects of the health system, for example outpatient or GP appointments. It would seem that managing bookings from waiting lists is a significant problem across many aspects of the health care sector and that is why so many applications arise in this area. This section will briefly consider appointment scheduling, to investigate whether the literature on this topic has useful insights for the problem of booking individual patients.

Appointment systems in healthcare have been considered by operational researchers since the 1960's. Jackson (1964) advocates the use of appointment systems in hospitals and general practice, to replace systems where patients arrived and then waited to be served. This reduces the time patients spend waiting, but if the doctor gets ahead of schedule then their time can be wasted waiting for the next patient, so making best use of the doctors' time should also be considered. This is addressed by Welch (1964), who introduced what is now known as the Welch rule, that if 2 patients are booked into the first slot, then the doctor will be slightly behind and have patients available if one does not arrive at some point in that block of appointments.

This balance of reducing patients waiting times, but avoiding doctors idle time within the order of the schedule on a given day is the main theme of appointment scheduling literature relating to healthcare. Klassen and Rohleder (2000) apply dynamic programming to analyse scheduling rules for this aspect of scheduling and they also apply simulation to the same problem (Klassen and Rohleder, 1996, 2002). Brahimi and Worthington (1991) use queuing models to suggest scheduling rules at this level. Ho and Lau (1999) also use simulation to explore the effects of changes in the length of appointments in the schedule given the variation in treatment times. Jerbi and Kamoun (2011) consider this aspect of scheduling using goal programming. This is just a sample

of the studies considering the structure of the day aspect of scheduling, from the point of view of balancing reducing patient waiting time and avoiding doctors' idle time. Even this small sample demonstrates that a range of methods have been used to consider this aspect of scheduling.

In surgical scheduling, this balance is less important, as surgeons see all patients before the start of the surgical session, so they must all arrive at the same starting time and wait to be seen. This is because the time taken to leave surgery, see individual patients as they arrived and return to sterile clothes in between operations would be detrimental to efficient use of operating theatres.

Klassen and Rohleder (1996) suggest booking the types of patients whose treatment times have lower standard deviations (i.e. are more predictable) first in the schedule. This may be useful in operating theatres, because staff in the rest of the hospital can be given a better idea of when the next patient needs to be ready. However, as the order of patients on the day of surgery is not decided at the time of booking, this is not important to the problem considered in this chapter.

Tai and Williams (2011) consider adaptations to the daily schedule to allow for patient arriving late, but again this is less relevant to our problem as all patients are expected to arrive in good time before the start of the session.

Su and Shih (2003) consider managing an appointment system where 'walk-ins' are expected in addition to the patients already booked for the session. This is similar to the arrival of emergency patients requiring surgery, but they are concerned with when gaps in the daily schedule should be left rather than the amount of time needed to allow for emergencies.

Patrick et al. (2008) consider scheduling with patients with multiple priority ratings. They assign time in each slot to different priority levels, which works well when a reasonable number of patients are treated in each slot, but is not so effective in theatre scheduling where one case may take up a large proportion of a theatre session.

Bowers (2010) and Gupta and Denton (2008) consider waiting lists and appointment scheduling in healthcare more broadly. Bowers (2010) points out that queues of the sizes found in healthcare act "as a buffer absorbing much of the stochastic variation". Therefore, getting rid of queues entirely would not be desirable as then the amount of treatment time available would need to match the variation in arrival rates.

Bowers (2010) also states that the some simulation models for hospital management are "essentially based on the assumption of a first in first out (FIFO) queue discipline. The FIFO assumption is reasonable in most simulations of systems involving inert entities but it may be less appropriate for healthcare simulations in which patients have very different priorities reflecting their individual needs." They recommend systems that consider the urgency of patients and prioritise accordingly. This suggests that exploration of booking rules should include prioritisation of urgent patients and that the extent to which this is achieved can be compared with the results of a FIFO scheduling rule.

Gupta and Denton (2008) also stress the importance of "reserve capacity for urgent appointment requests" in healthcare scheduling, confirming that it is important that any booking system allows space for the arrival of urgent patients. At the same time the system should aim to "realize high utilization of more-expensive specialists' time". This is the same balance of patient waiting time, compared to avoiding wasting doctors time considered in outpatient appointment ordering above. In this case, it is allowing space to have short waiting times for urgent patients, without allowing under booking of surgeons' time. Gupta and Denton (2008) also comment on the lack of implementation of studies on appointment scheduling, which is similar to that in theatre scheduling as discussed in Section 2.6.

While a large number of studies dealing with appointment systems consider healthcare related problems, there are studies considering appointment systems in general. Creemers and Lambrecht (2009) describe how queuing models can be used to analyse the expected waiting times and size of waiting list under varying conditions. This is particularly relevant in the service industry where companies need to consider how many servers to provide to deal with their customers. However, it is less relevant to the

surgical booking problem as we are trying to manage the waiting list with given resources, rather than consider the resources required.

The above discussion indicates that aspects of the problems considered by a number of the studies in the appointment scheduling literature are closely enough related to the problem under consideration in this chapter to provide useful insights. They suggest that it is worth considering how a range of algorithms compare to FIFO in their ability to deal with a mix of priority levels among patients and that similar issues arise in appointment scheduling in general and theatres scheduling in particular.

5.2.2.2 Introduction to General Scheduling Literature

As with the discussion of appointment scheduling above, this is not intended to be a thorough review of general scheduling literature. The objective is to consider studies in areas of scheduling with similar features to the problem of booking individual patients for surgery, to see what implications their results have for the booking problem.

Booking patients with different expected operation durations into theatre sessions of fixed duration is effectively a bin packing problem in the terminology used in general literature. Patients arriving over time requiring booking is effectively the same as objects requiring scheduling online, rather than all of the objects to be scheduled being known at the start of the process. Our problem also has due dates, by which patients should ideally be treated, so in scheduling terms the problem is an online bin packing problem with due dates and we are also concerned with the fairness with which patients are treated. Thus, the literature on dealing with problems concerning some or all of these aspects of scheduling may contain relevant methods and this is explored below, along with discussion of the methods used in scheduling literature.

5.2.2.3 Bin Packing Problems in the Literature

Bin packing problems are discussed by Chen et al. (1998) as approximations for scheduling problems. This, along with similarity of the need to fill operating theatre slots to classical bin packing problems, suggests that it would be worth considering the algorithms used to solve bin packing problems. The main difference from day-to-day scheduling is that all of the items to be packed are general known in advance. The evidence to 1998 suggests that the best algorithms for bin packing problems are variations on first-fit decreasing (FFD), which involves ordering the items in non-

increasing order of size and then packing them into the first available bin (Chen et al., 1998). This suggests that prioritising patients with longer expected durations should be explored. Wong and Lee (2009) consider two dimensional bin packing, and as for FFD the algorithms they consider start by ordering the objects in non-increasing order of length. Then they place them in the bin with the most space available, with algorithms differing over how the most space is determined. This suggests that booking into the available theatre slot with the most remaining space should be considered.

Shi and Ye (2008) consider online bin packing, where they refer to first fit, any fit and best fit algorithms, suggesting that exploration of these as rules for choosing a viable theatre slot should be considered. While they are considering online bin packing, it is not bins that close over time with jobs arriving over time, but bins that are continuously available, with items that must be packed into the bins after their release time. Thus, the algorithms they prove to be most effective are not relevant to booking patients.

While we have not identified an online bin packing studies considering due dates this brief discussion of bin packing has yielded some potential algorithms to consider.

5.2.2.4 Literature Concern Scheduling with Due Dates

In the scheduling literature, two classes of problem that deal with due dates are relevant, involving minimising the number of late jobs or minimizing the sum of the tardiness (the amounts by which due dates are exceeded) (Chen et al. 1998). In theatre scheduling we care both about the number of patients who are treated late and about the extent to which they are treated late, so we discuss both below.

For the problem of minimising the number of late jobs, Chen et al. (1998) describe algorithms by Moore (1968), which involve arranging jobs in order of due date. Then scheduling in this order and when one does not fit within its due date removing one of the jobs (e.g. with the longest processing time) to put at the end of the schedule. Scheduling jobs in order of due date is possible for hospital theatre scheduling when more than one job is scheduled at once and should therefore be considered. However, the extent to which jobs are late does matter, just moving patients who are going to be late to be very late is not acceptable, so it would not be sensible to test this strategy as part of an algorithm.

113

For minimising the sum of the tardiness Cheng et al. (2005) and Kolliopoulos and Steiner (2007) both provide complex algorithms using dynamic programming for offline scheduling with due dates, using relatively complex orderings of jobs before scheduling them. Both of their methods work well offline, but without knowledge of all the jobs to be scheduled they are not applicable to online scheduling.

Schmidt (1988) considered the problem of scheduling tasks with deadlines, starting with all tasks having identical deadlines; in which case sorting the tasks into non-decreasing order of processing time before booking is recommended. When the tasks have groups of deadlines, sorting within each group in the same way is recommended, before booking the groups in order. For booking patients the deadlines are grouped by day, so they could be booked in order of deadline in this way.

Huegler and Vasko (1997) recommend using swaps of jobs in a simulated annealing algorithm to improve on solutions obtained from simple heuristics. Angel and Bampis (2005) also use search methods, this time based on dynamic programming, to improve solutions. These work well for offline scheduling. However, for theatre scheduling where patients are booked firmly in an online problem, such search methods are not feasible.

Thus, generally the methods used for scheduling with deadlines make use of the knowledge of all of the jobs to be scheduled in finding a schedule. As this knowledge is not available for online scheduling such methods are unsuitable for our problem.

5.2.2.5 Literature concerning online scheduling

Thus far we have found some simple heuristic methods that could be applied to the problem of scheduling individual patients, although many of the methods used in the scheduling literature are not suitable for online scheduling. This section provides an overview of the methods that have been used for online scheduling.

Pruhs et al. (2004) review of online scheduling and list the following standard algorithms found in the literature:

- Shortest remaining processing time – runs the job with the least remaining work. This is not suitable for theatre scheduling as each job must be completed in one go and not interrupted.
- First in first out – schedule all jobs in arrival order, this would be feasible for booking patients.
- Shortest job first (or shortest processing time – SPT) – schedules the jobs in non-deceasing order of processing time. For booking patients that would equate to scheduling in non-decreasing order of expected operation duration.
- Highest density first – if the jobs have weights, then schedule them in weight order. This could be applied to booking patients by taking the due dates as weights.

The other algorithms they suggest assume that jobs can be interrupted, as for shortest remaining processing time, and as such are unsuitable when booking patients for surgery.

There is general agreement in the literature on online scheduling that delayed versions of SPT work best (Lu et al., 2003, Anderson and Potts (2004), Liu et al., 2009, Potts and Strusevich, 2009, Tao et al., 2010 and Liu et al. 2011b). Such algorithms involve delaying the scheduling decision until all of the jobs to be scheduled over a certain horizon are known, or as late as possible, in order to obtain the best schedule. This does not fit with theatre bookings having a target of scheduling all patients as it is decided that they require surgery. However, it does suggest that the effects of this target should be explored.

None of these algorithms have the bin packing aspect of theatre scheduling where the theatre runs for chunks of time rather than continuously. The bin packing literature suggests arranging items in non-increasing size order before placing them in bins, as do Khammuang et al. (2007) who do consider a different online scheduling problem with a bin packing element to it.

Thus, the extent of the effect of this bin packing element will determine whether it is better to book a group of patients in non-increasing or non-decreasing order of expected operation duration and this should be explored.

Liu et al. (2009) is the only study we have identified that considers due dates in an online scheduling problem and it uses similar methods to the online problems without due dates.

5.2.2.6 Literature concerning fairness in scheduling

As we are concerned with the fairness of the scheduling of theatre slots it is also worth considering the literature on fairness in general scheduling problems. This is a relatively new field of research and arises from concerns that while algorithms like SPT can do very well in terms of mean response time they may be unfair to jobs with large processing times (Wiermann, 2011).

A considerable proportion of the literature concerning fairness in scheduling relates to the transfer of data across computer networks; such as wireless networks (Kim and Han, 2005 and Bu et al. 2006) and the internet (Kelly et al. 1998). This area of literature does make some recommendations similar to those in the general scheduling literature: for example Zaharia et al. (2010) recommend delaying scheduling as much as possible to produce fairer schedules, the same type of recommendation that arises in the literature on on-line scheduling above. There is also a suggestion of grouping the items being scheduled based on their characteristics and then comparing how fairly the groups are treated relative to each other (Greenberg, 1992). This idea may be helpful in reducing the amount of computation required to undertake in order to evaluate the fairness of a theatres scheduling algorithm.

Generally the literature on fairness in data transfer is considering a rather different problem to theatre scheduling, because in a data transfer parts of two or more jobs could be undertaken simultaneously and pre-emption can occur, that is a job may be interrupted to allow another job to be worked on. Neither of these can occur within an operating theatre. Therefore, while the literature on fairness in data transfer and similar situations does provide some insights the rest of this section will focus on scheduling that does not involve data flows. This still includes studies that allow pre-emption and are thus considering a problem which while rather different from what is possible in theatre scheduling may still offer useful insights; for example Schwiegelshohn and Yahyapour (2000) also suggest considering groups of jobs when testing for fairness.

Another variation on fairness in scheduling involves long term fairness where the same jobs require repeated scheduling. This is particularly apparent in the carpooling problem, where a different selection from a group of travel together regularly and wish to share out the driving fairly (Ajtai et al., 1998). As very few patients return frequently for operations this is less relevant to our problem, although in areas of a hospital like radiography where some patients require repeated treatment it could prove useful.

Wiermann's (2011) review of fairness in relation to scheduling single server queues is relevant to our study as the study considers scheduling jobs of a variety of sizes, for a single machine. This is similar to scheduling patients for a single operating theatre, although no account is made of due dates.

Wiermann (2011) defines 3 types of fairness;
1. 'That it is more fair to serve jobs in the order that they arrive' which they refer to as temporal fairness.
2. 'It is quite acceptable to allow small jobs to bypass big jobs' so the small jobs do not have to wait for the big jobs, which they refer to as proportional fairness.
3. 'It is more fair to serve the more urgent jobs, regardless of job sizes or arrival order'.

The last of these is by far the most relevant to operating theatre scheduling, but unfortunately it is the one Wiermann (2011) does not consider. However, his insights into fairness in scheduling are still relevant.

With regard to proportional fairness he groups the jobs by size and considers whether or not each job size is treated fairly, by considering the expected waiting time for each group. This type of fairness is also considered by Sandmann (2011) and Avi-Itzhak et al. (2007), both of whom refer to it as 'slowdown fairness' because the aim is to be fair to jobs with longer processing times in terms of the extent to which they are slowed down by waiting for shorter jobs.

Wiermann (2011) demonstrates that first come first served (FCFS) is always unfair under this measure, while shortest remaining processing time (SRPT) is sometimes proportionally fair depending on the service distributions. SRPT assumes that pre-

emption is permitted; the closest to this that is possible in operating theatre scheduling is SPT. The other algorithms they discuss for proportional fairness assume pre-emption is permitted and are therefore not considered here.

Wiermann (2011) goes on to discuss 'proportional fairness in expectation' that is not only should algorithms be fair to each group of jobs but also within each group. This stems from concerns that larger jobs may experience a wider variation in waiting times. Thus, if we do compare groups of patients, the variation within groups should be considered as well as the average waiting times.

In terms of temporal fairness, Wiermann (2011) defines the 'politeness experienced by a job of size x … is the fraction of the response time of a job of size x during which the seniority of the job is respected'. FCFS, while being always unfair under slowdown fairness, is always polite under this measure as is jump to front (JTF) which is based on FCFS but allows jobs to jump to the front of the queue with a set probability.

Thus, a policy can be always unfair under one definition whilst also being always fair under the other. This suggests that to find a policy which is fair under both measures will be complex if indeed it is feasible.

Raz et al. (2004) study this problem using queueing theory and devise a resource allocation queueing fairness measure, which combines both temporal and priority fairness, based on the proportion of server capacity given to each job. Wiermann (2011) and Sandmann (2011) propose a simpler measure of combined fairness, which involves counting the number of times a job is discriminated against. This is done by the summation of the number of jobs that arrive no earlier than and are completed no later than the job under consideration and the number of remaining jobs with a size no smaller than that in question when it arrives and are complete no later than it. A calculation similar to this could be used to consider fairness with respect to due dates for hospital theatre scheduling.

In this section we have seen that while none of the studies relating to fairness in scheduling consider online scheduling with due dates, they still provide insights into what scheduling policies are fair and how to assess fairness. With regard to fair policies,

FCFS and SPT should both be considered. In terms of judging the fairness of algorithms, comparing the waiting times of different groups of patients could prove useful as could consideration of some form of discrimination measure.

5.2.2.7 Methods used in the scheduling literature

This section briefly reviews the types of method used in the scheduling literature with a view to identifying those that could be used in our study of the advanced scheduling of individual patients.

A significant proportion of scheduling problems are known to be generally NP-complete optimisation problems, meaning that they cannot necessarily be solved to optimality in a reasonable amount of time (Chen et al. 1998, Cheng et al, 2005, Lin, 2007 and Zhang et al. 2008b). Branch and bound and dynamic programming techniques are often applied to such problems (Huegler and Vasko, 1997, Chen et al. 1998, Angel and Bampis, 2005 and Cheng et al., 2005).

A range of simple heuristics is used, such as SPT and variations on it (Schmidt, 1988, Liu et al, 2009 and Wong and Lee, 2009). Simulation can be used to test these methods. Metaheuristics such as simulated annealing or tabu search are also used, for example by Zhang et al. (2008b).

In order to apply techniques like branch and bound, dynamic programming and metaheuristics, it is necessary to have knowledge of all of the jobs that require scheduling. In online scheduling such as our theatre scheduling problem, this is not the case. That leaves simple heuristics to be tested and a number of such have been suggested in the literature discussed above.

5.2.3 Discussion and Implications of Literature

Throughout this section a variety of scheduling algorithms have been suggested, which should be considered in our search to improve the process of booking patients for surgery. FIFO has come up in several sections and should therefore be considered. The literature on bin packing problems suggests algorithms such as booking into the slot for which the patients expected theatre time is the best fit to the remaining unbooked theatre time. It also suggests booking in decreasing order of expected treatment time, while the

literature concerning due dates and online scheduling recommend increasing order of expected treatment time. The literature on due dates recommends arranging jobs in increasing due date order before scheduling them. Also, the online scheduling literature recommends delaying the decisions as much as possible.

Some of the above could be used to complement each other, for example using a second rule for tie breaks, while others are completely contradictory. Thus, a thorough consideration of a range of algorithms is required.

The scheduling literature also suggests the use of more optimisation methods that apply in some cases and more complex heuristics. However, as discussed in Section 5.2.2.7 these are not suitable for online scheduling and will therefore not be considered further in this document.

Simulation is used for testing booking rules in a number of studies in the theatre scheduling literature relating to bookings, which suggests it would be appropriate for testing the types of algorithms identified here.

5.3 Methodology

5.3.1 Method Selection

As discussed above, the majority of studies considering this type of problem have used simulation. This is a sensible place to start because simulation:

- provides a visual representation of the problem, allowing clients to see how the model reflects reality;
- allows a variety of scenarios, in this case booking algorithms, to be tested in a short period of time;
- incorporates variability, such as that associated with arrivals, service times and time remaining to due date when a patient becomes available for surgery.

Given the apparent lack of implementation of academic work in this area the first of these points is particularly significant: if hospital staff can see the differences made by different algorithms on the schedules created, they are more likely to be convinced to adopt the recommended algorithms.

However, simulation can only test the algorithms that are given to it and does not make recommendations about generating completely different algorithms. The methods given in the scheduling literature described above are used as inspiration for the algorithms to be tested using the simulation model.

5.3.2 Algorithms to Consider

The list below gives a variety of algorithms that could be considered for each type of decision required in the advanced scheduling of individual patients;

1. At what stage to make bookings
    a. Book each patient as soon as it is known that they require treatment, which fits with the target described in the definition of the problem (Section 5.1) for booking patients, and gives patients the maximum possible notice of their operation date.
    b. Wait until the end of the day/week and book all patients together, which creates batches and allows some knowledge of other patients who require booking at the same time.
    c. Only book if the rest of algorithm determines that the booking should be made within the next x weeks of the current time, where the value of x remains constant throughout. This is inspired by the success of delay algorithms in dealing with online problems as illustrated in Sections 5.2.2.5 and 5.2.2.6 above. Varying the value of x allows exploration of the balance between the desirability of giving patients sufficient notice of their operation date and the scheduling improvements gained by delaying booking decisions.
2. What booking priority rule to use (this will only have an effect if there is more than one patient being considered for booking at a given time, as it determines the order in which patients are considered by the algorithm);
    a. First come first served/ First in first out – fair in terms of waiting times, but does not take account of varying urgency of patients.
    b. Last come first served – this is particularly unfair in terms of waiting times, but has some applications in more general scheduling.
    c. Most urgent first – this will ensure that urgency is considered, but if there are a large number of urgent patients, routine patients could wait an unacceptably long time.

d. In order of due date – this means that more urgent patients will be considered before routine patients arriving at the same time as them, and it also effectively makes routine patients become more urgent as their due dates approach.

e. In decreasing order of operation duration – inspired by the longest processing time algorithms discussed in Section 5.2.2.4.

f. A variation on due dates, using operation duration as a tie breaker.

g. A variation on operation duration, incorporating due dates as a tie breaker.

3. The selection of the order in which potential dates are searched and which feasible date is selected;

a. First available slot – book each patient into the earliest slot with sufficient unallocated theatre time for the expected duration of their operation. Given the bin packing aspect of the problem this is as close as possible to the FIFO scheduling rule (see Section 5.2.2.1) when combined with 2a) above.

b. Book to due date – book each patient into the first suitable slot moving back from their due date. If no such slot exists, then select the first available slot after their due date. This algorithm is based on a suggestion from the surgeon that he has considered booking in this way, it has the advantage of leaving spaces for the arrival of more urgent cases after the patient currently under consideration has been booked.

c. Exact fit – find a slot which the expected duration of the patients operation will fill exactly the available theatre time, taking account of patients already booked. If no such slot exists, then another rule will be required in addition. If more than one such slot exists, then the patient can be booked;

    i. As soon as possible

    ii. As late as possible within their due date

    iii. Randomly

d. Most empty slot – book the patient into the slot with the most available theatre time. If several slots have joint most time available, then the same rules as in c above could be used to decide between them. This is inspired by the scheduling algorithm discussed in Section 5.2.2.3 above.

To generate a booking algorithm a selection must be made from the stages at which to book given in 1; the booking priorities given in 2 (if 1a was not selected); and the method of choosing the booking date given in 3. It is possible to combine different rules, with groups of patients treated according to different rules or those booked within different time frames booked according to different rules. This is particularly relevant to 3c, where an alternative rule is required if there is no slot that the expected duration will fill exactly.

Following testing of the a limited number of combinations, it is possible to consider the reasons behind their effects and thus assess which combinations should be tested rather than testing all possible combinations, as the latter would be rather time consuming.

The straightforward performance measures against which these algorithms can be judged include;
- the percentage of patients (of each type) seen within their due dates;
- the percentage of available theatre time used, including allowing for turnaround time between operations;
- the percentage utilisation of theatres (the percentage of slots spent operating not including turnaround time as this is considered an important performance measure in hospitals);
- the number of overruns greater than allowable overtime;
- the number of cancellations or other rearrangements of operations.

These will assess the ability of the algorithms to respect due dates and use theatres efficiently.

Given that some of these algorithms involve booking patients as late as possible and the need to book urgent patients earlier than routine patients; it is also desirable to measure the following to test how fairly the algorithms treat patients:
- The expected waiting time for patients of varying urgencies, to check that more urgent patients are indeed seen faster.
- That patients with similar needs are booked in order, so a once a patient is booked another patient with a later due date and similar operation time is not

booked before them. This is testing for the first type of fairness discussed in Section 5.2.2.6.

5.4 Simulation Model

5.4.1 The Model

Figure 8, shows the setup of the simulation model. The system itself is quite simple; however the complexity comes in the visual logic used to implement the booking algorithms at the 'Booking' workstation.
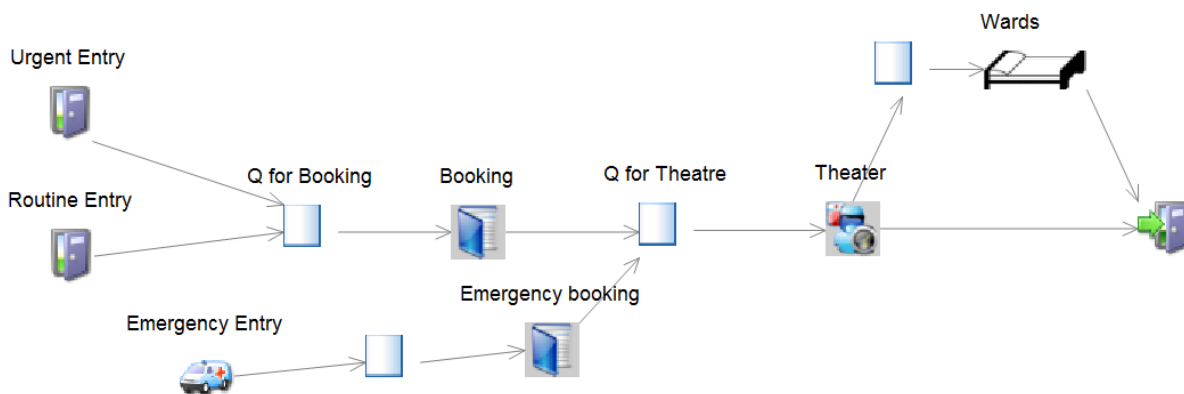
Figure 8: The simulation model

The model is set up as follows:

1. Depending on the urgency of their surgery patients enter via the 'Urgent Entry' or 'Routine Entry' when the decision that they require surgery has been taken;
2. The inter arrival times are sampled from the distributions given in the data;
3. The due dates of the patients are set at the entry points based on distributions from the data;
4. They go to the 'Q for Booking' where the depending on the algorithm in use they may be held for a certain amount of time and/or sorted by due date or expected operation duration;
5. The 'Booking' workstation uses visual logic to implement the algorithm and book the patients into a spreadsheet used to record bookings;
6. The entry process for emergencies is similar, although they are only held in a queue if another emergency is in the process of being booked and the booking rules are different as overtime may be used if necessary;
7. Once patients have been booked, they move to the queue for theatre, which only releases them to the theatre once their booked date has arrived;

124

8. The 'Theatre' workstation represents the operating theatre and applies a variation to the duration of the operation, which is sampled form a distribution of the difference between predicted and actual operation durations, it also records various data about the patient;

9. The model is also able to deal with the impact of the schedule on bed usage, by sending those patients who require an overnight stay to wards and those without an overnight stay directly to the exit.

Following discussion with theatre staff it was decided that once a patient was booked, their theatre slot could not be rearranged. This is because cancellations and re-bookings are particularly disruptive for patients, so the hospital would avoid them as much as possible. Also, allowing re-arrangement of patients might disguise how well or otherwise an algorithm is doing at creating a good schedule.

It is necessary to have at least a small amount of time between the decision to book a patient and their surgery taking place, to allow them to consider what is going to happen and give informed consent. The amount of time required will vary with the circumstances as emergency patients need to be treated quickly, for other patients we allow a week before we consider a slot to be suitable.

The algorithms' results are also be influenced by the amount of time considered available for booking each day, which needs to take account of the amount of time expected to be required for emergency patients out of the actual length of the theatre slots. Unless otherwise stated, the expected duration of the average number of emergencies per slot will be withheld from the time allowed for booking other patients while the rest of the theatre slot is available for booking.

5.5 Discussion

This chapter has briefly explored the aspects of the general literature on scheduling that are most relevant to the day-to-day theatre scheduling problem and used this to inform the suggestion of a range of algorithms for theatre scheduling. It has also proposed a simulation model for testing the algorithms. In order to conduct such testing, the simulation model needs to be populated with realistic hospital data, which will be undertaken in Chapter 6.

Chapter 6: Booking individual Patients: Case Study

This chapter follows on from the theoretical consideration of Chapter 5, in the form of a case study using a specific ophthalmologist's surgical data. Consideration is given to the ability to generalise these results by approximately fitting distributions to the data and comparing the results of the best booking algorithms identified so far, with these distributions input into the model and variations from some of them.

We start by considering the adaptions required to the simulation model proposed in Chapter 5 and discuss other aspects of the model set up. Followed by, discussion of the data preparation, for both the initial tests with empirical data and the fitting of algorithms for the variation from initial data. Then the results are set out and conclusions drawn from them.

6.1 Case Study to the Simulation Model Adaptions and Detailed Modelling

While this model was in development Stephen Lash, a local ophthalmic surgeon, contacted the University wanting to explore how he could improve his booking of surgical patients. Therefore, the model is tested using his data as a case study to explore the potential of different algorithms to improve the number of patients seen within their target times.

6.1.1 Adaptions to the model

As the cases treated by this surgeon are all (virtually all) treated as daycases, that is the patients do not stay overnight in hospital, there is no need to consider the effects of the schedule on wards so this part of the model is not needed for the specific case study, as the data does not contain any patients who stay overnight.

Lash's patients split more naturally into those requiring cataract surgery and those requiring Vitreoretinal (VR) surgery, than just urgent and routine. This is because the two types of patient have different arrival and treatment time distributions (as shown below in Section 6.3). So rather than having separate entry points for urgent and routine patients the model is adapted to have entry points for Cataract and VR patients, with

labels used to assign these patients as either urgent or routine based on the time remaining to their due dates.

Also there are surgical slots on at most two days per week, one on Tuesday mornings and the other on alternate Thursday afternoons, with these slots having different durations and overtime limits. In order to model this more accurately, a second theatre and queue are added to the model along with an entry for emergency patients going to those theatre sessions. In order to only have emergency patients arrive on the alternate Thursdays when there are slots, those who arrive on other Thursdays are routed straight out of the system.

The final addition to the basic model is a dummy queue and workstation above the queue for booking, which creates the possibility of running algorithms that look through all of those in the queue in priority order each day (by keeping those who have been checked in the dummy queue so that they don't go to the start of the queue for booking and become stuck in a loop). These adaptions to the model are illustrated in Figure 9.
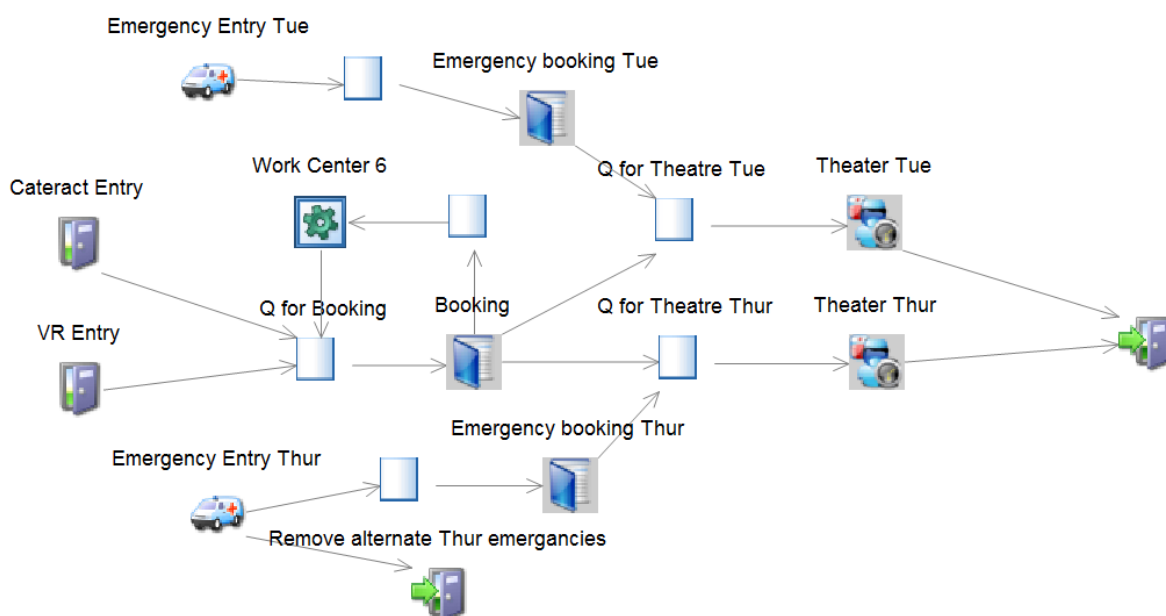
Figure 9: The adapted simulation model

6.1.2 Detail behind the model

The list below describes the actions taken by the settings and visual logic at each stage of the model;

1. Cataract Entry – Arrivals of cataract patients occur following a specified distribution. As each patient arrives they are assigned:

    a. A label containing a unique number (and the unique number counter is increased by one ready for the next patient).

    b. A label containing their arrival time.

    c. Labels (both numeric and text) to specify that they are a cataract patient

    d. A due data is created for each patient by adding to their arrival time a value sampled from the distribution of time remaining to due date when a cataract patient arrives.

    e. Based on the time remaining to the due date, a patient is assigned as either urgent or routine patients a numeric label with urgent patients assigned a value of 1 and routine patients a value of 2.

    f. An expected duration sampled from the data on the duration of cataract operations.

    g. An emergency tag indicating that they are not an emergency case.

    h. A booking priority value equal to their due date minus a number obtained by dividing the expected duration by a large number (the later part is to act as a tie break between those with the same due date). If there is still a tie then the patient who arrived first will be considered for booking first.

2. VR Entry – Exactly the same as for cataract entry except that there are separate distributions for VR patients throughout, and the relevant labels specify that they are a VR patient.

3. Queue for Booking – This queue holds all of the patients waiting to be considered by the Booking workstation. Unless otherwise specified by the algorithm they are released in order of the priority code described in 1h above (the reasons for this are discussed with the algorithms below in Section 6.4).

4. Booking – At this workstation, the Visual Logic relating to the relevant booking algorithm is implemented. If the algorithm books the patient, then they are routed to the queue for the theatre in which they have been booked; if not they are routed to the dummy queue above the booking workstation, where they will be held until the end of the day, as described above in Section 6.1.2 to prevent

the same patient getting stuck looping between the booking queue and workstation. Once the patient has been booked, data on the booking is recorded into the relevant spreadsheet for Tuesdays or Thursdays.

5. Workstation 6 – This is a dummy workstation and merely returns patients to the queue for booking when they have waited in the dummy queue described above.

6. Queue for Theatre Tuesday – This queue holds the patients that have been booked into theatre slots on Tuesdays and only releases to the workstation representing Tuesday's Theatre slot on the day that they have been booked into, and if the theatre is not currently busy.

7. Queue of Theatre Thursday – as for Tuesday but with those scheduled on Thursdays.

8. Theatre Tuesday – The operation duration is adjusted based on the distribution of variation between predicted and actual operation durations and the turnaround time is sampled from the turnaround time distribution. These figures are added together to give the time taken in theatre. The visual logic at this stage undertakes a number of steps to record information about the time the patient has waited to arrive at theatre.

    a. If the patient is an emergency, then the count of emergency patients treated is increased by one; the patients waiting time is added to the sum of emergency waiting times; if the patient is being seen within their due date then the number of emergencies seen in target is increased by one; if the patient has waited longer than the value recorded as the longest emergency wait then this is replaced with their waiting time.

    b. If the patient is a cataract patient, then the count of cataract patients treated is increased by one; the patients waiting time is added to the sum of cataract waiting times; if the patient is being seen within their due date then the number of cataract patients seen in target is increased by one; if the patient has waited longer than the value recorded as the longest cataract wait then this is replaced with their waiting time. For cataract patients, the same data as above is recorded in a spreadsheet called Fairness Testing, separated into different rows depending on the number of weeks to their target when they arrived.

    c. The same data is recorded with appropriate headings for VR patients as for cataract patients.

      d. The same data is recorded for urgent and routine patients as for emergency patients, with appropriate changes to the headings.

      e. Information on the duration of the operation and turnaround time are added to the spreadsheet storing information about patients treated on Tuesdays.

      f. A separate spreadsheet containing information about treated patients has the following data added to it;

          i. Unique number of the patient

          ii. Type of patient (cataract or VR)

          iii. The effective referral date of the patient

          iv. The due date of the patient

          v. Arrival time

          vi. Emergency tag (if the case was an emergency)

          vii. Expected operation duration

          viii. Actual operation duration

          ix. Time when they are treated

          x. Time spent waiting

          xi. Turnaround time at the start of their surgery

          xii. Booking priority

          xiii. The amount of time they had to their target when they arrived at the system

          xiv. An indicator storing whether they were seen before their due date

9. Theatre Thursday - Operates in the same way as for Tuesday, with a separate spreadsheet for recording patient information.

10. Once patients have been treated in a theatre, they got to the exit point and leave the model.

11. Emergency Entry Tuesday – Is very similar to the entry points for cataract and VR patients except that the type is set to emergency and no other level of urgency is set.

12. Emergency Entry Thursday – Is as for Tuesday, with the major difference being that as operations only occur every other Thursday if the simulation time at arrival rounded to the nearest week is even, then the patients are routed straight to a dummy exit.

13. The queues following the emergency entry points are dummy queues, which will only be needed if emergency patients arrive at the same time and one waits for the Booking workstation to book the other.

14. Emergency Booking Tuesday – Books each emergency patient to be treated on the day of arrival and sends them to queue of the Tuesday Theatre so that they can be treated with the other patients that day.

15. Emergency Booking Thursday – Works in the same way for Thursday's emergencies.

16. At the end of the warm up time, the following values are reset so that they only include data from the results collection period of the simulation;

    a. All of the data described in 8f and 9 (above) for individual patients.
    b. The numbers of each type of patient treated.
    c. The numbers of each type or patient treated by their due dates.
    d. The sums of waiting times for each type of patient.
    e. The maximum waiting times for each type of patient.
    f. The 'fairness testing' data mentioned in 8b and c above.

17. At the end of each run of the simulation, further results are calculated in End Run Logic;

    a. For each of urgent, routine, emergency, VR and Cataract patients
        i. The percentage of that type of patient seen treated by their due dates (from the count of those treated by their due dates and the total number treated for each type).
        ii. The average waiting time for that type of patient (from the sum of the waiting times and the number treated for each type).
    b. For patients treated on Tuesdays or Thursdays the following are calculated;
        i. The maximum length of a theatre session on that day of the week.
        ii. The minimum length of a theatre session on that day of the week.
        iii. The maximum utilization of a theatre session on that day of the week (where utilization is defined as the time spent actually operating).
        iv. The minimum utilization of a theatre session on that day of the week.

The number of sessions that went over the overtime limit for that day of the week.

vi. The average (mean) session length and utilization for that day of the week.

vii. The values needed to calculate the standard deviation of session length and utilization for that day of the week.

viii. Data items are set to take a selection of the Fairness Testing spreadsheet to the results collection.

ix. The spreadsheet with information on individual patients is used to count the number of times a patient with the same or greater amount of time until their due data who arrives later is treated earlier for each patient, as a measure of how fair the algorithms are.

18. On reset the values of all items changed during the running of the model are reset and the spreadsheets for storing data are all cleared.

Appendix B lists all of the different types of information stored in the model, with references to the list above to show when they are used.

The values output at the end of each simulation run or over which average data are returned at the end of a trial are as follows:

- The average and maximum waiting times for each type of patients – VR, and cataract as well as emergency, urgent and routine.

- The percentage of each type of patients treated before their due date.

- The number of session that ran 'too long' on either day of the week, along with the number of sessions that took place on each day of the week.

- The minimum, maximum and mean length of slots used on Tuesdays and Thursdays.

- The minimum, maximum and mean utilization of slots used on Tuesday and Thursday, where utilization is defined to be the amount of time spent operating.

- Separately for cataract and VR patients, the minimum, maximum and mean waiting time for those patients with initial time remaining to target of 4, 10 and 16 weeks.

- Sums of squares for each of the values for the 3 points above, to be used in calculating standard deviations.
- The number of occasions on which patients have been scheduled so that they exactly fill the remaining theatre time in the slot.
- A count of the number of times patients are overtaken, as a measure of fairness.

6.1.3 Setting up the simulation clock

In order to run the simulation model it is necessary to define the units of time over which the simulation will run, the length of the warm up period before results are collected, the length of time to run the simulation while collecting results and the number of times to repeat the simulation in each trial. This section will cover all of these factors including a brief description of the decision required and the reasons for the selections made in each case.

The units of time over which the simulation will run, is the smallest unit of time for which data can be collected. Simul8's online help desk describes the selection process thus:

"Simulate an amount of time that makes sense to your client in terms of the performance measure you are using. For example if you simulate the factory for a year and report to your client that you expect the factory to produce 14,500 boxes in a year this might be useful information in itself but might hide the information that any given week's production might be as low as zero or as high as 500. Conversely, your client may be unconcerned by information about the output in any given hour. Choose a time that makes sense to the client." (Simul8, 2012a).

In the case of our simulation, we wish to collect data such as the theatre utilisation and whether a day's surgery over ran the length of theatre slot available, so it makes sense to run the simulation in time units of one day. As the theatre slots each only occur on one day of the week and all of the week's arrivals can be simulated in one day, we use a day to represent each week.

The principle behind using a warm up time, before which no data is collected, is "simply to ensure that the simulation is not in some atypical start-up state caused by the

simulation starting empty" (Simul8, 2012b). For the problem we are considering, the surgeon's waiting list does not start empty, but always has some patients waiting to be seen so a warm up period is necessary.

The length of the warm up period should be sufficient to allow the simulation to reach a steady state. In order to determine an appropriate warm up period, the model was run without a warm up period for increasing durations and average waiting times and queue length recorded at intervals. The results are averages over 100 runs for the waiting times and 250 runs for the queue lengths to reduce the effects of data variations and are shown in Figure 10 and 11 below.
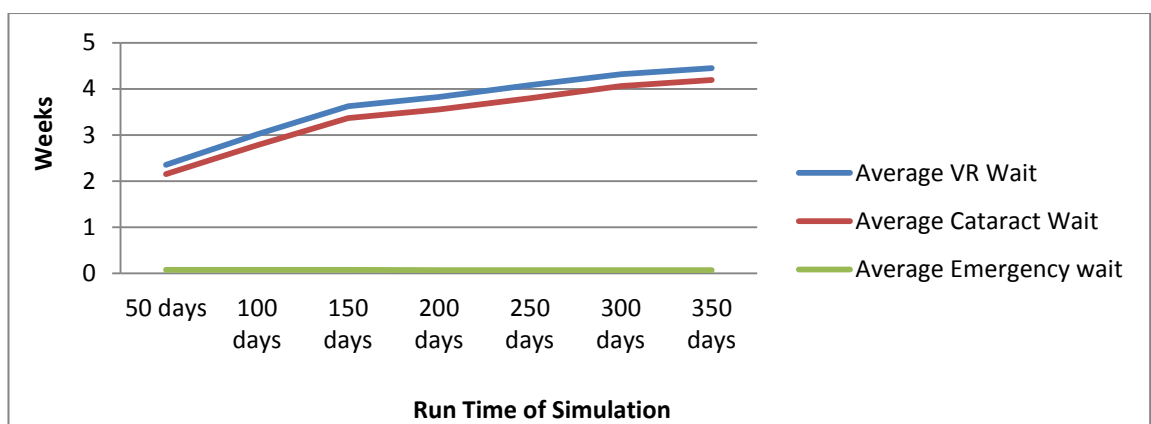
Figure 10: Investigating the time taken for the average waiting times to stabilise.
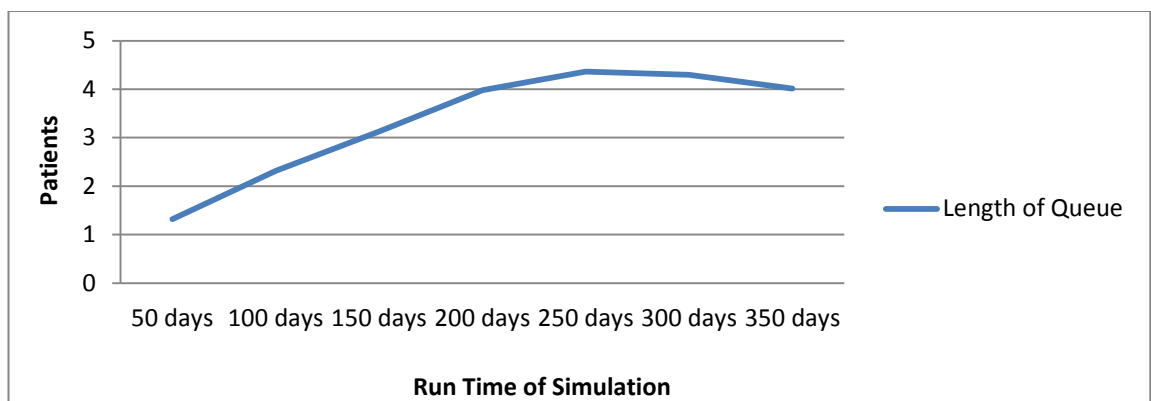
Figure 11: Investigating the time taken for the queue length to stabilise.

As can be seen on the graphs, the average waiting times and queue length are stabilising after 200 days. The amount of time taken for them to stabilise reflects the high traffic intensity of the system, which is slightly greater than 1 as can be seen from the results in Section 6.7 and is addressed by theatre sessions regularly running into overtime.

Consequently a warm up period of 300 days is selected for further simulations. Following the warm up period a results collection period of 300 days is used.

Next we consider the number of times each simulation will be run, so that the averages of the results over all of the simulation runs will be the averages we would expect in the long run of the real system. In order to be able to distinguish which is the better of two algorithms we will require the 95% confidence intervals for the values of the output variables to be non-overlapping (Simul8, 2012a). This will be more easily achieved if sufficient trials are done that the 95% confidence intervals are quite narrow to start with. To ascertain how many trails will be necessary to achieve this, simulations were run with varying numbers of trials and the results are given in Figure 12 and Figure 13 below.
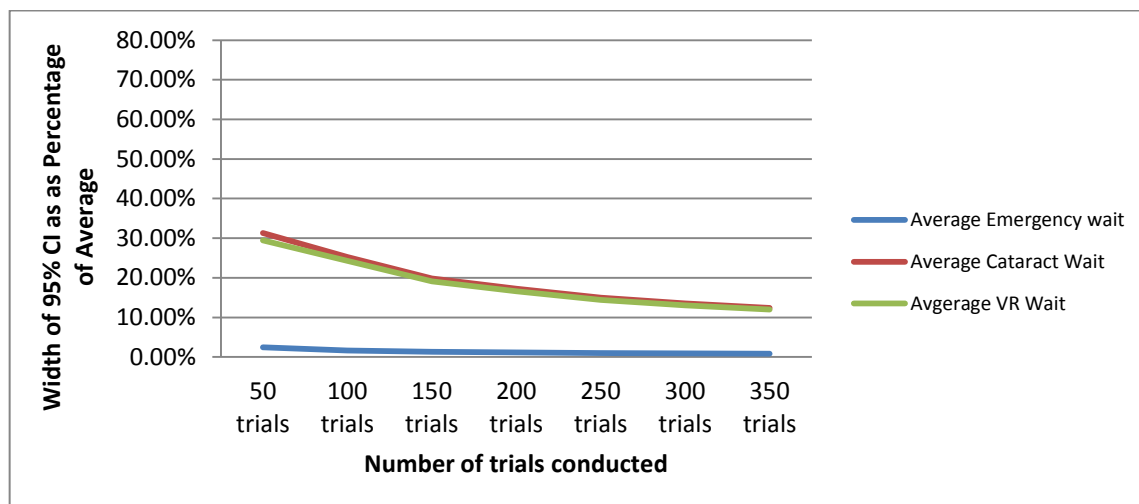
Figure 12: Investigating the effect of increasing the number of trials on the 95% confidence interval (CI) for waiting times.
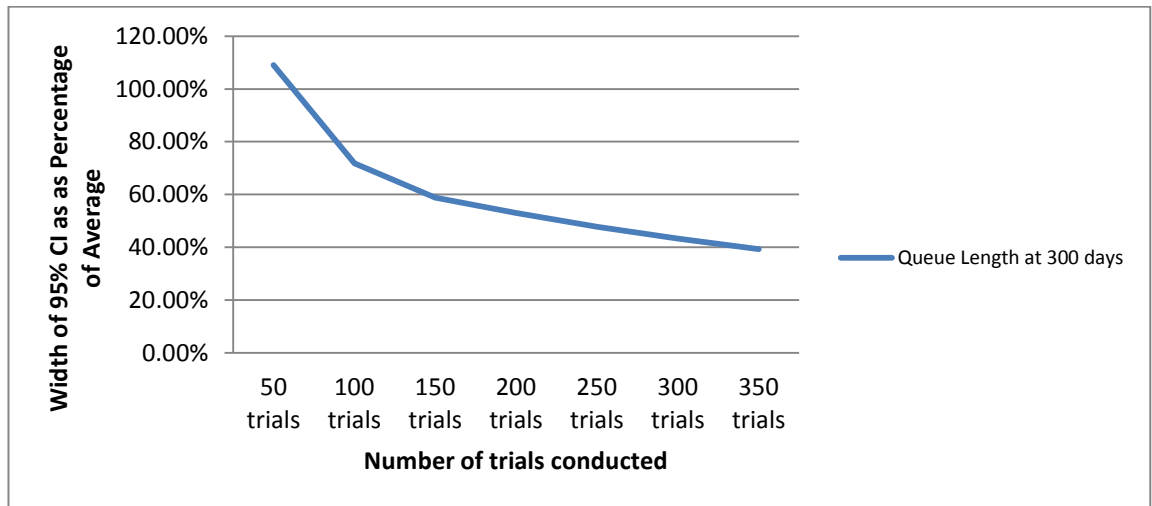
Figure 13: Investigating the effect of increasing the number of trials on the 95% confidence interval (CI) for queue length after the model has been run for 300 days.

Initially as the number of trials increases the confidence intervals decrease rapidly, after 150 trials the decrease has begun to level out for the variables considered. After 250 trials the improvement in these key outputs is considerably reduced, so 250 trials will be run for each variation on the simulation model.

To summarise, the simulation models will be run with time units of 1 day, with a 300 day warm up period, a 300 day results collection period and each trial will be repeated 250 times with the average and 95% confidence interval recorded for all of the output data.

This model can now be populated with data and used to test algorithms as described in the following two sections.

6.2 Data

This section describes the data used to test the model with the individual surgeon's case mix as a case study. Following consideration of the data required, the data sets obtained are first described, then the way the data required is calculated and the information thus obtained are discussed.

6.2.1 Data required

In order for the simulation model to represent a particular surgeon's case mix it needs to be set up with data representing his patients. In this case the data required includes:

- The inter arrival time distributions for cataract and VR patients arriving to the system in need of treatment.
- The distribution of time remaining until due dates for both cataract and VR patients; this includes the proportion who are urgent and the number of weeks in which they need to be seen, and for routine patients the time remaining on the 18 weeks from referral to treatment target.
- The distribution of theatre times for cataract and VR patient groups, for creating the predicted operation durations.
- The variation from expected operation duration for all patient types.
- The expected turnaround time between patients (the time needed to prepare the theatre between patients).
- The distribution of actual turnaround times.
- The expected arrivals of emergency patients requiring treatment in each slot.
- The distribution of operation durations for emergency patients.
- The frequency and duration of the theatre slots considered.
- The level of overtime considered acceptable for the theatre slots considered.
- The frequency with which theatre slots are closed, e.g. by annual leave.

Some of this data, for example the duration of theatre slots and acceptable overtime, can be obtained from discussion with the surgeon. The remainder needs to be extracted from recent hospital data.

6.2.2 Data obtained

The data was obtained in a number of data sets, the contents of which are set out below, before the subsequent section describes how the desired data was obtained from them. The data sets included unnecessary data as well as that required; only the relevant data is listed below:

- Theatre Times data, containing data from May to October 2010 (later data was not available due to a change in the database), including the following for each operation;

138

- o Operation date
- o The planned start and end times for sessions
- o The time at which anaesthesia commenced
- o The times at which the operation started and ended
- o The time at which the patient entered recovery (left the theatre)
- o The procedure or procedures carried out
- o The consultant in charge
- o Whether the case was elective or emergency surgery
- o Transformed NHS number (so that it would not be identifiable but could be used as a unique identifier when compared to other data sets)
- Sample data looking at the accuracy of predicting operation durations collected by the surgeon, as this is not routinely collected;
  - o Procedure
  - o Predicted and actual operation durations

- Inpatient data set including;
  - o Codes for the type of operation and specific procedure along with brief descriptions of the procedure
  - o The date of surgery
  - o The date of admission to hospital
  - o A code identifying the consultant and name of consultant
  - o The date of decision to admit
  - o The expected length of stay (almost all zero, as they do not stay overnight)
  - o Whether the patient was elective or emergency
  - o The waiting time of the patient
  - o A transformed version of the patients NHS number (as above)
- An outpatient data set;
  - o Referral date
  - o Appointment date
  - o Consultant
  - o Outcome of appointment
  - o Transformed NHS number (as above)

6.2.3 Data preparation

We start with the data derived from the theatre timing data.

For a significant number of cases, the anaesthetics for one case are started before the previous case had left theatre, so the anaesthetics are included in the turnaround times rather than the operation times. Also, some minor cases involving injections can be dealt with by a fellow (training to be a surgeon) in the anaesthetics room while the surgeon is operating on minor cases. As these do not affect the other operation durations, etc., they have been completely excluded from the analysis, by looking up and removing their procedure codes.

Starting times for the surgeon under consideration are shown in Figure 14; these are the difference between the expected and actual start time for his theatre slots.
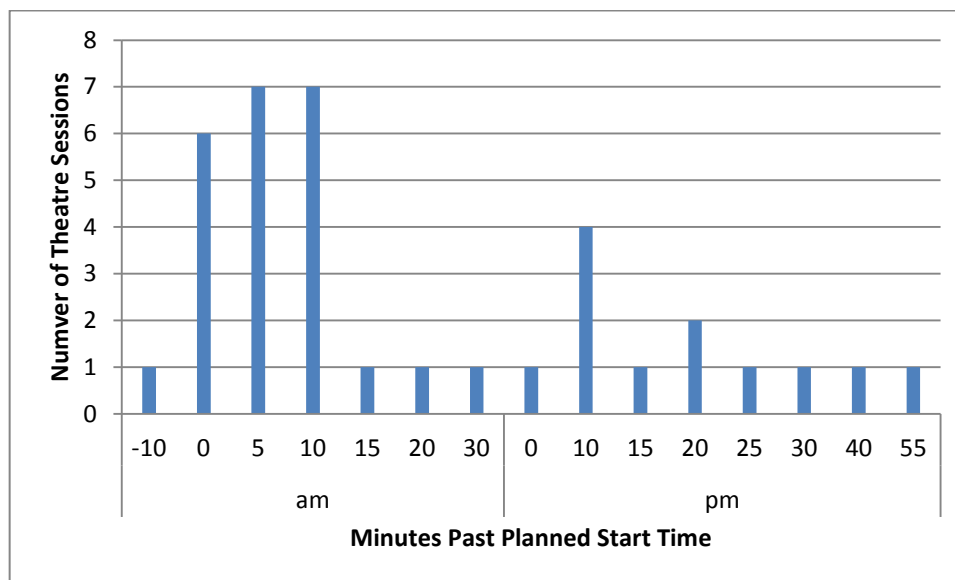
Figure 14: Morning and afternoon starting times – variation from planned start of session.

The division between pm and am slots is for afternoon and morning theatre sessions. This shows that the majority of the sessions start within 10 minutes of the expected start time. There are a number of reasons why a session might start late, such as an overrunning morning session delaying the start of an afternoon slot. For the simulation model, we are concerned with knowing when sessions start, not why they might start late.

Operation durations for the surgeon under consideration are shown in Figure 15.
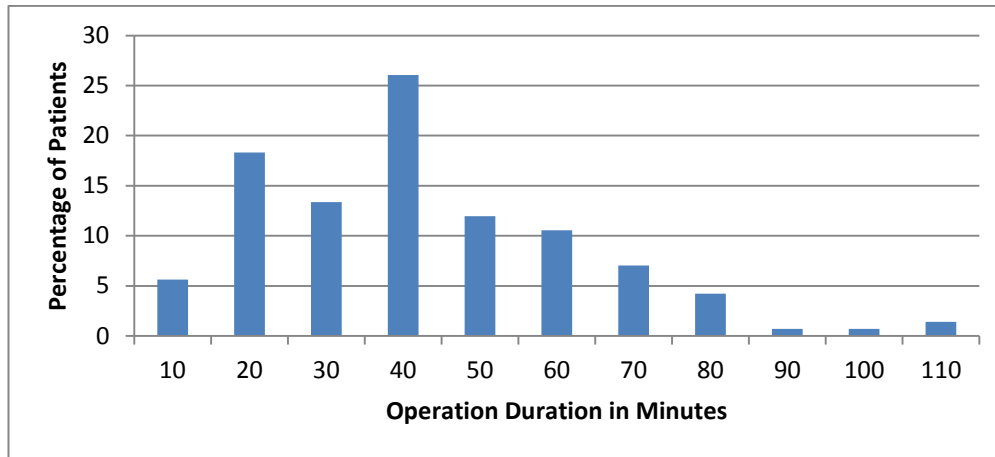
Figure 15: The frequency with which different operation durations occur in the data.

Figure 15 shows a smooth distribution to the durations of operations. Following discussions with the surgeon, we looked at separating those having cataract and VR surgery in the data. The results of this are shown in Figure 16.
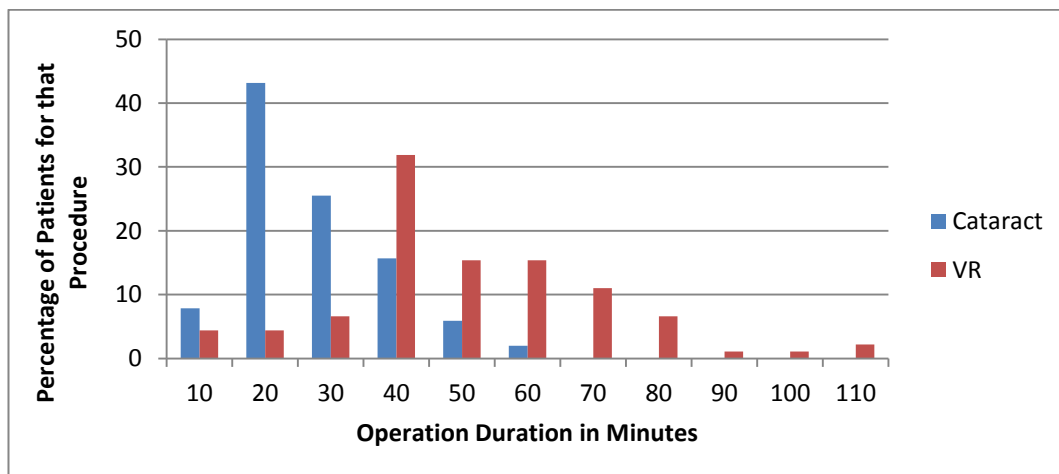
Figure 16: Comparison of the distributions of operation durations for cataract and VR patients.

It is clear that the distribution the durations of cataract operations is different from the distribution of durations for VR patients.

The durations of emergency operations are shown in Figure 17 below. The number of such operations over the time period for which data was available is low. Therefore it would be hard to fit a distribution to this data, but it can still be sampled within a simulation model. Also due to the low number of emergency patients in the sampled

time period, it was not considered sensible to separate out their turnaround times, so these are included with the elective cases.
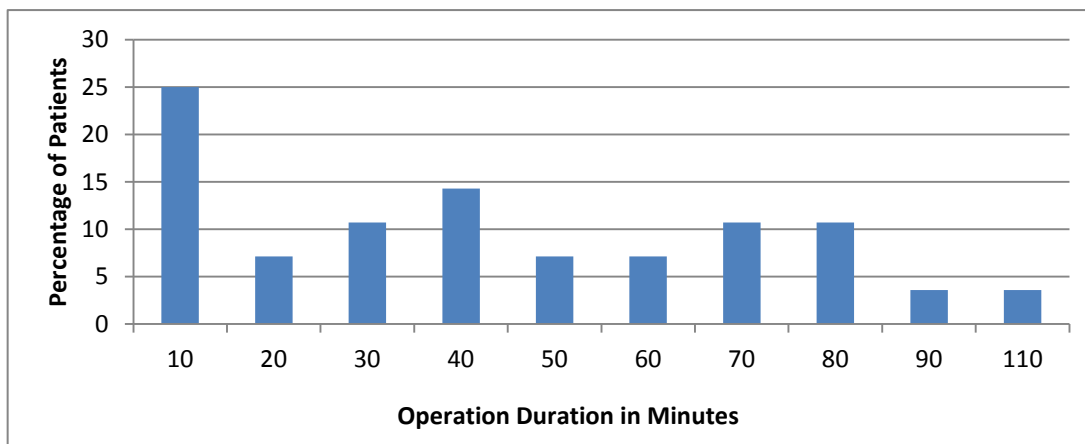
Figure 17: The distributions of operation durations for emergency patients.

Turnaround time is defined as the time between the end of one operation and the start of the next; this includes anaesthetic time and recovery time as well as cleaning and preparation of the theatre between patients. Figure 18 below shows the distributions of turnaround times.
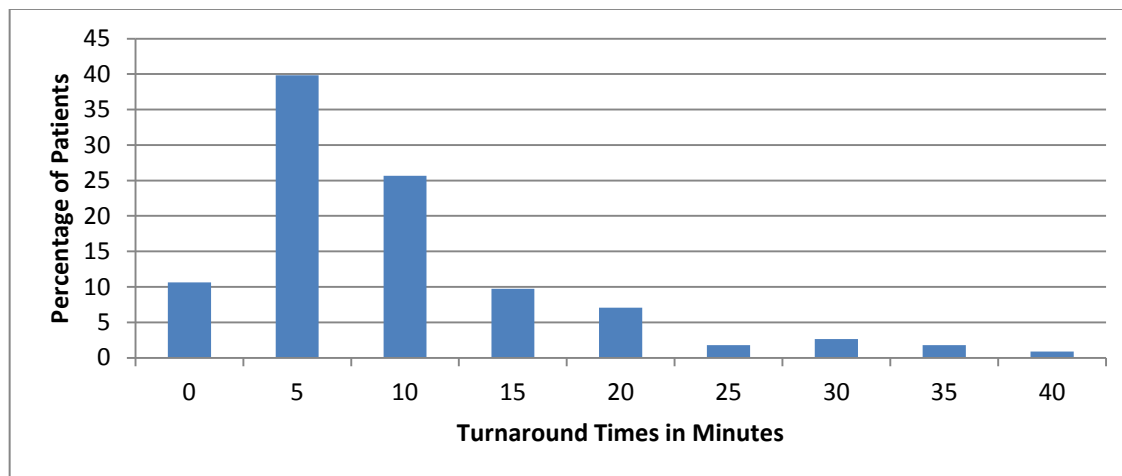
Figure 18: The distribution of turnaround times.

Figure 18 shows that the majority of turnaround times are 10 minutes or less, but some take considerably longer. When these results were shown to the surgeon involved he felt that they were considerably longer than expected, so further consideration of their composition was required. Figure 19 below shows the distribution of anaesthetic times.
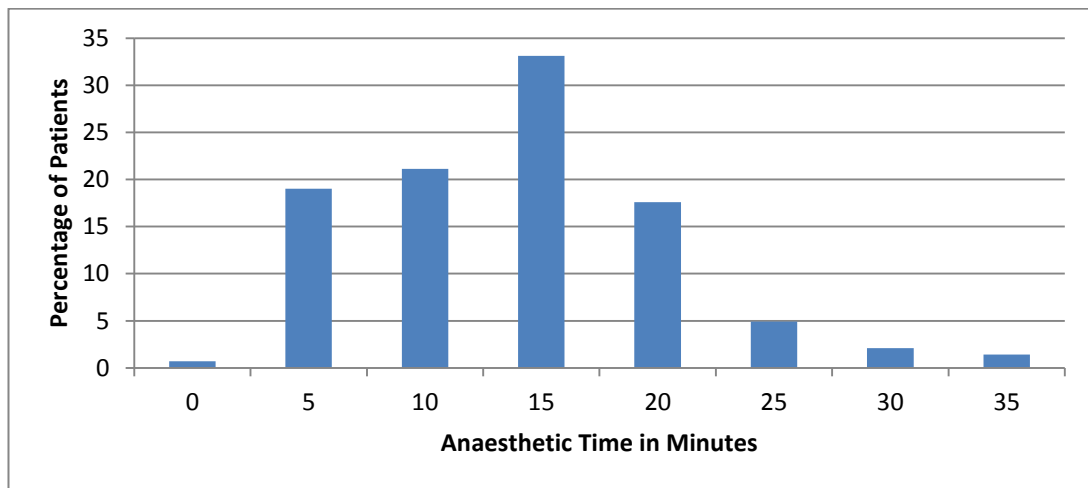
Figure 19: The distribution of the time taken to administer anaesthesia.

Figure 19 shows that in some cases the anaesthetic can take considerably longer than average to administer, enough to explain the longer turnaround times. Given that the modal anaesthetic time is longer than the modal turnaround time, it is clear that time savings are being made by starting one patient's anaesthesia before the previous operation is completed. Following discussion of these results the surgeon was happy the distribution of turnaround times above is appropriate.

Based on this data, a mean turnaround time of 10 minutes is used in the simulation model as the expected turnaround time and then the actual turnaround time is sampled from the turnaround times given above. This is rounded up from the actual mean of 7.4 minutes calculated from the data, so that 75% of turnaround times will be less than the time allowed when planning operations.

From the predicted and actual operation duration data, it is possible to estimate the variation from predicted durations. Using these times directly is not so useful, because if for example the expected duration of an operation is 20 minutes and sampling from this data gives an adjustment of minus 30 minutes then that gives us an actual duration of -10 minutes, which is not possible. So the data values were recalculated looking at the percentage change in duration as shown in Figure 20.
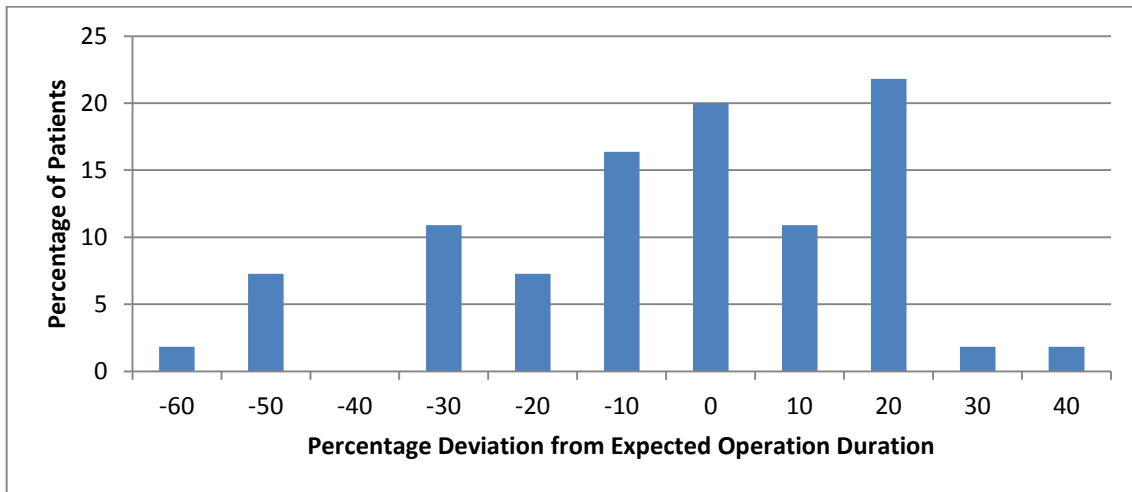
Figure 20: The distribution of percentage deviations from the expected to actual duration of operations.

Originally, the intention was to use the transformed NHS numbers to link records between the inpatient and outpatient data sets, to obtain the referral dates for the surgical cases from the outpatient data. This matched only 121 out of 642 inpatient data records to outpatient data and graphing the time from referral to surgery yielded the results below;
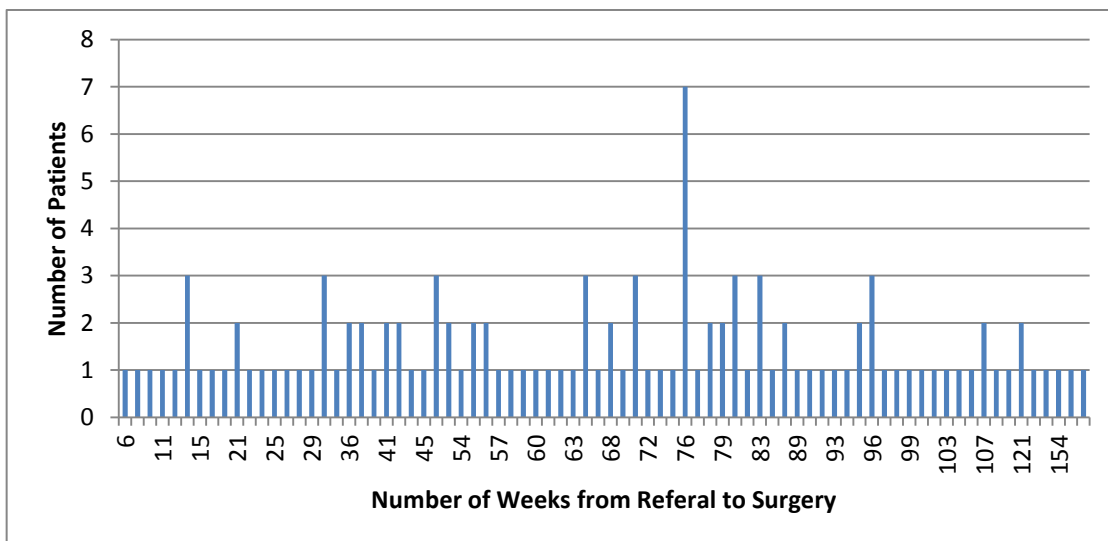
Figure 21: The frequencies with which different times from referral to treatment occurred when looking at combined inpatient and outpatient data sets.

As Figure 21 suggests that the majority of patients wait much longer than the 18 weeks from referral to treatment target. This graph tells us that the linking of the inpatient and outpatient data sets to find the time remaining to target has not been successful. Thus, it

is not possible to obtain a time remaining to due date by linking the inpatient and outpatient data sets.

This may be due to many patients being monitored as outpatients before their conditions reach the stage where surgery is required, which creates significant gaps between referrals and decisions to operate. Meaning that for those patients the clock will start on the 18 week target when the decision to operate is made or the urgency will be based on clinical need.

As it did not prove possible to combine the inpatient and outpatient data sets in order to obtain the target dates, the best approximation for this is to assume that all patients were treated on their due dates and take the time waited as the time to due dates from when the decision to operate is made. This will result in overly constrained due dates as at least some of the patients will have been treated before their due dates. Figure 22 shows the spread of waiting times for both cataract and VR patients.
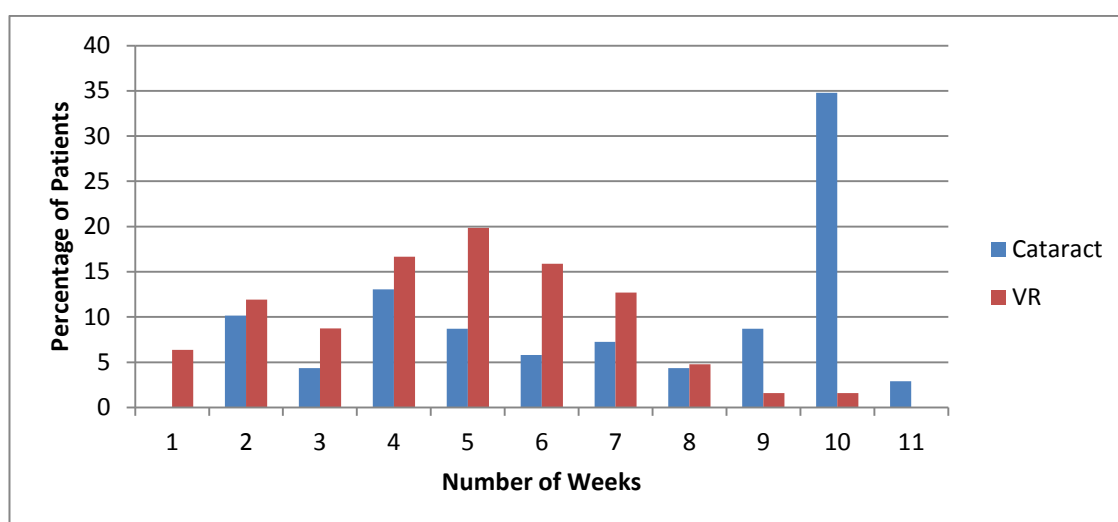
Figure 22: The waiting time distributions for cataract and VR patients, with times rounded to the nearest 2 weeks.

This gives the best approximation that the data allows for the number of weeks from the decision to treat until the patient is due for treatment. Of the VR patients whose waiting time rounded to zero weeks are included in those with a 2 week target in the simulation model as otherwise they should be emergencies. The cataract patients who waited 20 weeks are included in the 18 week target as the maximum target is 18 weeks. The peak in cataract patients at 18 weeks is due to the 18 week target.

The inpatient data set can be used to calculate the number of decisions to operate each week for cataract and VR patients, by using the decision to admit date and not including the more recent cases (as some of those may not have been treated yet). Figure 23 below shows that there is a reasonable amount of variation in the number of patients added to the waiting list each week.
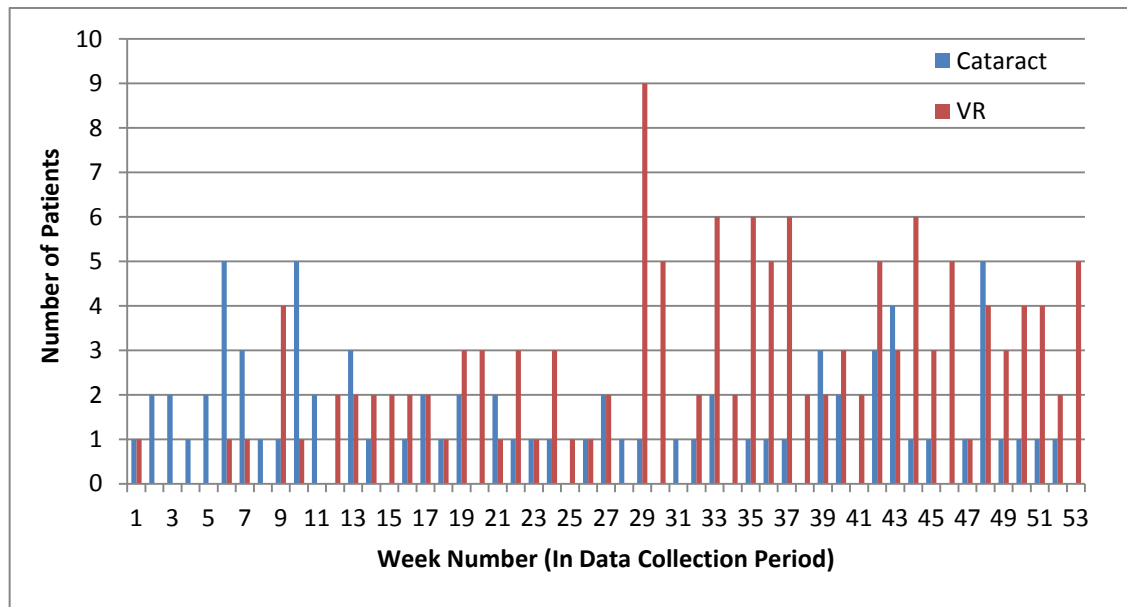
Figure 23: The number of patients added to the waiting list each week.

Table 3 gives the frequencies with which different inter arrival times occur.

Table 3: The frequencies with which different inter arrival times (in weeks) occurred for elective surgery

Inter Arrival Time (weeks)	Frequencies	
	Cataract	VR
0.1	0%	2.273%
0.11111	2.273%	0%
0.14286	0%	0%
0.16667	9.091%	6.818%
0.2	11.364%	9.091%
0.25	9.091%	11.364%
0.33333	18.182%	15.909%
0.5	29.545%	29.545%
1	20.455%	25.000%

146

This data can be used to provide empirical distributions for the inter arrival times of cataract and VR patients to input into the simulation model.

The following contains similar data for the number of emergencies; this is based on the number treated each Tuesday and Thursday. It is based on the number of emergencies treated, which is not exactly the same as the inter arrival times, because sometimes emergency patients can safely wait until the next day to be treated if there is spare capacity in the bookings for the next day, but as data was not available to determine the frequency with which this occurs, it is assumed that emergency patients will be treated on the day of arrival in our model.

Table 4: The frequencies with which different numbers of emergency patients arrived during the sessions considered

Inter arrival time for Emergencies (weeks)	Frequencies	
	Tuesday	**Thursday**
0.2	3.448%	3.448%
0.25	13.793%	0%
0.33333	10.345%	13.793%
0.5	44.828%	31.034%
1	27.586%	51.724%

The frequency and duration of the theatre slots considered was obtained from discussions with staff and is also apparent in the data, every Tuesday morning from 8.30 am to 12.30 pm and alternate Thursday afternoons from 1.30 pm until 5 pm.

The level of overtime considered acceptable for the theatre slots was obtained from staff – half an hour for morning slots and 1 hour for afternoon slots.

The frequency, with which theatre slots are closed (the timetabled surgeon does not operate in the slot), was obtained from the inpatient data by looking at the days when operations did not occur. On both Tuesdays and Thursdays, slots are not used approximately 10% of the time.

Thus, all of the data required listed in Section 6.2.1 have been obtained or estimated as well as possible from the data available. Given that the delay from planned start times were not dissimilar from the turnaround times, the process is simplified in the model by treating the start times as turnaround times.

6.3 Fitting Distributions to the Data

Thus far we have worked with empirical distributions to ensure that the actual data is represented as closely as possible. Working with fitted distributions will greatly facilitate the exploration of variations from the original data, and depending on how closely the distributions fit the data is expected to give the first variation from the original data.

This section will work through the various data required for the simulation, showing which distributions were tested and explaining the choices made for the best fit. For each distribution (unless otherwise stated), a visual assessment of the distribution(s) most likely to fit the shape of the distributions is conducted and Minitab® Statistical Software used to fit the parameters of that distribution to the data. The fitted distribution is then compared to the actual distribution graphically, and if two distribution have been tested an assessment is made as to which is the better fit for the data.

As discussed in Section 6.2.3 turnaround times are used to represent start times for the first operation of the day, so we do not need to model the starting times. This means that the data (as for Section 6.2) to which distributions are to be fitted includes:

- The inter arrival time distributions for cataract and VR patients arriving into the system in need of treatment.
- The distribution of due dates for both cataract and VR patients.
- The distribution of theatre times for cataract and VR patient groups.
- The variation from expected operation duration for all patients.
- The distribution of turnaround times.
- The expected inter arrival time distributions for emergency patients during the theatre slots considered.
- The distribution of operation durations for emergency patients.

We will now take each of these in order fitting appropriate distributions.

148

6.3.1 Arrival distributions

The inter arrival distribution for cataract patients has a strong left skew and long tail, which suggests that the log normal distribution would be a reasonable fit for the data.
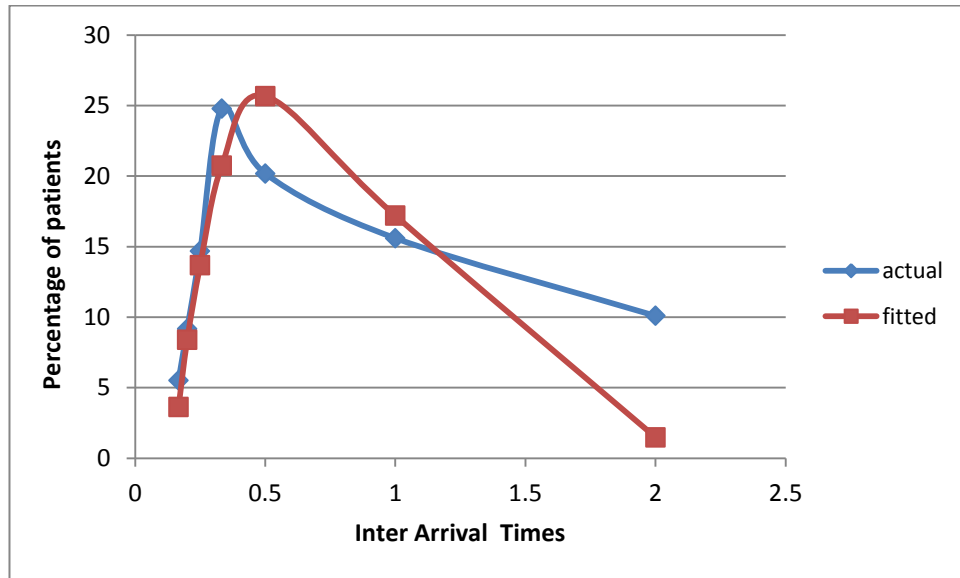
Figure 24: Graph showing the fitted log normal distribution compared to the arrival rate for cataract patients.

The mean of the distribution is 0.4421 and the standard deviation is 0.2718. The first data point of the fitted distribution is higher than the actual distribution, indicating that the cumulative distribution up to that point has included more patients than for the actual distribution.

The skew on the fitted distribution is close to that of the original distribution, but the tail of the fitted distribution drops more rapidly than that of the original.

Thus, the fitted distribution is similar to the original but not a very close fit. As we are looking to explore variations from the original distribution this is acceptable.

The VR inter arrival distribution has a less of a skew, which suggests that the normal distribution would be a reasonable fit for the data; we test both normal and log normal below.
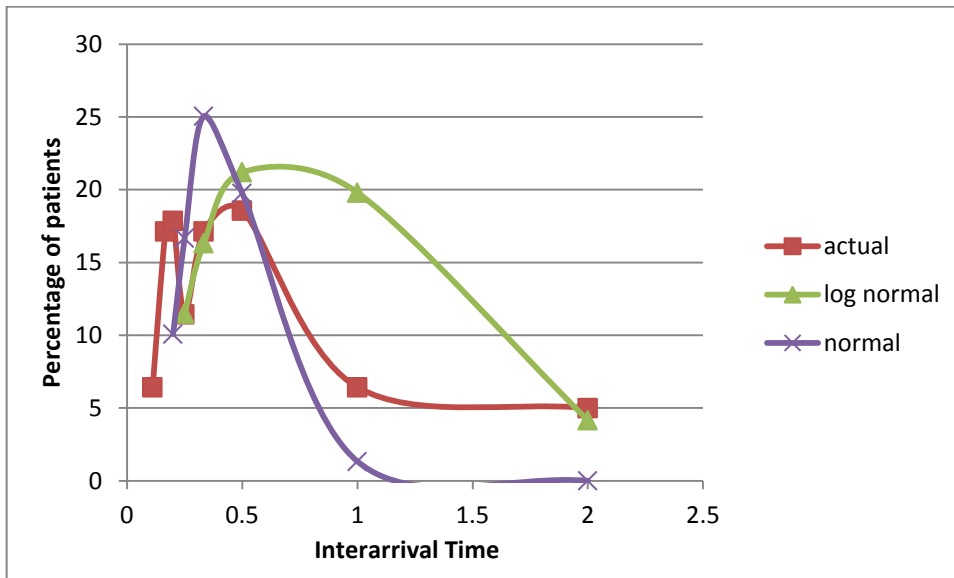
Figure 25: Graph showing the fitted distributions fitted to the distribution of VR arrivals.

Neither the normal or log normal is a close fit for this data set, but the normal has its peak closest to the values taken by the actual data, whereas the log normal is further from the actual data across two thirds of the graph. As neither of the graphs is a good match and it is the distribution, the empirical distribution will continue to be used for the VR entry data.

6.3.2 Due date distributions

The cataract due date distribution has two distinct sections suggesting that a mixture of two distributions will be required to model it. The shape of the distribution, particularly the second peak, suggests that two normal distributions would work. In order to fit normal distributions we have used solver in excel to fit the means and standard deviations of both distributions as well as the proportion of each to use. The resulting distribution is shown in the graph below.
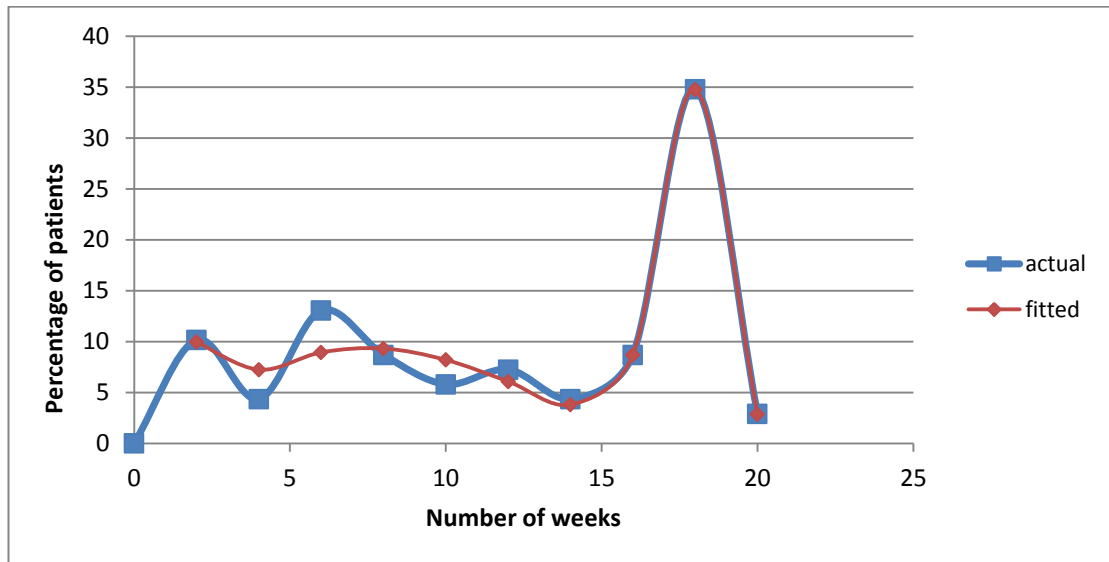
Figure 26: Graph showing the fitted combination of normal distributions compared with the actual distribution.

The mean of the first normal distribution is 6.492 and its standard deviation is 4.814, the mean of the second normal distribution is 16.788 and its standard deviation is 0.7756. The combined distribution is formed by taking 56.02% of samples from the first distribution and 43.08% of samples from the second distribution.

As can be seen in Figure 26 this gives a good match for the distribution of due dates for cataract patients, particularly the second peak.

The distribution for the due dates of VR patients is much closer to the shape of a single normal distribution.
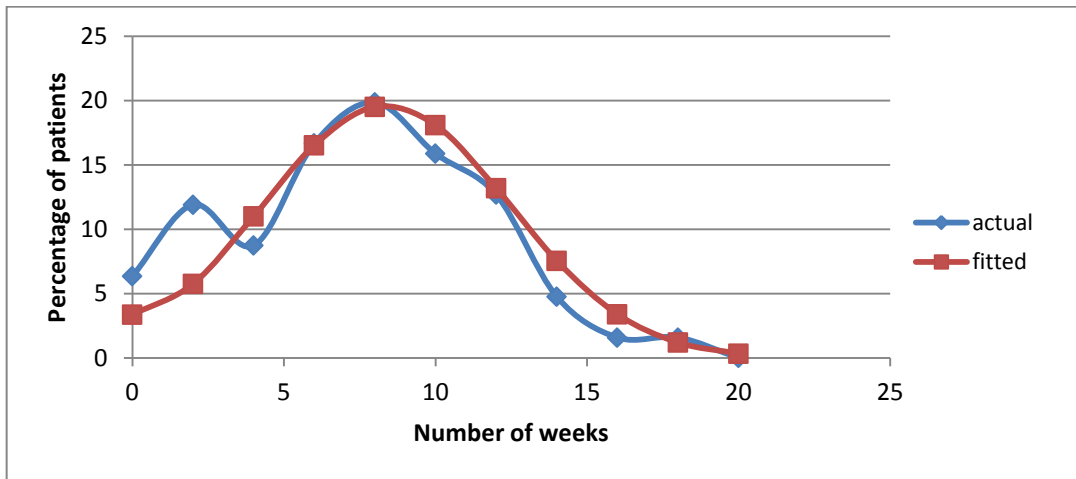
Figure 27: Graph showing the fitted normal distribution for the due date distribution of VR patients.

Figure 27 shows that a normal distribution with mean 7.376 and standard deviation 4.032 is a reasonable fit for the distribution of due dates for the VR patients.

6.3.3 Operation durations

The graphs below show that normal distributions are not exact fits for the distributions of operation durations for either cataract or VR patients. However, the normal distributions shown the graphs follow the general trends of the empirical distributions and the shape of the distributions does not suggest any other distributions that might be better.
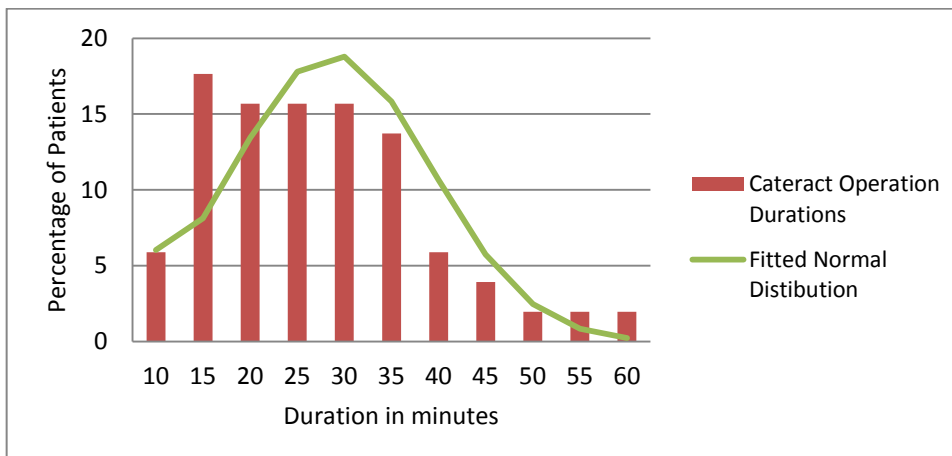
Figure 28: Graph showing the fitted normal distribution for cataract operation durations.

Figure 28 shows that a normal distribution with mean 26.21 and standard deviation 10.44 is an acceptable fit for the distribution of cataract theatre times.
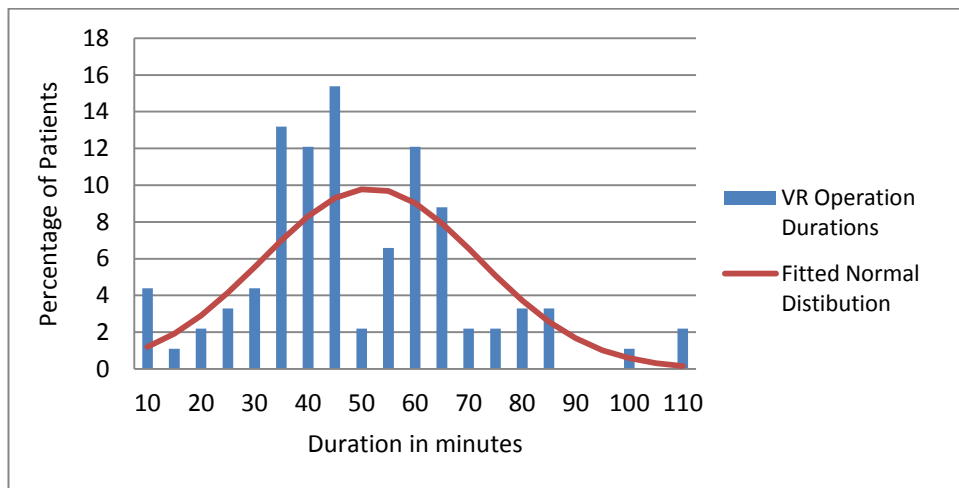
Figure 29: Graph showing the fitted normal distribution for VR operation durations.

Figure 29 demonstrates that the normal distribution with mean 49.21 and standard deviation 20.27 is an acceptable fit for the distribution of cataract theatre times. Since it is possible that the dip in the actual data between 45 and 60 minutes may be due to actual times being rounded to these figures in the original data, the fit to the actual theatre times may be better than the data suggests.

6.3.4 Variation from expected durations

The distribution of the percentage of difference from the predicted theatre times is not a particularly smooth graph, but appears to be approximately normally distributed.
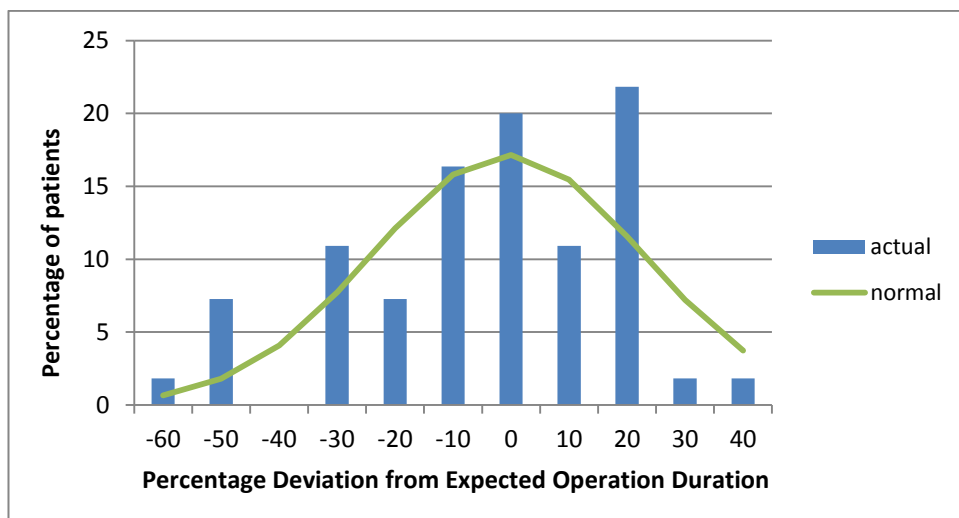
Figure 30: Graph showing the normal distribution fitted to the percentage difference between actual and predicted operation durations.

Figure 30 shows that the normal distribution with mean -5.625 and standard deviation 23.07 is a reasonable fit for the distribution of the percentage difference form the predicted operation durations.

6.3.5 Duration of turnaround times

The theatre turnaround times show a strong negative skew so again a log normal distribution is fitted.
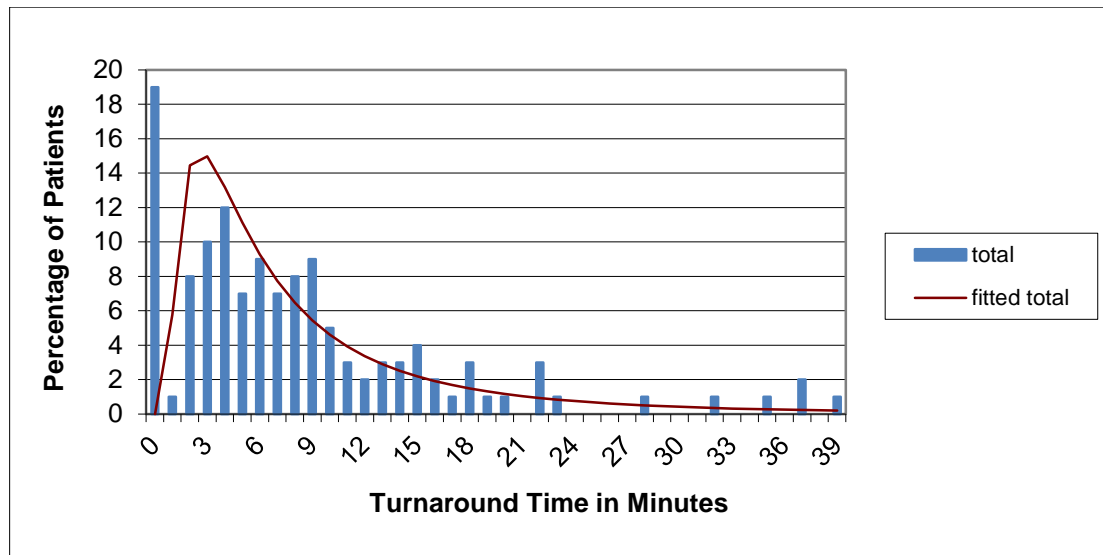
Figure 31: Graph showing the fitted log normal to the theatre turnaround times.

Figure 31 shows that the log normal distribution with mean 8.32 and standard deviation 8.176 is a reasonable fit for the distribution of turnaround times.

154
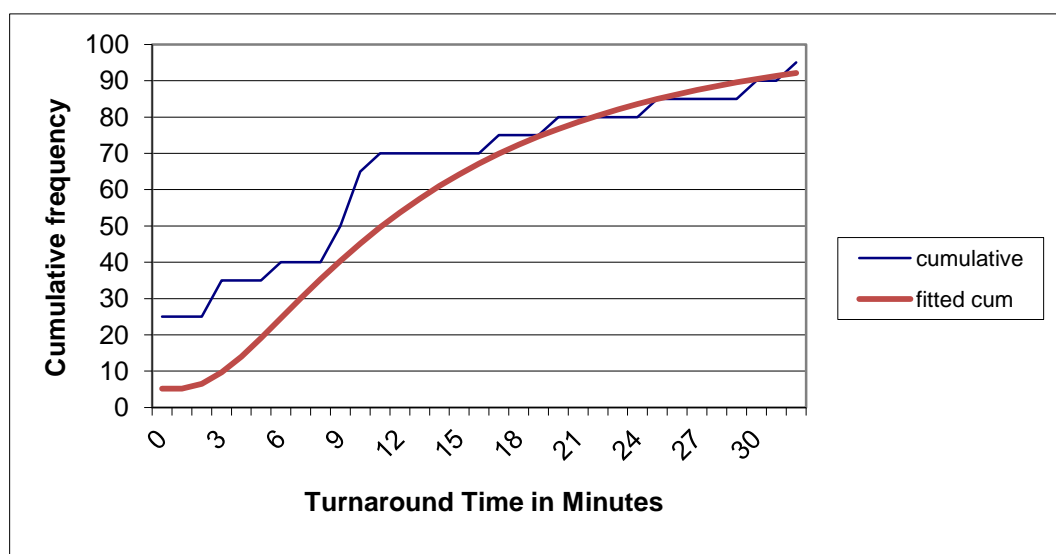
Figure 32: Graph showing fit of log normal distribution to emergency turnaround durations..

Figure 32 shows that the log normal distribution with mean 1.091 and standard deviation 0.3652 is an acceptable fit for the distribution of turnaround times. Due to the small number of data items graphing the fitted distribution against the actual distribution does not provide a clear graph, so the fit is illustrated with the cumulative distributions.

6.3.6 Emergency arrivals distribution

The inter arrival distribution for emergency patients is a steep curve so the exponential distribution is fitted to it below.
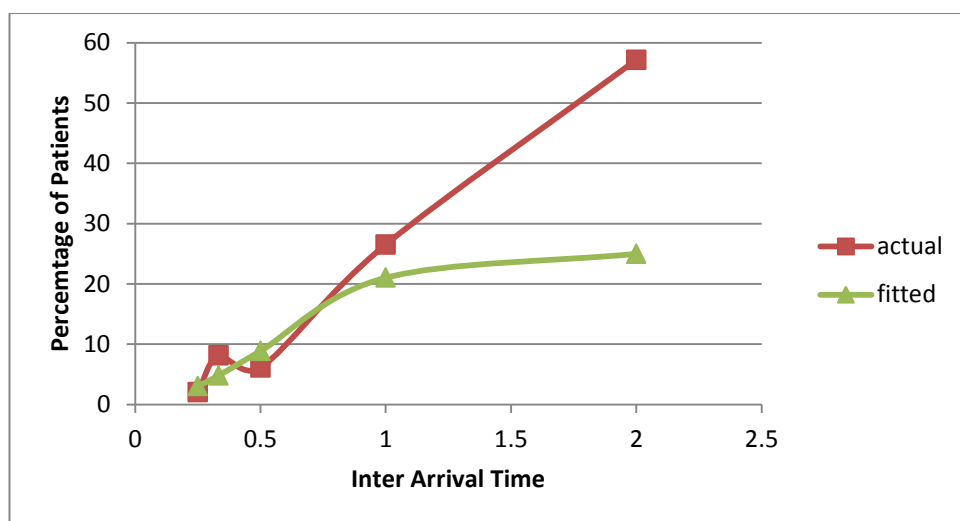
Figure 33: Graph showing the fitted exponential distribution for emergency inter arrival times.

As can be seen from Figure 33 the exponential distribution is not a very good fit for the emergency inter arrival times as it fails to capture the steepness of the curve. As it does not seem possible to fit a suitable distribution, we continue to use the empirical distribution for emergency inter arrival times.

6.3.7 Emergency operation time distribution

The distribution of emergency operation durations is also negatively skewed so again a log normal distribution is fitted.

Due to the small (discrete) number of data items for each duration graphing the fitted distribution against the actual distribution does not provide a clear graph, so the fit is illustrated with the cumulative distributions.
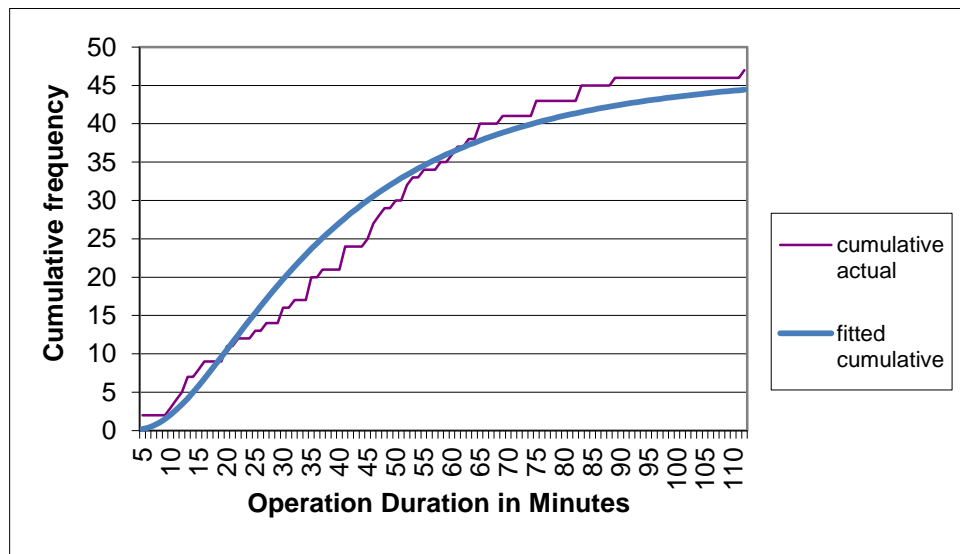
Figure 34: Graph showing log normal distribution fitted to emergency operation durations.

This shows that the log normal with mean 46.02 and standard deviation 25.39 is a reasonable fit for the distribution of emergency operation durations.

A more rigorous approach to the fitting of distributions would test goodness of fit using statistical tests. In this case the modelling has already been conducted with the empirical data, so the intention is to create data which is similar but not the same as the original data, in order to test the sensitivity of the results. As such, checking the fit of the distributions by eye is sufficient.

Thus, distributions are fitted to replace all of the empirical distributions except for the VR and emergency inter arrival times. Next these will be used in the simulation model to explore any changes to the effectiveness of the algorithms.

6.4 Implementing the algorithms

This section will discuss how the algorithms for testing have been selected from those discussed in Section 5.3.2 along with how they are implemented. The references in brackets refer to the bullets in Section 5.3.2. The subheadings relate to groups of

algorithms and the algorithms are discussed in the same groups in the results in Section 6.5.

6.4.1 Booking as soon as possible

The first algorithm considered is booking each patient as soon as it is known that they require treatment (1a), in the order of arrival (2a) and into the first available slot (3a). This is a straightforward algorithm to implement and is intended to be used as a baseline against which to judge the more sophisticated algorithms. It is as close as possible to treating patients in a first in first out (FIFO) manner, given the need to fit their expected operation durations into the theatre slots, and is referred to as 'FIFO no P' in the results section.

The second algorithm is the same as the first but with patients booked in priority order of due date using expected operation duration as a tie breaker with longer operations given higher priority (2f), and is referred to as 'FIFOp' in the results section. As two patients arriving at the same time will not occur often, it is expected that this will not have much if any effect.

The third algorithm is as the second but with patients batched and booked at the end of the week rather than as the decisions to operate are made (1b, 2f, 3a), and is referred to as 'FIFO batching' in the results section. This method is intended to explore the effects of such batching.

The fourth and fifth algorithms are as the second, but only allowing patients to be booked in to the next 2 or 6 weeks ahead of the current date (1c, 2f, 3a), these are referred to as 'FIFO xw' where x is 2 or 6 in the results section. They are to explore the effects of delayed booking algorithms.

6.4.2 Booking to due date

The sixth algorithm is based on the surgeon's suggestion of booking patients as late as possible to allow for the arrival of more urgent patients. It involves starting from the patients due date and considering slots in reverse order until one is found with sufficient space to book the patient into that slot. If no suitable slot is found between the due date

157

and the current time then the first suitable slot after the due date is used (1a, 2f, 3b). This is referred to as 'ToDD' in the results section.

A potential drawback with 'DD' is that space in some slots could be left unused, resulting in more cases to treat later on and greater delays. To address this, the next algorithm considered is booking to the due date unless there is sufficient space the following week in which case the patient is booked into that space (1a, 2f, adapted version of 3b), it is referred to as 'ToDDornext' in the results section.

We also consider delayed booking versions of 'ToDDornext' where patients are booked only into the next 2, 6 or 10 weeks (1c, 2f, adapted version of 3b), referred to as 'DDornextxw' with x replaced with 2, 6 and 10 respectively in the results section.

Next a variation on the 6 week version is tested with patients booked in due date order, with tie breaks split on a first come first served bases. This is intended to test if using operation duration for the tie breaks is making a difference (1c, 2d, adapted version of 3b) referred to as 'Ddorn6WDDpriority only' in the results section.

6.4.3 Booking to a percentage of due date

The next set of algorithms considered is as ToDDornext, but booking to 80% of the time from the current time to the patients due date, followed by the same with 70%, 50% and 40%, referred to as '80%DDornext' etc. in the results Section (1a, 2f, variations on 3b). Batching is also tested on the 80% version (1c, 2f, variations on 3b) and variations on the amount of time available for booking on Tuesdays are considered to explore the effects on overtime requirements.

6.4.4 Discussion of booking priorities

Up to now only 2a, 2d and 2f have been considered as options for the booking priority. Last come first served (2b) is not considered because it would be particularly unfair with similar patients being treated in reverse order of arrival. Most urgent first (2c) is also not tested because by using due dates routine patients effectively become more urgent over time and are thus treated fairly. Arranging patients in order of operation duration (2d and g) is not considered separately as due dates are given higher priority than operation durations. As the initial results show, using due date order with operation

duration as a tie breaker does better than the other priority orders, so only this booking priority (2f) is used in subsequent algorithms.

The initial results from the algorithms show that delayed algorithms (1c) do well, with batching (1b) not doing better than booking as patients arrive into the system (1a), so only (1a) and (1c) are considered henceforth. While (1c) clearly does better than (1a), it is better for patients to be given good notice of their operation dates, so 1a is still considered to see if a variation on the booking order can improve it significantly.

6.4.5 Booking to most empty slots or exact fits

The remaining algorithms consider variations and combinations of booking to the most empty slot and/or to a slot that the patient fills exactly (3c) and (3d), with patients booked either as soon as they arrive to the system (1a) or only into the next X weeks (1c), and in order of due date priority with tie breaks going to the patient with the longest expected booking duration (2f). For algorithms based on slots that the patients operation duration fills exactly, a small degree of tolerance is allowed, so that if the total expected operation durations of the patients will slightly over or under run, the allowed booking time for that slot they will be considered an exact fit for that slot.

These algorithms include exploration of booking patients to due date unless there is a slot that the patient fills exactly with variations on the number of weeks ahead that can be booked (1c, 2f, 3c ii). Whether the search for an exact fit should start from the due date and move backwards or treat those which are exact fits in FIFO order is explored.

Rather than searching for the first slot in which the patient fits moving back from their due date if there is no exact fit, searching for the emptiest slot available is considered. Variations on this are explored with the search for the exact fit conducted so that tie breaks will be for the slot closest to the due date (DD) or closest to the current time (FIFO order) and the same for tie breaks on the most empty slot.

In all of these algorithms if there is a slot the next day available for booking in which the patients expected duration will fit the patient is booked into that slot, so that time will not be wasted as discussed in Section 6.4.2.

In the names of the algorithms 'n' refers to this consideration of the next available day, 'exact' to looking for exact fits, 'most empty' to searching for the emptiest slot, 'DD' to searching back from the due date for that aspect of the algorithm, 'FIFO' to searching forwards from the current time and 'xw' as before to only booking into the next x weeks. These abbreviations are necessary to keep the length of the algorithm's names reasonable for fitting onto graphs and tables in the results section.

For all algorithms, if it is not possible to book a patient into a slot before their due date then they are booked into the first available slot after it.

6.5 Validation and Verification

In order to ensure that the model both represents the system correctly and performs the computations correctly, it is necessary to validate the model and verify its accuracy. The importance of validation and verification and their definitions are discussed in Section 4.5.

6.5.1 Verifying the model

Verification of this model involves testing that all elements of the simulation model are behaving as expected, as follows:

- Observing the visual output of the simulation model when it is running slowly to check that the patients are flowing through it correctly;
- Running the model with fixed distributions to test that it performs the bookings as expected;
- When the visual logic is set up for each algorithm, running that algorithm in debugging mode to check that it is working as expected;
- For each algorithm, additional data are output to the spreadsheets to enable detailed checking of the order in which bookings are made during debugging;
- The booking spreadsheets for each day are checked to ensure correct adherence to the booking limits.

6.5.2 Validating the model

Validating the model involves ensuring that it provides a sufficiently accurate representation of the real system for its intended use. In this case, this is achieved by;

- Prior to building the model, hospital staff, including bookings clerks were interviewed to give us a clear understanding of the problem;
- Once the model was set up, it was discussed with a surgeon and other staff to check that it represents the system correctly;
- The results of the model are checked against data from the real system to ensure that they follow similar patterns.
- Results such as the percentage utilisation where checked with the surgeon to ensure that they were as expected.

These checks ensure that the model operates as expected and represents the real system, subject to the limitations listed in the next section. Some of the adaptions to the system ensure that the model will allow clear comparison of the intrinsic performance of the algorithms, for example, not allowing cancellations under any circumstances.

6.6 Assumptions and Limitations

During the construction of a model assumptions are made and there are areas of the real system that it is either not possible to model accurately (for example due to data limitations) or not desirable to model accurately due to the time required in comparison to their importance to the way in which the model is to be used.

The following assumptions and limitations must be born in mind when considering the results of this model:
- In the real system, depending on clinical necessity, emergency patients, for whom there is insufficient space in the theatre slot on the day they arrive, may be treated in overtime or wait until the next day. As only two days of the week are considered in the model and no data on the urgency of emergency patients is available, this cannot be accurately represented. Therefore, it is assumed that all emergency patients are treated on the day they arrive. In some cases, this will have resulted in longer sessions than would exist in the real system.
- The model does not allow cancellations or rebooking to occur. The surgeon, on whose data it is based, does not like cancelling and rearranging patients as this is disruptive to the patients, so he avoids doing this almost entirely. Thus, the model represents the real system. It is also preferable to test the algorithms

without allowing for cancellation and rebooking as this could mask poorly performing algorithms.

- The model does not included clinical cancellations or patients not attending for other reasons. This happens to a very limited extent and would reduce the theatre utilisation because some operations would not take place.

- As mentioned at the end of Section 6.2.3, it was not possible to obtain correct target dates for patients from the data, so the distribution of waiting times has been used instead. Since some patients will have been seen before their due dates this means that the times to due dates in the model are shorter than in reality for some patients. This will mean that the percentage of patients treated within targets from the model will be less than in reality and results should be interpreted with this in mind.

- It is assumed that it is desirable to give patients plenty of notice of their appointment dates, so that they can make necessary arrangements for having their operations. Nevertheless, algorithms that would only give 2 weeks' notice are considered. Shorter notice periods are not considered as 2 weeks is the shortest time to due date possible in the data input, so the only more urgent patients who will arrive less than 2 weeks before a slot are emergencies. Waiting for emergencies to be present in the system would require scheduling on the morning of the current day, which is gives to short notice period for elective patients.

- In the model, the difference between expected and actual start times is sampled from the distribution for turnaround times, which is reasonable because the distributions both average around 10 minutes and the anaesthetic times that significantly affect turnaround times will have a similar effect on the start of sessions.

- It is assumed that the scheduling rules are applied rigorously, which is desirable in order to test their performance. In reality it is likely that the person doing the bookings would make adjustments to allow for the variability in arrivals, any attempt to allow for this would be complicated and potentially mask poorly performing algorithms.

- No consideration is made of the effects on the number of patients in beds on the schedule. This is because it is very rare for ophthalmology patients to stay overnight, so this is not an important consideration when they are booked.

162

- It is also assumed that the patient case mix, inter arrival times, distribution of turnaround times and operation durations remain the same as for the period at which data was collected.

These assumptions and limitations should be considered when interpreting the results of the model. They have been discussed with the surgeon involved and are considered reasonable for the manner in which the model is intended to be used.

6.7 Results

This section contains the results of running the algorithms. All of the results are for running trials which compile data for 250 runs of the model with different random data. The average values for all performance measures over all of those trials are then computed. This avoids the possibility of the results being skewed by one algorithm working particularly well on one data set but not on others. The reasons for selecting 250 runs are discussed in Section 6.3.1.

The results are given for groups of algorithms, so that these can be compared. Then the best algorithms are compared directly at the end of the section. For each group of algorithms, there is a short description of the group followed by a graph showing the percentage of Emergency, Routine and Urgent patients treated on or before their due dates under each algorithm. On this graph, the higher the bars the better, as we want all patients to be treated on or before their due dates.

Furthermore, a graph for each group of algorithms is presented with a view to analysing how effectively urgent patients are prioritised, that is how fairly the algorithm respects due dates. This graph gives the average waiting times for cataract patients who have 4, 10 or 16 weeks to go until their due dates when they join the system, and the same applies for VR patients. If the patients have been treated fairly then the 4 weeks to due date patients will wait less time than the 10 weeks patients, who in turn will wait less than the 16 weeks patients. Also, if the majority of patients have been treated on time, then the averages will be less than the respective numbers of weeks.

After the graphs are tables giving other results, starting with the mean slot lengths (the average of the actual theatre time used each day) for both Tuesdays and Thursdays as

percentages of the actual slot length (theatre time available that day). The percentage of sessions on each day that overrun the allowable overtime is also given. Note that, the allowable overtime for Tuesdays is shorter than that for Thursdays, as the Tuesday slot is a morning session, so the values for the two days are not expected to be equal. In addition, is the utilization of the theatres as defined by the hospital, which is the percentage of the time spent performing operations. Lastly the tables contain the number of overtakes, which is the number of times that after a booking a patient that another patient with the same or later due date is booked earlier than the first patient.

6.7.1 Booking as Soon as Possible

The results that follow are for algorithms where patients are booked into the next slot which has sufficient space remaining for their expected operation durations; that is the algorithms described in Section 6.4.1. The results for these algorithms are shown in Figure 35, Figure 36 and Table 5.
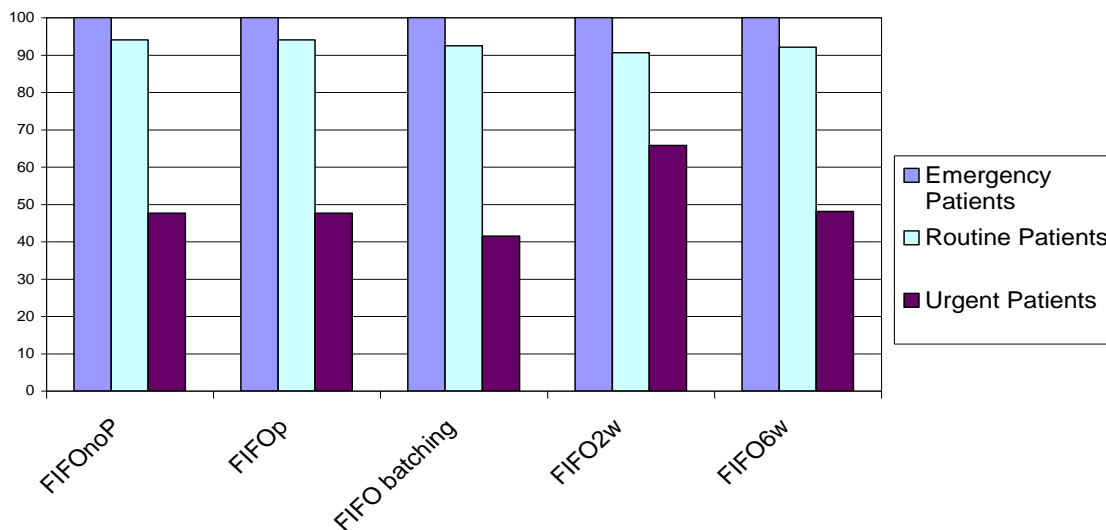
Figure 35: The percentage of each type of patient seen within their due dates for algorithms that book into the first available slot.

Figure 35 shows that prioritising the order in which patients are booked makes no difference if they are all booked straight away and that batching makes little difference. It also shows that when booking is delayed, more urgent patients are treated within their due dates, with the shorter booking period allowing more patients to be treated by their due dates. With a first in first out booking algorithm, booking only into the next 6

164

weeks makes a very small difference compared with booking at any point in the future, in terms of average waiting times for each type of patient.
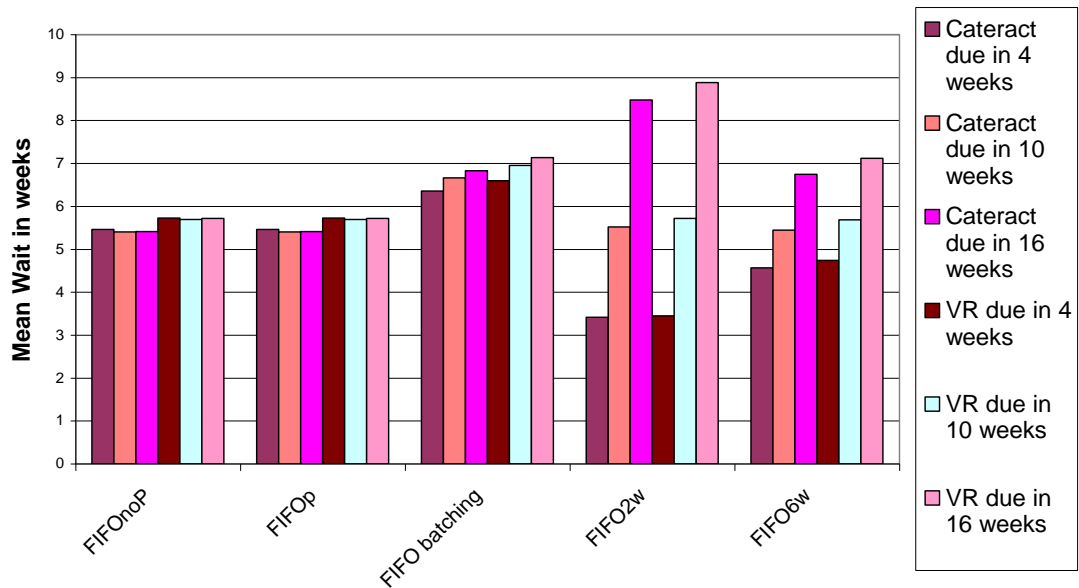
Figure 36: The waiting times for patients grouped by their initial time to due date for algorithms that book into the first available slot.

Figure 36 shows that the delayed algorithms do better at allowing urgent patients to be booked within their due dates and create a better spread of waiting times according to time remaining to due date.

Table 5: Other results for algorithms that book into the first available slot

| | Algorithms | | | | |
	FIFOnoP	FIFOp	FIFO batching	FIFO2w	FIFO6w
Mean slot length Tuesday %	101.4	101.4	101.5	101.4	101.4
Mean slot length Thursday %	111.8	111.8	111.6	111.6	111.7
% Tuesdays overrunning	29.5	29.5	29.5	29.6	29.7
% Thursdays overrunning	17.1	17.1	17.5	17.3	17.1
Mean Utilization on Tue %	71.4	71.4	71.4	71.5	71.4
Mean Utilization on Thur %	79.1	79.1	79.0	78.8	78.9
Number of Overtakes	800.1	800.1	767.2	524.0	664.6

Table 5 shows that there is little difference between the algorithms considered in terms of the use of the theatre slots and the number of overruns. It also shows little difference in the number of overtakes, except when booking is only allowed 2 weeks ahead in which case there are fewer overtakes. There are also fewer overtakes when only those

arriving in one day are batched and booked in due date order at the end of the week. This suggests that delay in booking allows greater fairness in terms of patients being booked in due date order.

6.7.2 Booking to Due Dates

The results that follow are for algorithms where the majority of patients are booked into the slot closest to their due date which has sufficient space remaining for their expected operation durations; that is the algorithms described in Section 6.4.2. The results for these algorithms are below in Figure 37, Figure 38 and Table 6.
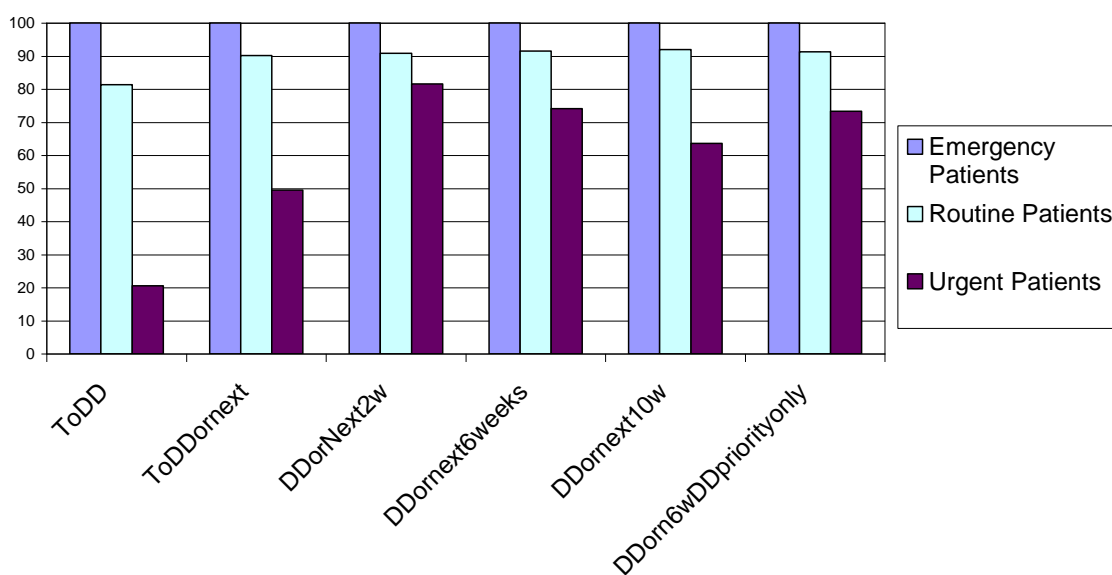
Figure 37: The percentage of each type of patient seen within their due dates for algorithms that book most patients as close as possible to their due dates.

Figure 37 shows that booking all patients as late as possible close to their due dates performs badly in terms of patients being seen by their due dates. Looking into the data in more detail reveals that this is due to some spaces in theatre being unused and thus pushing all of the operation dates further back than for FIFO. Booking to due date unless there is space left in the next slot, does slightly better than FIFO and as for the previous group of algorithms the delayed algorithms are significant improvements. Booking in order of due date priority followed with arrival order as a tie break does marginally worse than due date priority with expected operation duration as a tie break. Therefore, it is better to use the latter order for considering patients for booking.
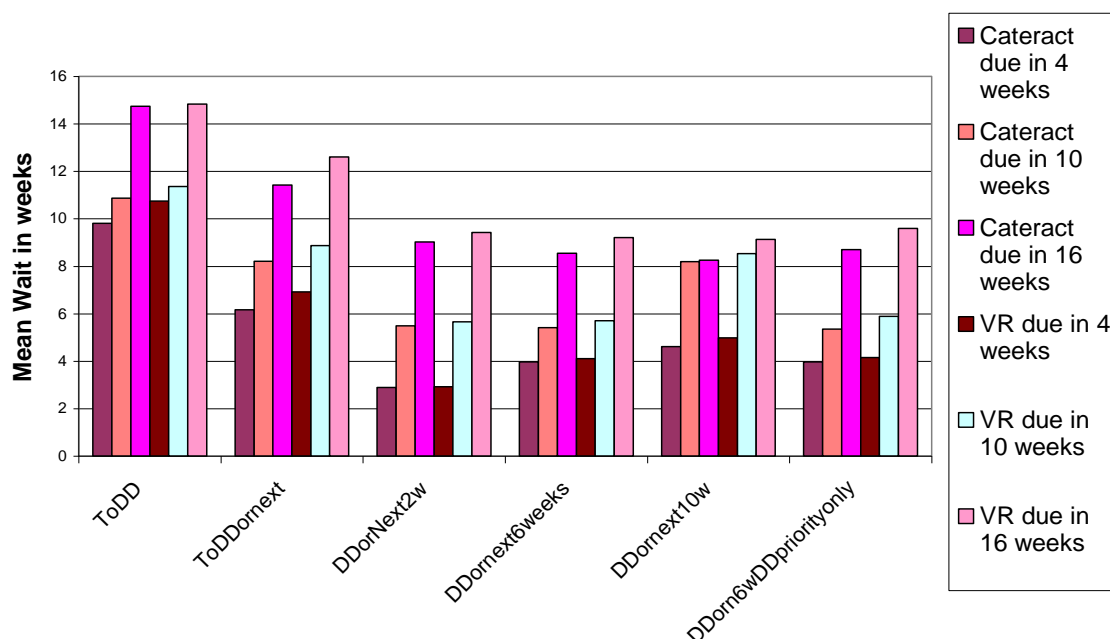
Figure 38: The waiting times for patients grouped by their initial time to due date for algorithms that book most patients as close as possible to their due dates.

Figure 38 shows that booking all patients as late as possible close to their due dates has patients waiting longer than the other algorithms in this section. It also shows that all of the algorithms involving booking most patients as close as possible to their due dates do well at giving shorter waiting times to higher priority patients.

Table 6: Other results for algorithms that book into most patients as close as possible to their due dates

		Algorithms				
	ToDD	ToDDor next	DDor next2w	DDor next6w	DDor next10w	Ddorn 6wDD priority
Mean slot length Tuesday %	99.3	100.3	101.5	101.2	100.8	101.1
Mean slot length Thursday%	115.6	113.5	111.6	111.9	112.5	112.1
% Tuesdays overrunning	28.2	28.3	29.6	29.4	29.0	29.3
% Thursdays overrunning	18.8	17.6	17.3	17.3	17.5	17.5
Mean Utilization on Tuesdays %	70.4	70.9	71.5	71.3	71.0	71.2
Mean Utilization on Thursdays %	80.6	79.7	78.8	79.1	79.5	79.2
No. of Overtakes	1637.3	2405.8	464.2	1539.5	2817.9	1599.3

167

Similarly to Table 5, Table 6 shows that there is little difference between the algorithms considered in terms of the use of the theatre slots and the number of overruns. It is also similar to Table 5 in that the most significant difference in the number of overtakes occurs when booking is only allowed 2 weeks ahead in which case there are significantly fewer overtakes. The rise in the number of overtakes compared to Table 5 is a result of some patients being booked before those with the same due date because they arrive later and all of the slots close to their due dates have been booked.

6.7.3 Booking to a Percentage of Due Dates

The results that follow are for algorithms where the majority of patients are booked into the slot closest to a percentage of the time from their arrival to their due date; that is the algorithms described in Section 6.4.3. The results for these algorithms are presented below in Figure 39, Figure 40 and Table 7.
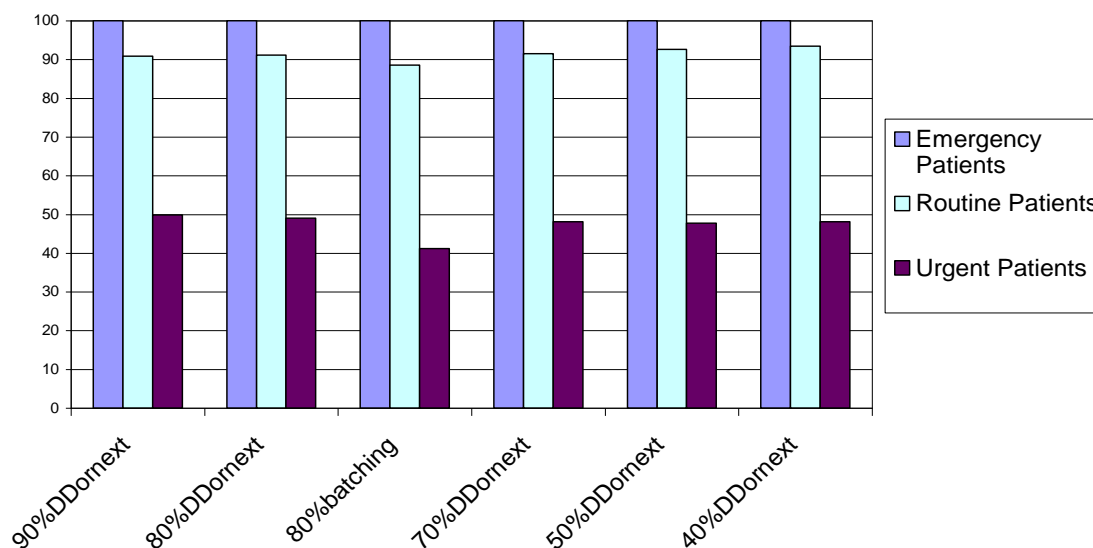
Figure 39: The percentage of each type of patient seen within their due dates when booking the majority of patients as close as possible to a percentage of their due dates.

Figure 39 shows a slight decrease in the number treated by their due dates as the percentage of target in which they are booked decreases for urgent patients, with a slight increase in the number of routine patients treated within their due dates. This suggests that booking to the full due date is better in terms of treating urgent patients within their targets.
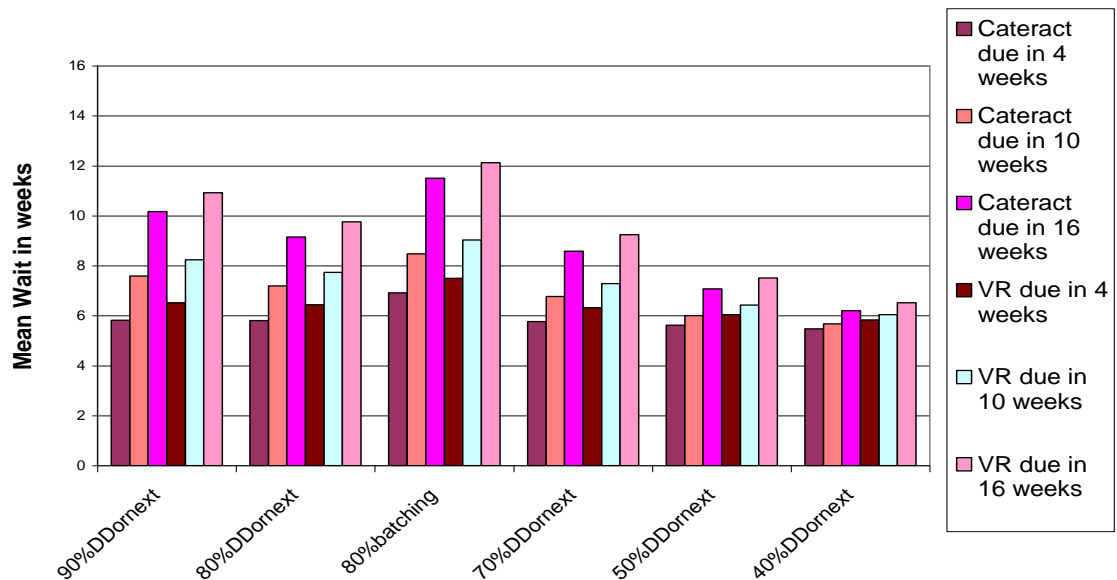
Figure 40: The waiting times for patients grouped by their initial time to due date when booking most patients as close as possible to a percentage of their due dates.

Figure 40 shows that as the percentage of due date decreases so does the average waiting time, but the prioritisation of urgent patients is less apparent. This agrees with the results shown in Figure 39 that urgent patients do less well as the percentage of due date to which patients are booked decreases.

Table 7: Other results for algorithms that book into most patients as close as possible to a percentage of their due dates

	Algorithms					
	90%DD or next	**80%DD ornext**	**80% batching**	**70%DD ornext**	**50%DD ornext**	**40%DD ornext**
Mean slot length Tues %	100.4	100.4	100.5	100.5	100.7	100.9
Mean slot length Thur %	113.4	113.3	113.2	113.0	112.7	112.3
% Tuesdays overrunning	28.4	28.5	28.5	28.5	29.0	29.1
% Thursdays overrunning	17.4	17.7	17.6	17.7	17.2	17.2
Mean Utilization on Tue %	70.9	70.9	71.1	71.0	71.1	71.2
Mean Utilization on Thur %	79.7	79.7	79.3	79.5	79.3	79.2
Number of Overtakes	2025.2	1724.5	1338.4	1507.9	1149.1	957.9

Table 7, like the previous tables of results, shows that there is little difference between the algorithms considered in terms of the use of the theatre slots and the number of overruns. It is also shows a reduction in the number of overtakes as the percentage of due date to which patients are booked decreases.

6.7.4 Changing the Booking Limits for Theatre Slots

The results that follow are for algorithms where the majority of patients are booked into the slot closest to 40% percentage of the time from their arrival to their due date, with variations on the amount of the theatre slots that are available for booking. Specifically, we consider the algorithms described at the end of Section 6.4.3. In the algorithm reducing the time available for bookings on Tuesdays, the time available is reduced by 5 minutes. Keeping this reduction in the time available on Tuesdays and additional 5 minutes is made available on each slot on a Thursday for the Tuesday reduced Thursday increased algorithm, The results for these algorithms are below in Figure 41, Figure 42 and Table 8.
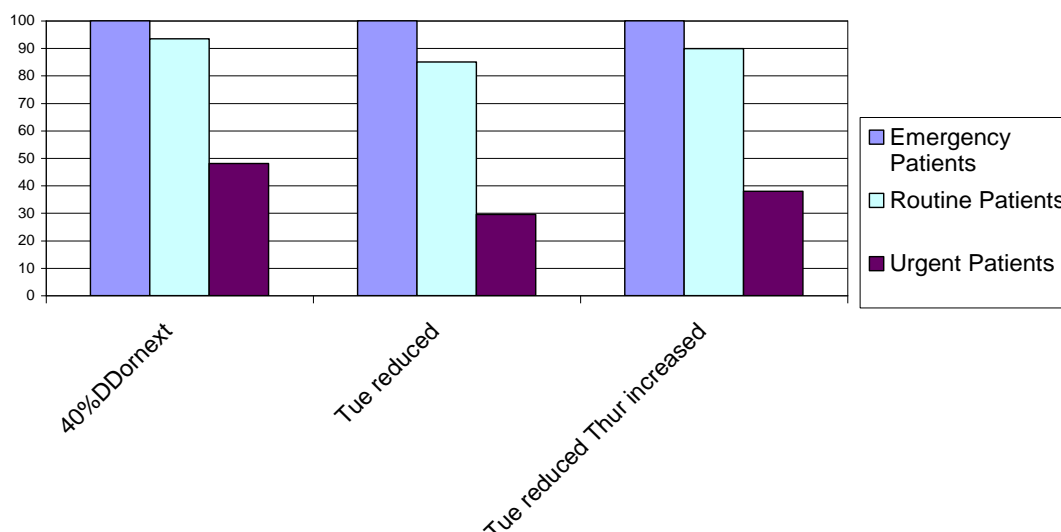
Figure 41: The effects on the percentage of patients treated by their due dates of changing the amount of each theatre slot available for booking into.

Figure 41 shows that, as might be expected, reducing the amount of time available for bookings reduces the number of patients treated by their due dates and increasing the time available has the opposite effect. It should be noted that as there are only slots every other Thursday the effects of changes to Thursdays are not as significant as the effects of changes to Tuesdays.
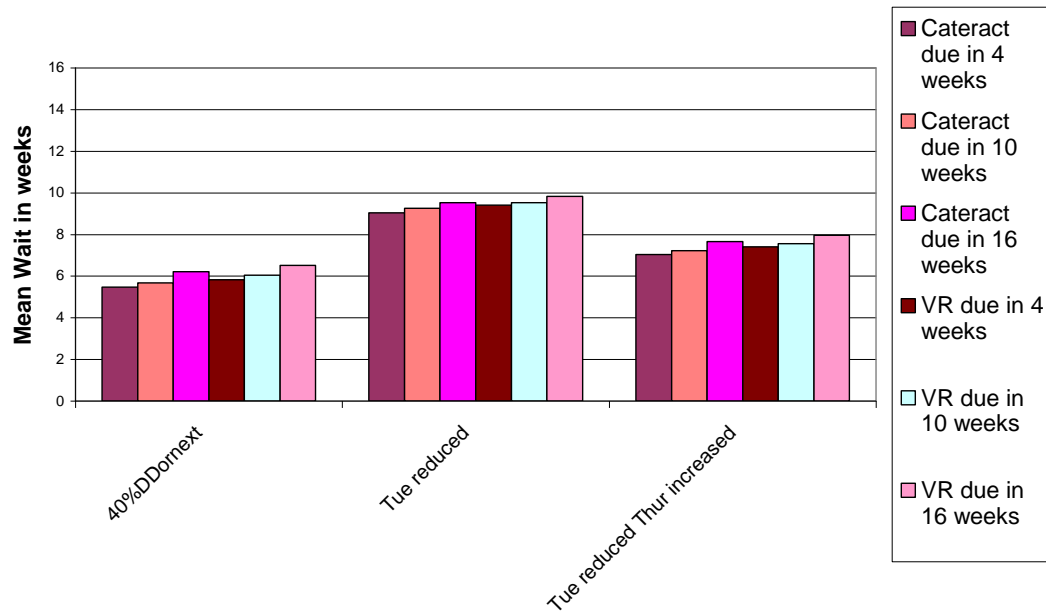
Figure 42: The effects on fairness of changing the amount of each theatre slot available for booking into.

Figure 42 shows that changes to the amount of time available for booking affects waiting times, but not the fairness towards more urgent patients.

Table 8: The effects on the other results of changing the amount of each theatre slot available for booking into.

	Algorithms		
	40%DDornext	**Tuesday reduced**	**Tuesday reduced Thursday increased**
Mean slot length Tuesday %	100.9	99.1	98.7
Mean slot length Thursday %	112.3	114.4	116.5
% Tuesdays overrunning	29.1	26.3	26.1
% Thursdays overrunning	17.2	18.4	20.9
Mean Utilization on Tue %	71.2	69.9	69.7
Mean Utilization on Thur %	79.2	80.6	82.1
Number of Overtakes	957.9	941.4	944.6

Table 9 shows the effects that changing the amount of theatre slots that are available for booking has on the use of those theatre slots and the percentages of overruns. One might expect that reducing the amount of each slot available for booking would create significant reductions in the slot lengths and number of sessions overrunning; however, our results show that the effect is quite small. This is due to the increase in waiting list resulting in more patients to choose from and therefore in a closer fit to the slot size

171

available. The table also shows that such changes have little effect on the numbers of overtakes occurring.

6.7.5 Booking to Exact Fits or Due Dates

The results that follow are for algorithms where the majority of patients are booked into the available slot closest to their due date, except that booking into a slot where that the patient fills the remaining available theatre time as an exact fit is included where possible. As before, variations on the number of weeks into which booking is allowed are explored, as is the tie break if there is more than one exact fit; these are some of the algorithms described in Section 6.4.4. The results for these algorithms are below in Figure 43, Figure 44 and Table 9.
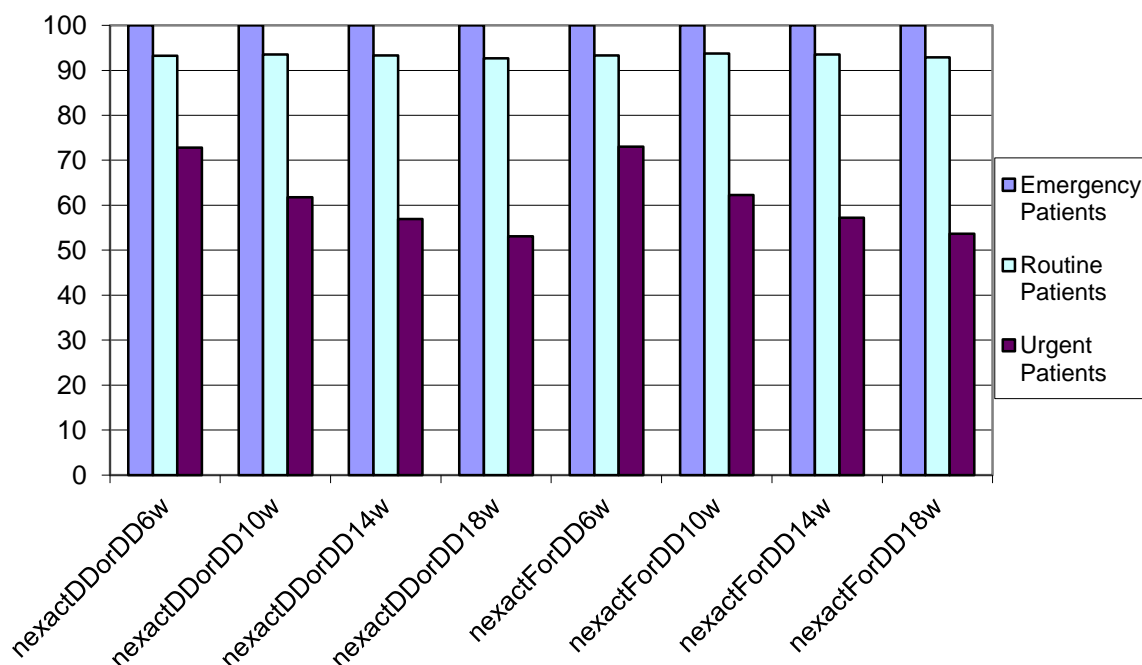
Figure 43: The percentage of each type of patient seen within their due dates for algorithms that book into slot exactly filled by the patients expected operation duration.

Figure 43 shows that algorithms that book into a slot such that the patients expected operation time fills the time available for booking into that slot if such a slot exists actually do better in terms of the number of urgent patients seen by their due date if bookings are considered over a shorter horizon. It also shows that breaking ties for exact fits by considering the exact fit that is closest to the due date or closest to the current time makes only a slight if any difference.

172
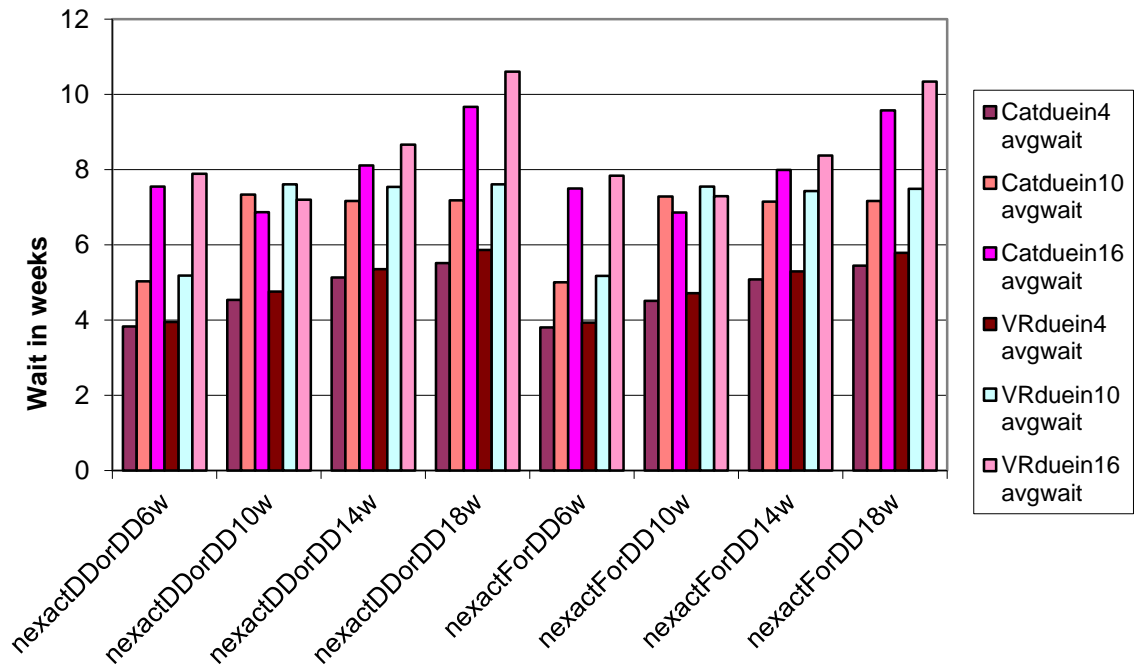
Figure 44: The waiting times for patients grouped by their initial time to due date for algorithms that book to exact fits where such exist.

Figure 44 shows that breaking ties for exact fits by selecting the slot closest to due date creates a slightly better spread of waiting times based on initial time to due date. It also shows that considering booking into the full 18 weeks of the longest due dates increases such fairness.

Table 9 shows little difference between these algorithms in terms of other results, except that, as before, the delayed algorithms reduce the number of overtakes.

Table 9: Other results for the algorithms that book into exact fits if they exist

	Algorithms							
	nexactDDorDD6w	nexactDDorDD10w	nexactDDorDD14w	nexactDDorDD18w	nexactForDD6w	nexactForDD10w	nexactForDD14w	nexactForDD18w
Mean slot length Tuesday %	100.8	100.1	99.7	99.5	100.8	100.1	99.8	99.5
Mean slot length Thursday %	113.0	114.2	115.0	115.7	113.0	114.2	115.0	115.5
% Tuesdays overrunning	29.2	28.5	28.4	27.9	28.9	28.6	28.3	28.2
% Thursdays overrunning	18.2	18.8	19.1	19.3	18.3	18.7	19.2	19.2
Mean Utilization Tue %	71.0	70.5	70.4	70.7	71.0	70.5	70.5	70.7
Mean Utilization Thu %	79.8	80.8	80.9	80.4	79.8	80.8	80.8	80.2
Number of Overtakes	1813	2958	3188	2911	1819	2968	3226	2970

6.7.6 Booking to Exact Fits or Most Empty, Varying the Search Order

The results that follow are for algorithms where patients are booked into a slot they fit exactly or the emptiest slot before their due date; specifically, we consider more of the algorithms described in Section 6.4.5. The differences between these algorithms are in the order in which the slots are considered. In the tests above, if there were two slots equally empty or both exact fits, then the one closest to due date was chosen. In the first of these algorithms, the exact fits are closest to due date but the most empty are the closest to simulation time if there are such ties. For the second, the most empty tie break is closer to due date and for the third both tie breaks are closest to the current time. The results for these algorithms are below in Figure 45, Figure 46, Table 10 and Table 14.
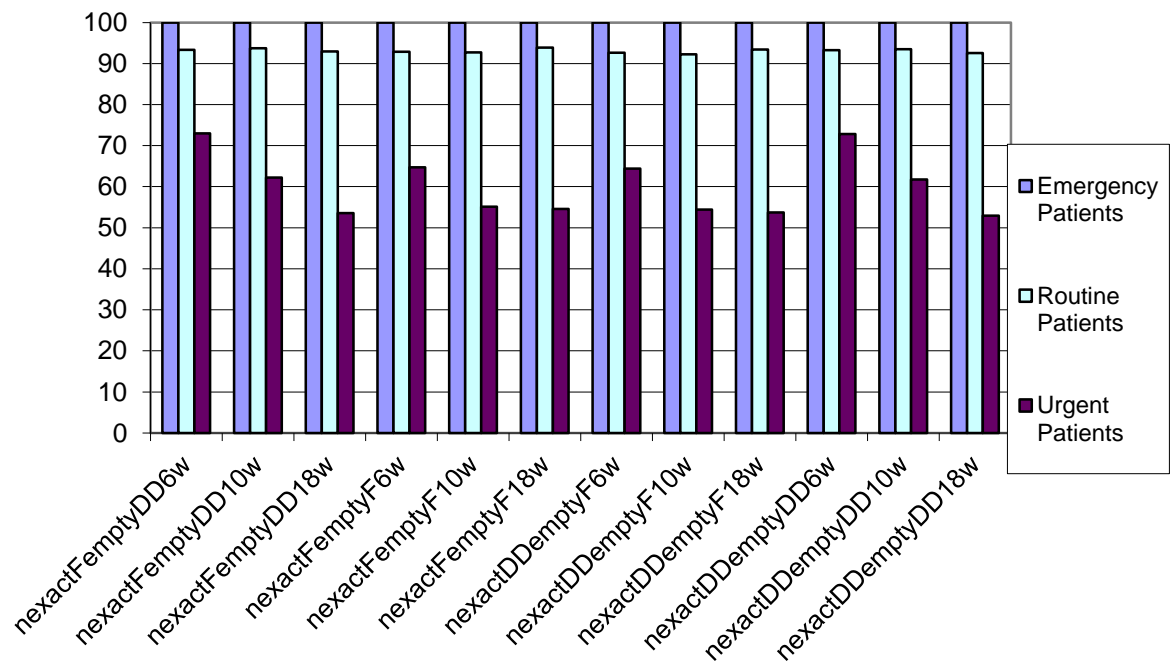
Figure 45: The percentage of each type of patient seen within their due dates for algorithms that book into the emptiest slot within due date or an exact fit, with potential slots considered in different orders.

Figure 45 shows that the algorithms considering booking into the emptiest slot in due date order do better than those considering the first available slot for this aspect, but that the gap reduces as a longer booking horizon is considered. Comparing these values with those in Figure 43, reveals that considering the emptiest slot rather than the slot closest to the due date after looking for exact fits only slightly changes the numbers or urgent patients who are seen before or on their due dates.
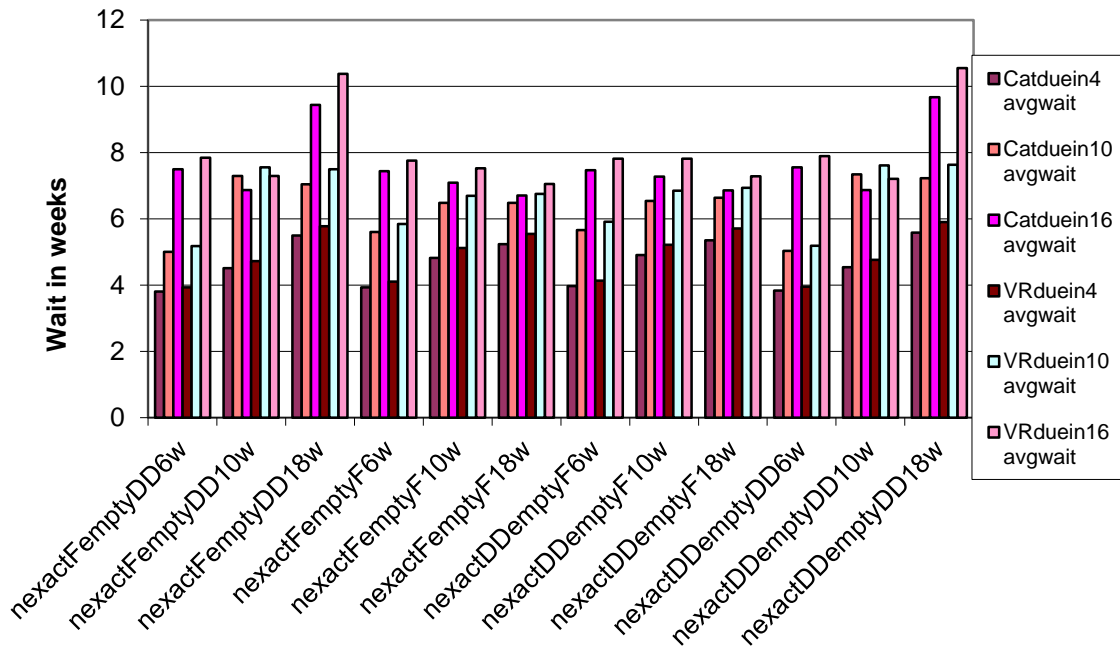
Figure 46: The waiting times for patients grouped by their initial time to due date for algorithms that book into the most empty slot within due date or an exact fit, with potential slots considered in different orders.

Figure 46 shows that the greater the extent to which booking is done close to the due date the wider the spread of waiting times compared for those with different initial times to due date. It also shows that the spread is greater, but with longer waiting times, if the booking horizon is longer.

Table 10: Other results for algorithms that book into the emptiest slot within due date or an exact fit, with potential slots considered in different orders.

	Algorithms					
	nexactFemptyDD6w	nexactFemptyDD10w	nexactFemptyDD18w	nexactFemptyF6w	nexactFemptyF10w	nexactFemptyF18w
Mean slot length Tuesday %	100.8	100.1	99.5	100.9	100.6	100.6
Mean slot length Thursday %	113.0	114.2	115.7	112.4	112.9	113.1
% Tuesdays overrunning	28.9	28.6	28.1	29.2	28.9	28.8
% Thursdays overrunning	18.3	18.7	19.4	17.9	18.0	18.0
Mean Utilization on Tuesdays %	71.0	70.5	70.7	71.3	71.0	71.0
Mean Utilization on Thursdays %	79.8	80.8	80.3	78.9	79.4	79.5
No. of Overtakes	1819	2968	2929	1813	2003	2054

Table 11: Other results for algorithms that book into the emptiest slot within due date or an exact fit, with potential slots considered in different orders, continued.

	Algorithms					
	nexactDDemptyF6w	nexactDDemptyF10w	nexactDDemptyF18w	nexactDDemptyDD6w	nexactDDemptyDD10w	nexactDDemptyDD18w
Mean slot length Tuesday %	100.8	100.6	100.6	100.8	100.1	99.5
Mean slot length Thursday %	112.5	113.0	113.1	113.0	114.2	115.6
% Tuesdays overrunning	29.4	28.8	28.8	29.2	28.5	28.1
% Thursdays overrunning	17.7	18.2	18.0	18.2	18.8	19.4
Mean Utilization on Tuesdays %	71.2	71.0	70.9	71.0	70.5	70.7
Mean Utilization on Thursdays %	79.1	79.6	79.7	79.8	80.8	80.3
No. of Overtakes	1808	1970	2005	1813	2958	2854

Table 10 and 11 shows only minor differences between the algorithms. This suggests that, while the tie breaks have an effect, it is relatively minor.

6.7.7 Comparing the Best Algorithms

This section compares the algorithms that performed best over the previous sections, with the results displayed in Figure 47, Figure 48, Figure 49 and Tables 12 and 13.

Figure 47: Comparison of the percentage of each type of patient seen within their due dates for the best performing algorithms.

Figure 47 shows that in terms of the percentage of patients (particularly urgent patients) treated within their due dates delayed booking algorithms do best. To test the significance of the variations Figure 48 gives the 95% confidence intervals for the number of patients treated within target for the same set of algorithms as above. This shows clearly that the difference made by booking only a limited number of weeks into the future is significant, but that the difference reduces as the number of weeks increases. It also shows that the differences between the algorithms are more significant when the booking horizon is more limited.
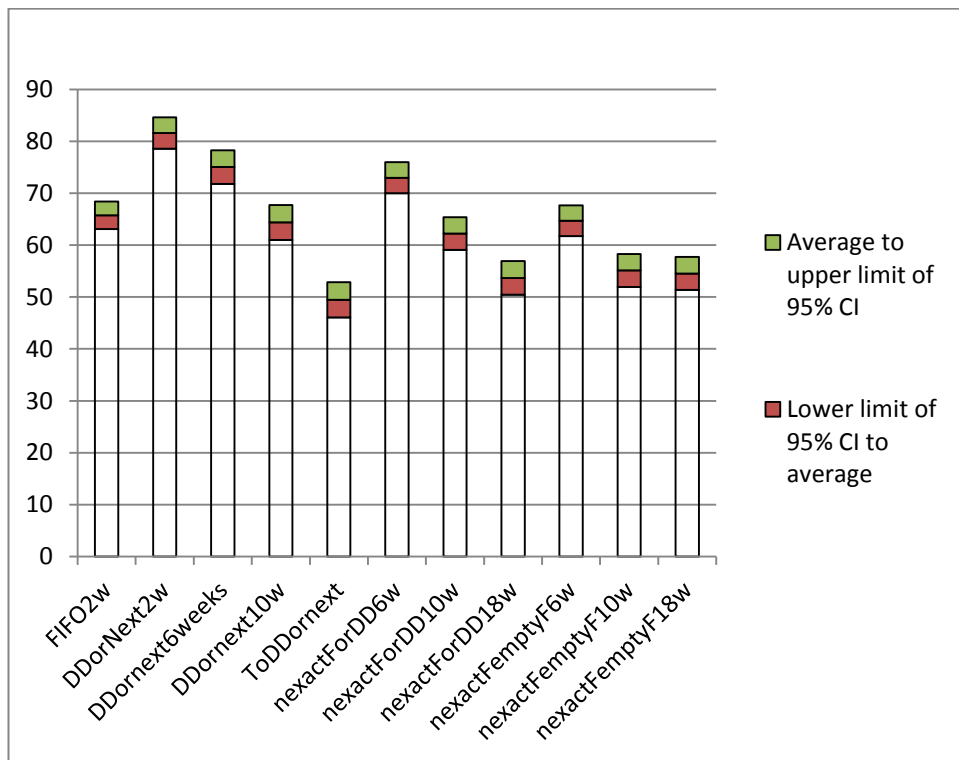
Figure 48: Confidence intervals for the percentage of urgent patients treated within target dates for the best performing algorithms
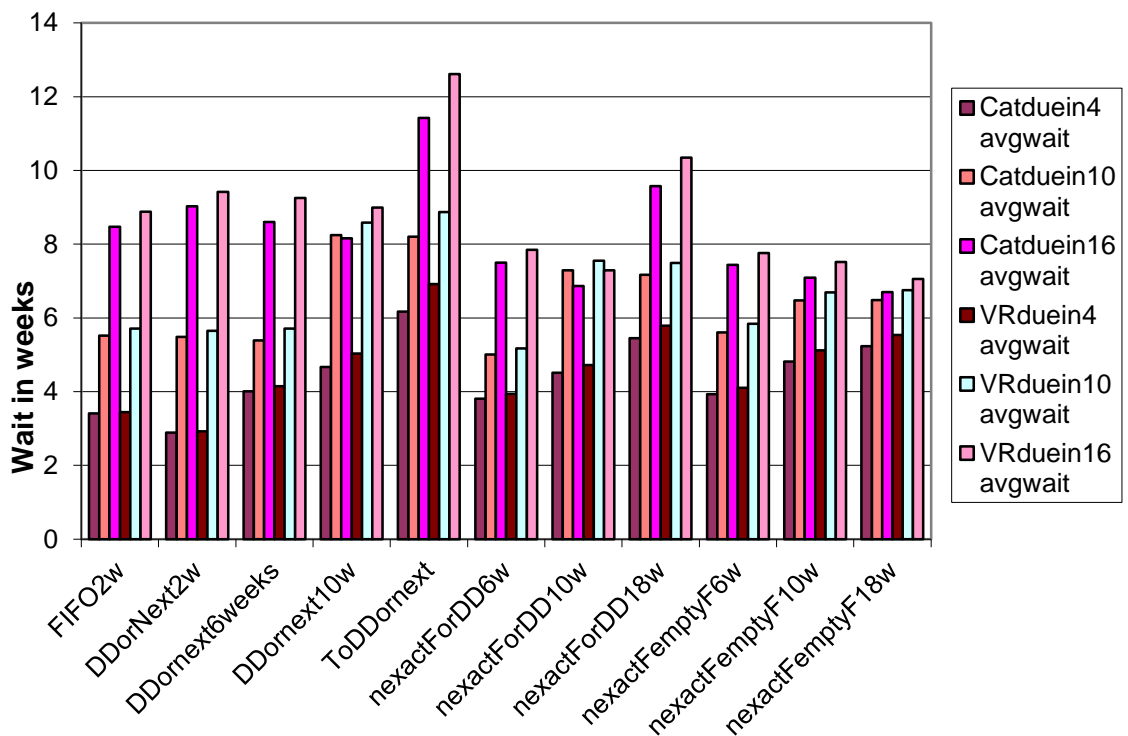
Figure 49: The waiting times for patients grouped by their initial time to due date for the best performing algorithms.

Figure 49 shows that all of the algorithms selected for consideration in this section do well in terms of prioritising urgent patients, particularly when compared to just booking into the next available slot as soon as the patient enters the waiting list.

Table 12: Other results for the best algorithms.

	Algorithms				
			DDornext		
	FIFO2w	DDorNext2w	6weeks	DDornext10w	ToDDornext
Mean slot length Tuesday %	101.4	101.5	100.9	100.3	100.3
Mean slot length Thursday %	111.6	111.6	112.3	113.6	113.5
% Tuesdays overrunning	29.6	29.6	29.2	28.4	28.3
% Thursdays overrunning	17.3	17.3	17.4	17.9	17.6
Mean Utilization on Tuesdays %	71.5	71.5	71.1	70.7	70.9
Mean Utilization on Thursdays %	78.8	78.8	79.1	79.5	79.7
No. of Overtakes	524	464	1581	2924	2406

Table 13: Other results for the best algorithms, continued.

	Algorithms					
	nexactFo rDD6w	nexactFo rDD10w	nexactFo rDD18w	nexactFe mptyF6w	nexactre mptyF10 w	nexactFe mptyF 18w
Mean slot length Tuesday %	100.8	100.1	99.5	100.9	100.6	100.6
Mean slot length Thursday %	113.0	114.2	115.5	112.4	112.9	113.1
% Tuesdays overrunning	28.9	28.6	28.2	29.2	28.9	28.8
% Thursdays overrunning	18.3	18.7	19.2	17.9	18.0	18.0
Mean Utilization on Tuesdays %	71.0	70.5	70.7	71.3	71.0	71.0
Mean Utilization on Thursdays %	79.8	80.8	80.2	78.9	79.4	79.5
No. of Overtakes	1819	2968	2970	1813	2003	2054

Table 15 and 13 show that as discussed previously there is little to choose between the algorithms in terms of theatre utilisation, use and overruns.   There are significant differences when the number of overtakes are considered, so if this is an important criterion, then the algorithms with the most delayed bookings clearly do best.

Overall these results show that there is not a great deal of difference between the best algorithms. The relative importance of the performance measures as well as the constraints on the booking system should be considered in choosing between them. Delaying the decision to book patients as close as possible to the surgery taking place, improves the performance of the booking algorithms, but the extent to which this is possible is limited by the need to give patients time to prepare for their operations. Booking patients as late as their due dates allow, leaves spaces into which more urgent patients can be booked and this increase the proportion of urgent patients treated on time. As the period of time between booking and surgery taking place increases then looking for slots which will be filled exactly by a patients expected operation duration becomes effective in terms of treating patients on time, but it does reduce some fairness measures.

6.8 Exploring Data Variations

The previous section determined which algorithms should be considered for one particular surgeons' case mix. This section explores if the same results hold for variations from this case mix. In order to do this, we use the fitted distributions from Section 6.3 to explore any changes to the results, before going on to vary these distributions and observe the effects on the comparative results of the algorithms considered previously.

6.8.1 Testing algorithms with fitted distributions

The distributions described in the previous section are used in the simulation model in this section testing the best algorithms from Section 6.7.7 to see if the fitted distributions have affected the comparative effects of the algorithms. The results are set out in the same format as those in Section 6.7.

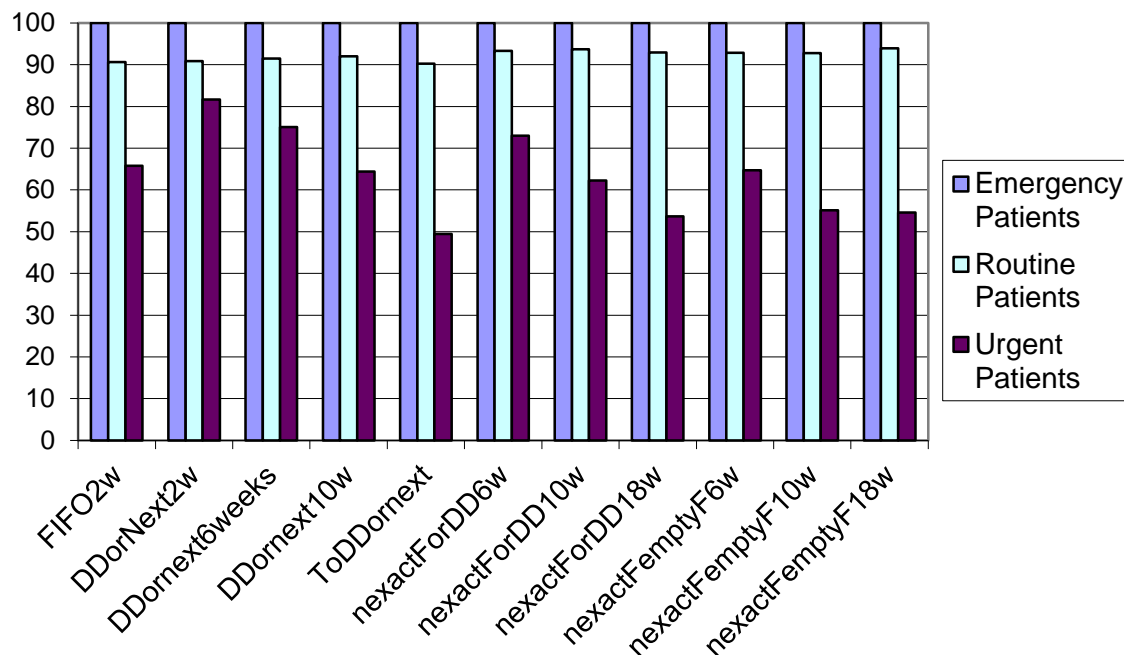**Figure 50: Comparison of the percentage of each type of patient seen within their due dates for the best performing algorithms, when applied to the fitted data.**

This shows a similar pattern to Figure 47, with the delayed booking algorithms performing better in terms of the percentage of urgent patients seen within their due dates. However, the proportion of urgent patients treated in target is higher and the effects of delaying booking are less significant.

Comparing the combined durations of all operations for sample data sets for the original and fitted distributions reveals that for the data sampled from the fitted distributions a smaller volume of patients were generated. This is also reflected in a change of traffic intensity as shown by the mean slot lengths falling below 100% in Table 14. This means that there were fewer hours of operations to fit in the theatres and explains why more patients could be treated within their due dates for the new data. The reduced differences between the algorithms is also to be expected because there is less scope for improvement when there is more flexibility to fit patients into theatre slots.

Figure 51: The waiting times for patients grouped by their initial time to due date for algorithms for the best performing algorithms, when applied to the initial variation of the fitted data.

Figure 51 shows that, as before, the algorithms where patients are booked close to their due dates are fairer in terms of more urgent patients having shorter waiting times.

Table 14: Other results for the best algorithms, when applied to the fitted data.

	Algorithms									
	FittedFIFOp	FittedFIFO2w	FittedFIFO6w	FittedFIFO10w	FittedFIFO18w	fittedDDornext1w	fittedDDornext2w	fittedDDornext6w	fittedDDornext10w	fittedDDornext18w
Mean slot length Tuesday %	83.5	83.5	83.5	83.4	83.5	87.4	87.3	86.5	85.6	85.1
Mean slot length Thursday %	92.8	92.8	92.7	92.8	92.8	91.2	91.2	92.4	94.0	94.9
% of Tuesdays overrunning	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
% of Thursdays overrunning	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Mean Utilization on Tues %	67.5	67.6	67.6	67.6	67.5	71.3	71.2	70.5	69.8	69.9
Mean Utilization on Thurs %	75.5	75.4	75.4	75.5	75.5	74.2	74.1	75.4	76.6	76.2
Number of Overtakes	951	674	897	942	951	561	526	1905	3447	3071

Table 15: Other results for the best algorithms, when applied to the fitted distributions, continued.

	Algorithms							
	fittednexactForDD2w	fittednexactForDD6w	fittednexactForDD10w	fittednexactForDD18w	fittednexactFemptyF2w	fittednexactFemptyF6w	fittednexactFemptyF10w	fittednexactFemptyF18w
Mean slot length Tuesday %	85.2	83.7	82.4	81.5	84.1	83.8	83.8	83.8
Mean slot length Thursday %	89.9	91.9	94.1	95.7	91.6	91.6	91.6	91.6
% of Tuesdays overrunning	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
% of Thursdays overrunning	0.1	0.1	0.1	0.1	0.1	0.0	0.1	0.0
Mean Utilization on Tues %	68.9	67.5	66.3	65.9	68.0	67.8	67.8	67.8
Mean Utilization on Thurs %	73.2	75.3	77.4	77.9	74.6	74.6	74.6	74.6
Number of Overtakes	582	2062	3446	3406	623	812	830	832

The results given Figure 50 Figure 51, Table 14 and Table 15 are better than those for the original data, for the reasons explained above. As for the previous data the only result that varies significantly is the number of overtakes with the delayed booking algorithms again being fairer on this aspect of the problem.

Overall these results demonstrate that even when there is greater flexibility in the fit of operating times to theatre time available, the algorithms identified when sampling from the empirical data do best. The difference made by selecting one algorithm over another is reduced, but using the algorithms suggested when using the first data set would still be good choices.

6.8.2 Testing algorithms with a variation on fitted distributions

As discussed in the assumptions and limitations (Section 6.6) using the time from decision to treat to surgery to gain the distribution of time to due date, is likely to have resulted in some due dates being tighter than they should be. This is because some patients will have been treated before their due dates. In order to see if this may have an effect on the results the this variation on the fitted data, included increasing the due dates by one week whilst maintaining a maximum of 18 weeks, it also adds 2 minutes and 30

seconds to the operation durations to increase bring the volume of surgery closer to the original data. The results are given in Figure 52 and Figure 53.
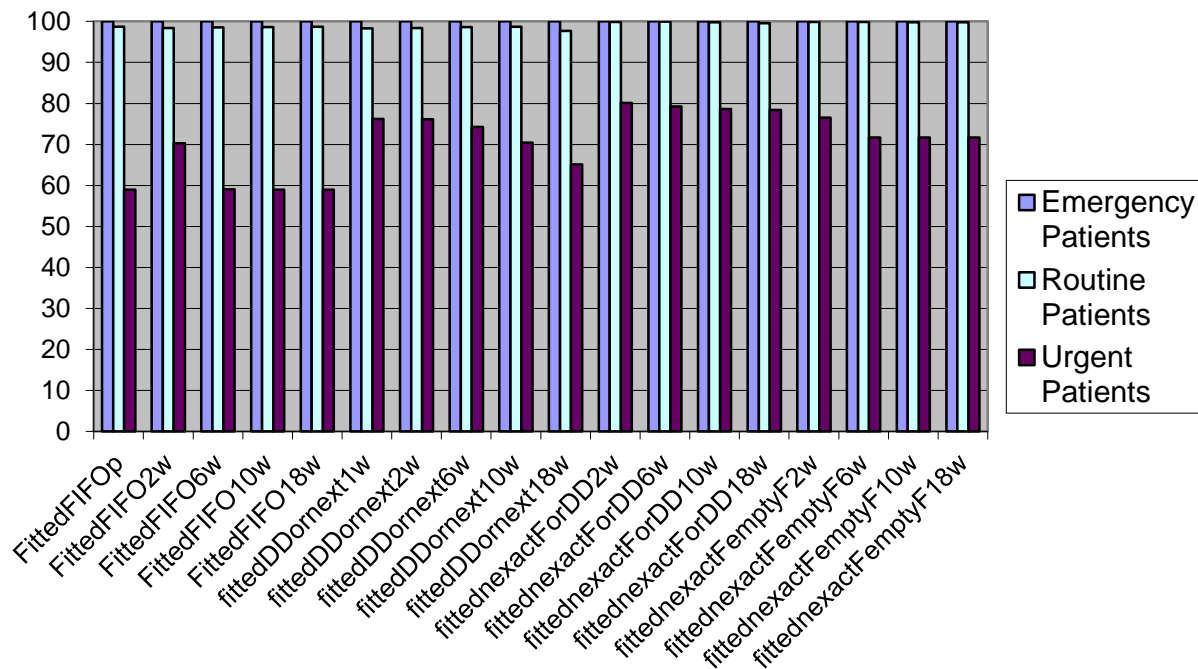

Figure 52: Comparison of the percentage of each type of patient seen within their due dates for the best performing algorithms, when applied to the second variation of the fitted data.

Figure 52 is very similar to Figure 50, except that the number treated in target has increased, indicating that the change in the data does not change the comparisons drawn previously between the algorithms.
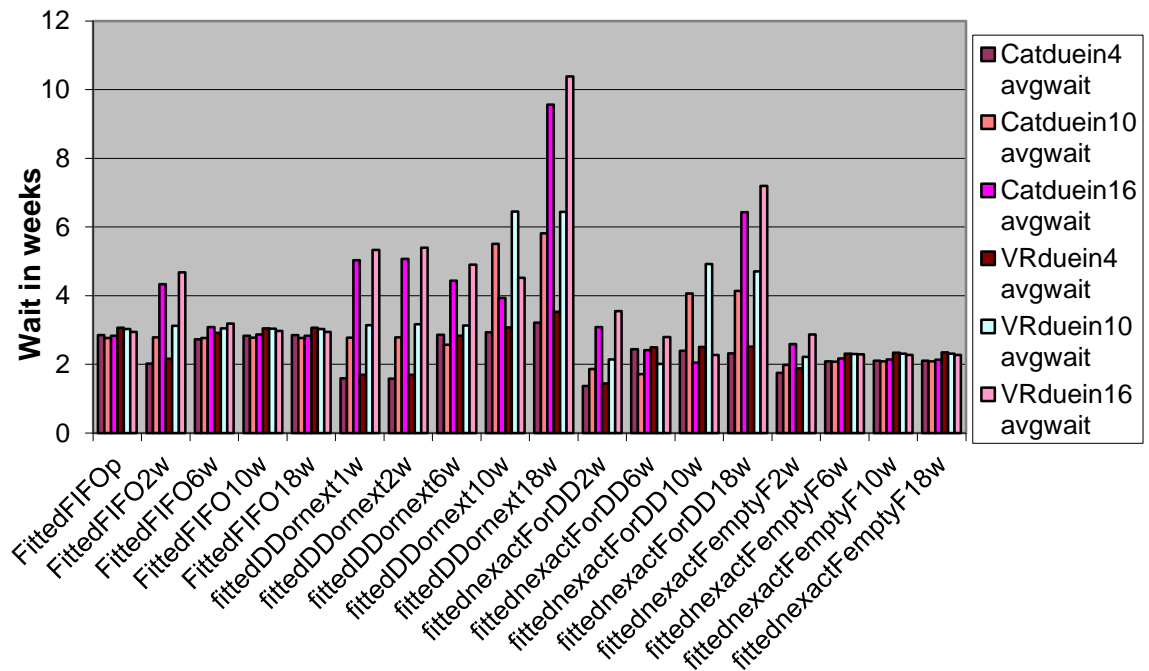
Figure 53: The waiting times for patients grouped by their initial time to due date for algorithms for the best performing algorithms, when applied to the second variation on the fitted data.

Figure 53 demonstrates that again the variation of the data has not affected the comparisons between the algorithms. This is also true for the other results shown in the tables in Appendix C.

6.8.3 Final variations on algorithms

Thus far, the algorithms that have consistently performed best have been those where booking is delayed until just a few weeks before each theatre slot. However, delaying booking as much as possible contradicts the desirable target of booking patients at the time when the decision that they need surgery is taken. This suggests that it is necessary to choose between giving patients as much warning as possible of the date of their surgery and meeting due date targets for all patients.

This raises another possibility: could some patients be booked at the time of decision to treat while others are booked at shorter notice. The advantage to patients of the latter is that they could experience shorter waiting times. Thus, some (rather than all) of the routine patients could have delayed booking so that they are only booked a fixed number of weeks in advance to create additional flexibility in the system. However, if some patients are being booked on shorter notice than others, should some of the available time in the operating theatre be held back to become available for booking only when those patients become available for booking?

In order to fully explore these questions, we ran further simulation trials, using the same algorithm nexactFemptyF as it did marginally better than the other algorithms when booking 18 weeks ahead for most of the data variations. To this algorithm, additional code was added to assign a proportion of those whose due dates are over 10 weeks from their arrival dates to be held back for booking on shorter notice. The proportions held back in this way are 0.1 (10%), 0.2 (20%), 0.3 (30%), 0.4 (40%) and 0.5 (50%), and these patients are only allowed to be booked for treatment up to 2, 4 or 6 weeks from the time of booking. Also, consideration is given to keeping back some of theatre time available in each slot to be available for booking only 2, 4 of 6 weeks before the slot occurs. The amounts of time tested for holding back in this way were 15, 30 and 45

minutes. All combinations of these 3 variables were tested and the abbreviated names of the algorithms are written as follows 0.4on6weekshold30, where 0.4 is the proportion of patients who will be booked on shorter notice, the maximum time ahead that they will be booked is 6 weeks and 30 minutes of theatre time will be held back to be booked at most 6 weeks before the theatre slot occurs. The results of these tests are given in the series of graphs below.

Figure 54 shows that the highest percentages of all types of patient treated within due dates is achieved when a larger section of each theatre slot is only made available for booking a few weeks before slot will take place. Keeping a proportion of patients to be booked only a limited number of weeks in advance only improves the percentage in target if only a small amount of time is restricted for booking close to when it takes place.

Keeping a section of each theatre slot back to be booked only a few weeks before it takes place is increasing the numbers of routine patients treated within their targets as well as the urgent patients. For the urgent patients, it is apparent that the increase in numbers treated by their due dates is due to slots still being available within the next few weeks when the most urgent patients arrive. Routine patients who arrive after a period of above average routine arrivals when all of the slots near their due date have been booked, would under other algorithms have to be booked after their due dates. When some theatre time has been kept back for booking, these routine patients can be booked into that time and treated with limited notice of their operation dates (or if this is inconvenient could choose later dates).
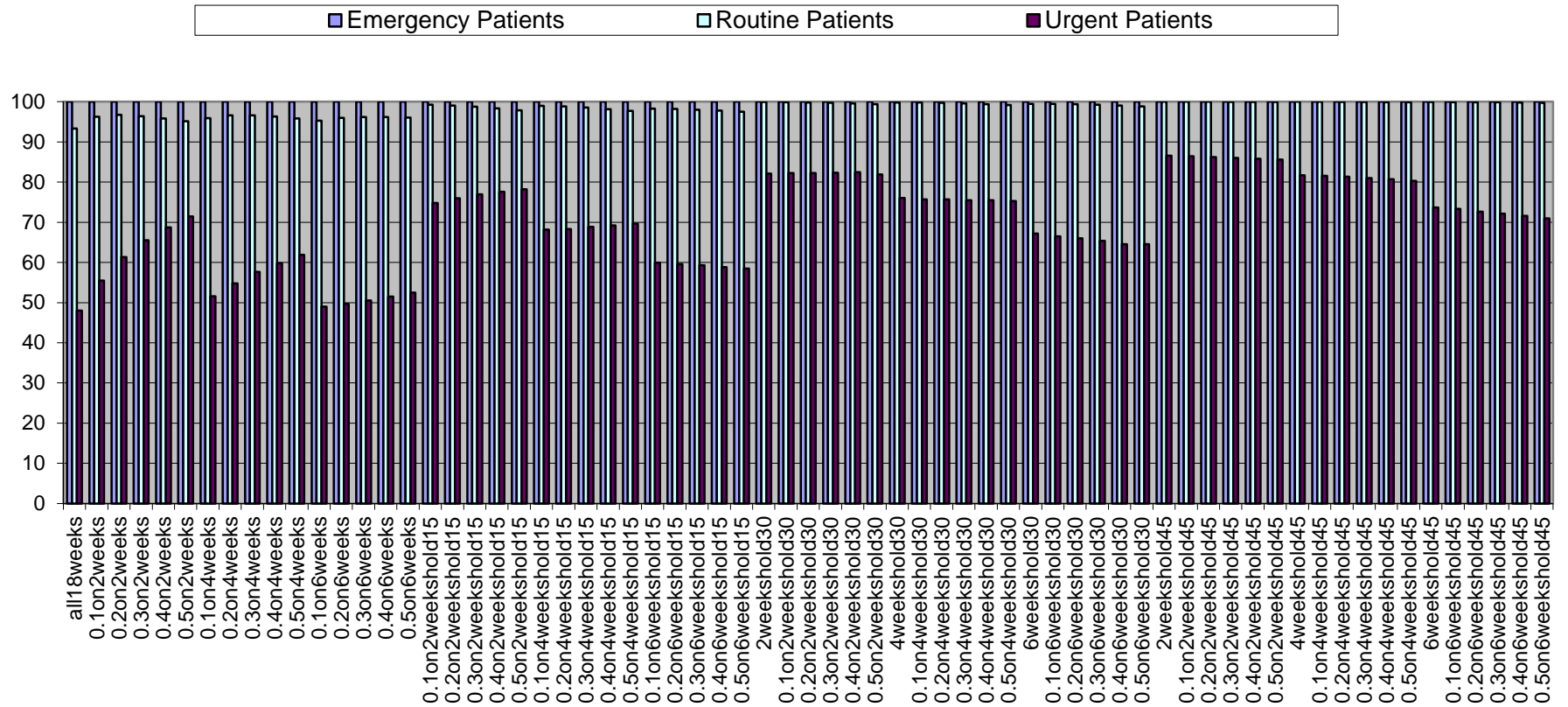
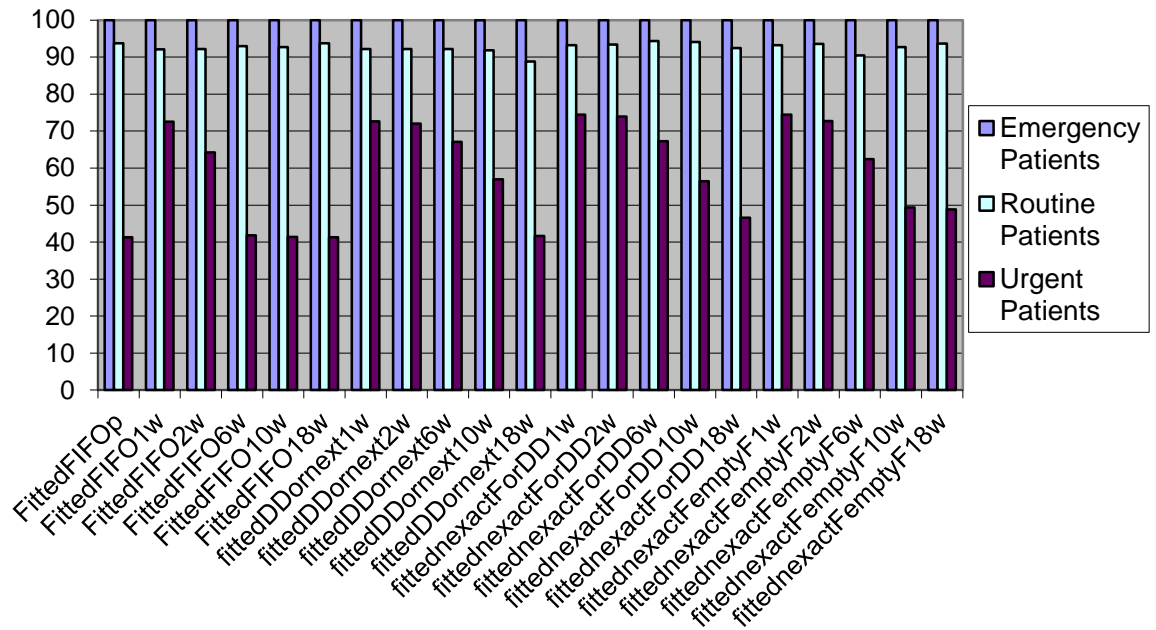**Figure 54: Comparison of the percentage of each type of patient seen within their due dates for the algorithms with some patients only booked on short notice and some theatre time held until close to its dates.**

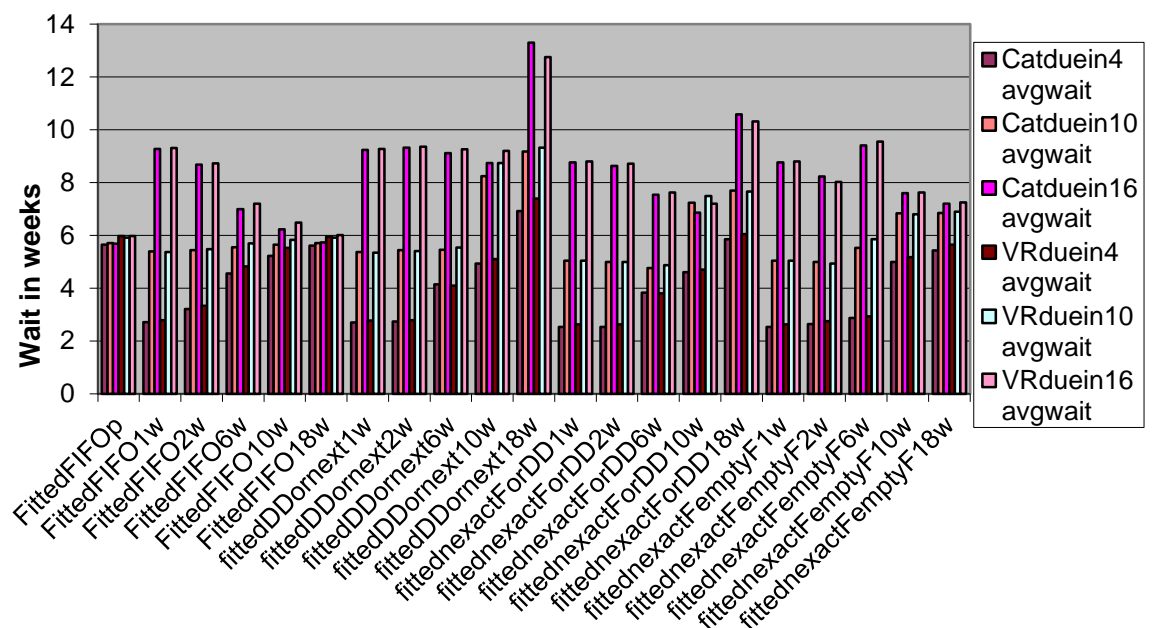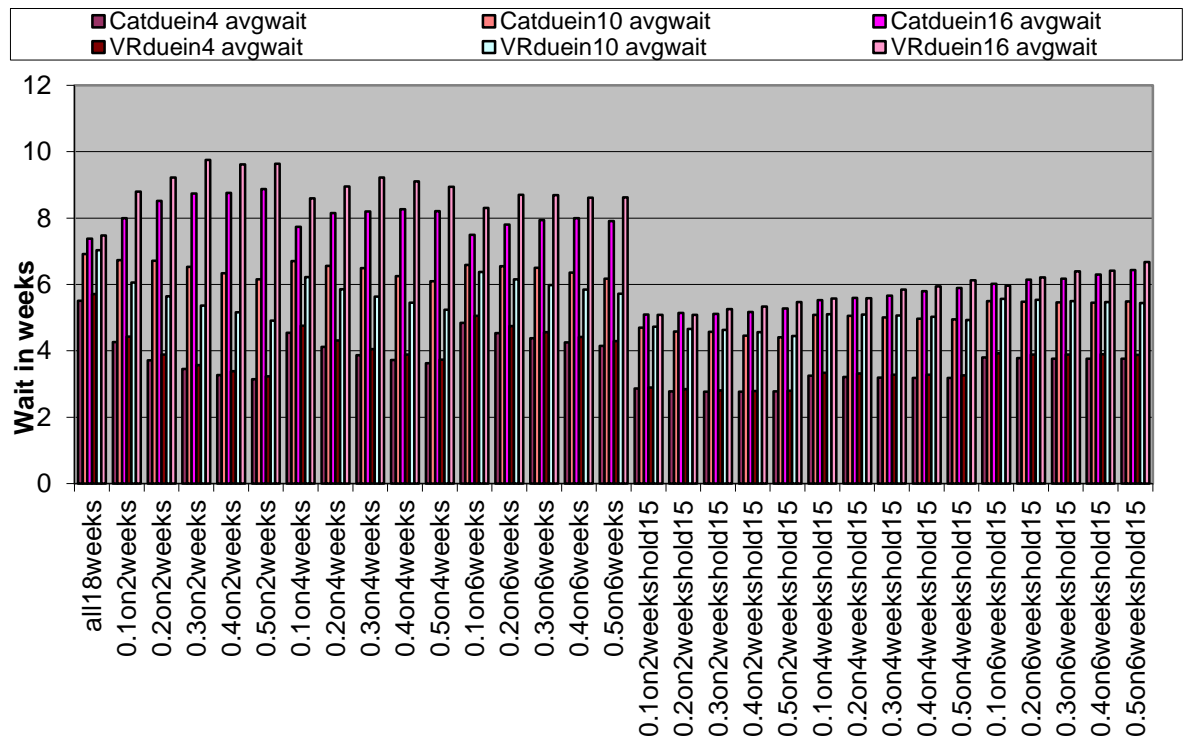**Figure 55: The waiting times for patients grouped by their initial time to due date for algorithms for the algorithms with some patients only booked on short notice and some theatre time held until close to its dates.**

The fairness testing data has been split into two graphs in order to allow the large amount of data to fit into the space available. Figure 55 and Figure 56 show that most urgent patients are waiting the shortest times under all of these algorithms, but the average waiting time for those who had 10 or 16 weeks left to due date are closer together than for other sets of algorithms.
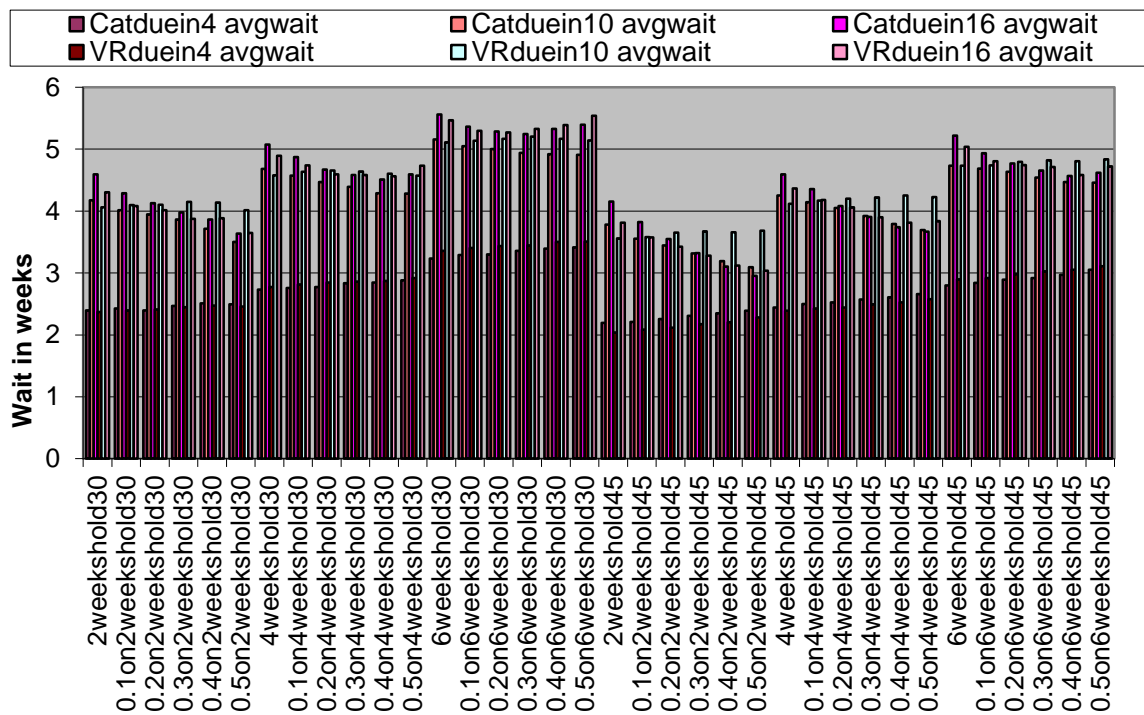
Figure 56: The waiting times for patients grouped by their initial time to due date for algorithms for the algorithms with some patients only booked on short notice and some theatre time held until close to its dates.

Given that most of the data shown in the table for all of the other sets of algorithms does not vary much between the algorithms, such a table is not included for these algorithms. However, as there is variation in the number of overtakes, a graph for these is included below.
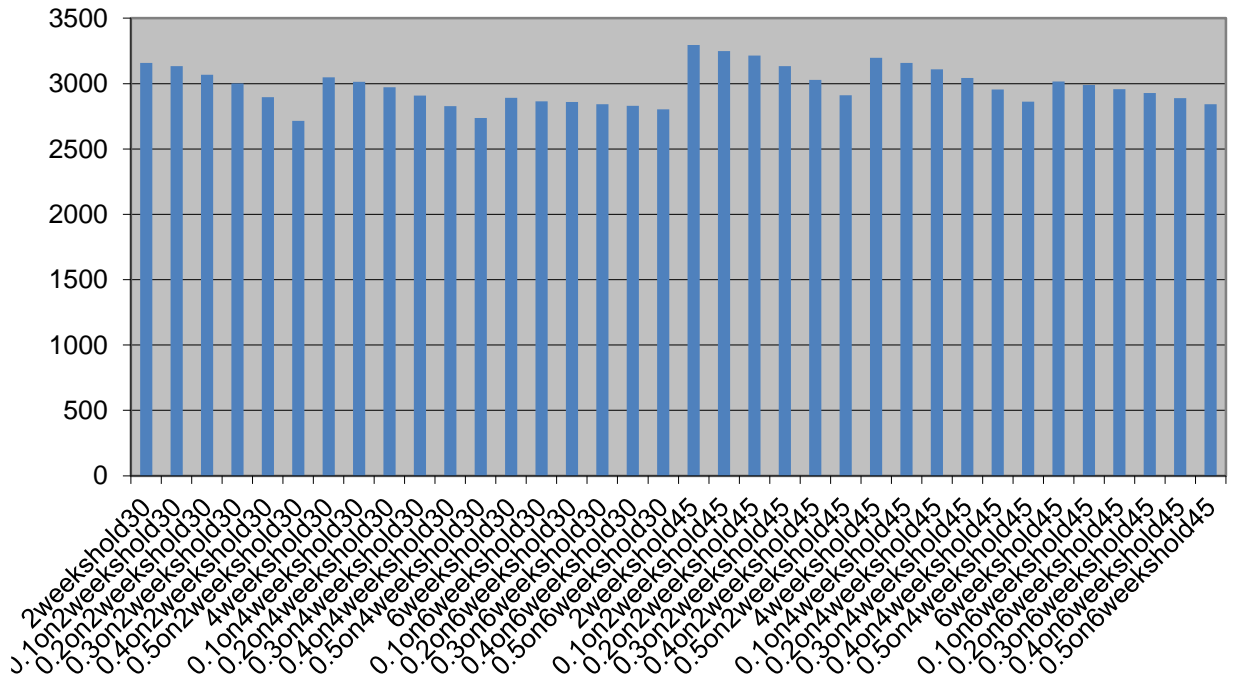
Figure 57: Graph showing the number of overtakes occurring for the algorithms with some patients only booked on short notice and some theatre time held until close to its dates.

Figure 57 shows that there is some variation in the number of overtakes but not a great deal for these variations in the algorithms.

6.9 Conclusions

This chapter illustrates how the problem of booking individual patients can be explored using simulation, with a case study used to explore the effects of different algorithms.

Our method of searching through potential algorithms has been designed to be as comprehensive as possible. The initial searches through the literature on appointment scheduling and scheduling in general in Section 5.2, as well as discussions with hospital staff, generated a variety of variations for various aspects of potential algorithms. Our process of testing algorithms has worked through the different aspects of the algorithms, taking forward the options that produced improvements in the results. When testing aspects that could take values on a continuous range, for example, the percentage of a theatre slot to hold back for late booking, we sampled at regular intervals across the range of possible values. This has provided a comprehensive exploration of all of the

variations considered. It does not preclude the possibility of their being other algorithms that would produce better results, but it limits this as much as reasonably possible.

The results given illustrate that the booking algorithm used can make a considerable difference to the waiting times of patients and the ability of the hospital to treat them within target dates. They also show that the best algorithm to use depends on the objectives of the team and the importance placed on giving patients plenty of notice of their operation dates.

When comparing the algorithms hospital managers should consider whether or not fairness each type of fairness is important compared to treating patients within the times required by their clinical need or waiting time targets.

If patients can be given only a week or two's notice, then delayed algorithms where patients are only booked a couple of weeks ahead of surgery provide the best results in terms of the percentage of patients treated before their due dates. Such delays are also fairest to the most urgent patients allowing them to be treated quickly and in terms of avoiding 'overtakes'. This agrees with the results from general scheduling literature when dealing with online problems, of which this problem is a variation.

If it is considered acceptable to book some patients (predominantly those who are most urgent and so have to be booked on short notice anyway), then restricting a portion of the available theatre time so that it only becomes available for booking a few weeks before it each slot takes place, also achieves high numbers of patients treated within due dates. This approach allows patients to be booked at the time of decision that they require treatment and have a larger proportion treated within their due dates. The disadvantage of this method of booking is that it is less fair in terms of the number of patients who are treated before those with earlier due dates and in terms of the waiting times of less urgent patients correlating with their urgency. Thus, if the 'fit' of operation durations to theatre times is tight, then restricting the theatre time available for booking will make a substantial difference in terms of treating patients by their due dates. However, if the amount of theatre time available is such that there is enough flexibility to treat patients by their due dates anyway, then this policy may create unnecessary unfairness in the system.

If either form of short notice is not considered permissible, then the amount of notice required influences the choice of algorithm. To avoid unused theatre time, if there is a space available the following week, then patients should be booked into it. Otherwise, they should be booked as late as possible to allow space for more urgent arrivals to arrive later. If the period of notice required is close to 10 weeks, then where there are slots that patients will fill exactly they should be used, although this will increase the number of overtakes.

In summary, the best algorithm depends on the priorities of the decision maker, particularly on how much the booking decision can be delayed or the ability to book into theatre time released only close to the time of the surgery slots. This research provides guidance on which algorithms will suit different sets of priorities.

Throughout this modelling close links have been maintained with the surgeon whose data it is based on and he has been making changes to his scheduling patterns based on the results.

7. Concluding Remarks

This section summarises the entire project demonstrating its impact on and contribution to hospital operating theatre scheduling; discusses how the field could develop in future; and includes concluding remarks based on the thesis as a whole.

7.1 Summary of contributions

The literature review reveals that a range of operational research techniques can be applied to different aspects of operating theatre scheduling and that the problems involved are sufficiently complex that is necessary to consider them separately. Based on this, we also separated the problem into separate stages. The literature review also identifies a lack of evidence of implementation of studies in this area of research.

In order to explore possible reasons for the lack of implementation and to ensure we started with a clear understanding of the challenges involved in theatre scheduling from the point of view of hospital staff, we started with a qualitative study. This uses cognitive mapping to illustrate the issues and their interconnectedness, as well as allowing us to test our understanding with theatre staff. Combined with the information from the literature, it allowed us to identify that challenges to implementation include flexibility to take account of the characteristics of a particular hospital and that staff are very busy with immediate concerns, so lack the time to investigate models that may help them in the longer term.

The literature review identifies three main levels of theatre scheduling problems, strategic, tactical and day-to-day bookings. We considered the strategic problem to be more political and hospital specific, and hence felt more impact could be achieved by focussing on the tactical and day-to-day challenges.

At the tactical level, we explore the generation of master theatre timetables and our results demonstrate the reduction in the maximum number of beds required that can be achieved by considering the impact of the theatre timetable on this aspect of the hospital. Unlike previous studies, our model allows surgeons preferences for theatre slots to be taken into account where possible. The model also allows for consideration of days when surgeons are unavailable, variations in types of theatre and the types of surgery

195

that can be conducted in them, limited equipment availability and varying the length of the cycle over which the timetable is repeated. Previous studies have not incorporated this full range of factors.

The tactical model is set up with an Excel user interface, as hospital staff members are generally familiar to with Excel this makes it more accessible to them; it also makes it possible to adapt the model to a different hospital or to changing needs in a straightforward manner. The user interface allows users to adjust the weights given to the various objectives without needing to reload all of the data. This allows them to explore a variety of solutions so that they can take account of tacit information that could not be included in the model when selecting a new timetable. This model provides a significant improvement on previous technique used in local hospitals to develop new timetables. In particular, due to the relative speed at which the model suggests timetables, it can quickly and without bias demonstrate the extent of the differences attributed to a specific timetable.

The value of the qualitative modelling became apparent when we transferred to working with a surgeon in a different hospital to consider day-to-day scheduling, and our understanding of theatre scheduling transferred to the new scenario.

In the day-to-day scheduling part of this project, we focus on the advanced booking of individual patients for surgery. At this stage, we use simulation to explore a range of algorithms for booking patients, with the algorithms derived from a mixture of scheduling literature and ideas from hospital staff. This reveals that, as in online scheduling, more efficient schedules can be achieved by delaying scheduling at least part of the scheduling as close to the time of surgery as possible. In surgery, the extent to which the decision can be delayed is limited by the need to give patient adequate warning to make arrangements to attend hospital for their surgery.

The data collection for this part of the project revealed the importance of considering all aspects of the time taken for operations when booking patients, for example allowing for the time required to 'turnaround' the operating theatre between patients. The testing of algorithms also demonstrated the intuitively obvious point that it is important to use theatre slots as fully as possible.

The testing of algorithms also reveals that the most efficient schedules are obtained when patients are considered for scheduling in order of increasing due date, prioritising in non-decreasing order of expected operation duration when due dates are equal.

All of the work in this thesis has been undertaken working closely with partner hospitals to ensure that the problems addressed are real challenges faced in hospitals and that sufficient factors are considered to make the results useful. At the tactical level, we have incorporated a wider range of factors and greater flexibility than other studies on this aspect of theatre scheduling. At the day-to-day scheduling level we have conducted a wider exploration of potential scheduling rules than other studies, drawing on the wealth of experience in the broader field of machine scheduling and suggestions from surgeons for inspiration.

The different stages of this project presented different challenges and constraints, therefore requiring different methodologies. As a whole, this thesis demonstrates that a range of methodologies can be applied to different stages of a problem to develop effective solutions.

7.2 Future research on hospital operating theatre scheduling

This section makes suggestions for further research applying operational research methods to the challenges surrounding hospital operating theatre scheduling.

As in previous studies, we have treated the tactical and day-to-day scheduling problems in isolation. In reality, the schedules produced will interact with each other, for example the day-to-day scheduling rules used could change the bed usage for a theatre slot of a particular surgeon, which would change the results of the tactical level schedule. It would be particularly interesting to conduct a long term study with a hospital implementing both sets of results. Then, using data since the change in day-to-day scheduling rerun the tactical model and adjust the theatre timetable, and so on. This approach has the potential to iteratively improving the entire timetabling system.

Given that there are seasonal variations in the demand for certain types of surgery and other pressures on hospitals, such as the winter flu season, it would be interesting to

explore the benefits of varying the theatre timetable to better fit seasonal variation in demand. To some extent this could be achieved by running different timetables at different times of year, but could a truly dynamic model taking account of factors like seasonal variation and changes in the numbers of referrals to hospitals do better? This would be a significant change to the system to which hospital staff members are accustomed, so the challenges involved may be too substantial a barrier to implementation.

At the day-to-day scheduling level, it would be interesting to see what improvements could be gained by using more complex algorithms that take account of the current 'busyness' of the system. With this approach, less urgent patients are booked earlier when the variation in arrivals has resulted in fewer cases in the system, this increases the fairness. However, the complexity of the system should be balanced against the need for transparency in the system and the easy with which current hospital systems can implement it.

Exploration of other specialities with different case mixes and including consideration of the impact of the number of beds available on day-to-day theatre scheduling would also be worthwhile.

It would also be valuable to explore the effects on the scheduling of urgent patients of the routine patients becoming more urgent as they reach come close to the 18 week waiting limit as discussed in Section 5.1.

The most significant challenge for the application of operational research to theatre scheduling remains getting the results implemented as widely as possible. It is our hope that the inclusion of qualitative modelling demonstrated in this theses, provides an example that will assist with addressing this challenge. Following the production of this thesis we will continue to work with local hospitals to implement its recommendations.

Appendix A

The table below expands on the meaning of the concepts given in Figure 1, where they are not apparent from the text given on the diagram.

Concept as on Figure 1	Explanation
1. Meet Targets	Meeting the combined targets including waiting times and cancellation targets.
2. Reduce waiting lists/times	Reducing the time patients wait between the decision that they requires surgery and their operations taking place.
3. Increase Productivity	Increasing the number of operations taking place in a each theatre slot.
4. Reduce Cancellations	Reducing the number of patients whose operations are cancelled.
5. Advertise via practice managers	Sending information to GP's via practice managers.
6. Introduce regular GP newsletter/ booklet of surgeon info/website	
7. Raise GP awareness of when to refer (consider availability) and fitness for surgery	
8. More well being activities in the community	Increase the activities like falls clinics aimed at reducing the requirement for surgery.
9. More filtered by physio	Patients receiving physiotherapy which reduces the number requiring surgery.
10. Increase face to face contacts with GPs e.g. quarterly meetings cycling through specs	Increasing contact between surgeons and GPs to improve communication.
11. Provide information on procedures for GPs to inform patients	
12. Reduce Outpatient numbers	
13. Increasing proportion of out patients need surgery	
14. Raise GP awareness of capacity and surgeons specialisations so refer	

appropriately	
15. Reduce demand for surgery	
16. Allow patients to opt in to receive info. by email	
17. Introduce contacting patients 48 hours before op	Contacting patients before their operations to check that they are fit and able to attend.
18. Cancel op. if DNA pre-assessment	If patients do-not-attend their pre-assessment then cancel their operations.
19. Include more in pre-assessments	
20. Reduce number of appt inconvenient	Reduce the number of cancellations of surgery because the appointment was inconvenient to the patient.
21. Reduce no of DNA	Reduce the number of do-not-attends
22. Reduce no. of Pre-op guidance not followed	
23. More opportunity for patients to understand what is involved / consent at pre-assessment	
24. Reduce number of self heal/die on list	
25. Reduce patient cancellations	
26. Reduce no. of operation not required	Reduce the number of cancellations due to a clinical/patient decision that the operation is no longer required
27. Reduce no. of unfit for surgery	
28. Reduce clinical cancellations	
29. Whole hospital work later on specific days	Either by having extended theatre slots or additional evening theatre slots with the backup of other hospital systems.
30. Increased staff flexibility	
31. Book in front of target i.e. shorter waits	Booking patients to be seen well within their waiting time targets
32. Improve specialty/quality mix of theatre staff / re-design roles	
33. Run 3 session days	
34. More all day lists	

35. Plan case mix around TSSU turn around so equipment not such limiting factor	Improvements to the service for sterilising theatre equipment.
36. Reduce turnaround times for equipment	Reduce the times between one operation ending and the next beginning.
37. Book at decision to operate	Booking patients theatre day at the time when the decision to operation on them is made.
38. Where appropriate have two Reg/staff grade running surgeries with consultants moving between them.	
39. Increase capacity	
40. Equipment availability less of a limiting factor	
41. Proper utilisation of additional kit	
42. Same Anaes + Surg operating list to reduce delays	Having the same team working together.
43. Increase proportion of sessions that start on time	
44. Reduce over booking/overruns	
45. Increase time available for infection control	
46. Reduce occurrence of missing notes/ admin error	
47. Computerise notes	
48. Use capacity when surgeons on leave etc.	Ensuring other surgeons use the theatre time if the surgeon its allocated is on leave or on a training course etc.
49. Have notes ready	Having information on patients ready at the start of the surgical slot.
50. Get ordered list to wards sooner	Ensuring wards know in advance the order in which to send patients for surgery.
51. Ability to book patients further in advance	
52. Have coordinator to oversee bookings/theatres	
53. Constraints on system e.g. theatre type required	

54. More beds available	
55. Bring more patients in early to make sure they get beds	Bringing in patients the night before their surgery is booked, purely to ensure that they have a bed after surgery (to avoid cancellation due to lack of beds).
56. Increase efficiency of bed usage – needs further consideration to determine how to do this effectively	
57. Increase matching of bookings to available time	
58. Have necessary facilities open, before surgery due to start	
59. Increase awareness of list availability	
60. Lists are available on central drive	
61. Use a diary system, booking only if bed, theatre time, equipment all available	
62. Ability to plan to smooth demand for beds	
63. Reduce use of beds by medical patients	Use of surgical beds by patients from not requiring surgery due to lack of beds elsewhere in the hospital.
64. Improve predictability of medical case load	
65. Development of bed model for medicine	
66. Dedicate lists LA or GA to make best use of anaesthetists	Having surgical lists where all patients are either requiring local anaesthetic or general anaesthetic.
67. Procedures should be coordinated in advance	
68. ?Designate sessions of all major/minor cases?	
69. Ability to include bed usage in planning	
70. Optimise Theatre Timetable	
71. Improved ability predict to LOS	Improve ability to predict length of stay in

	hospital.
72. Ability to predict bed usage	
73. Increase understanding of variability of demand	
74. Understand seasonal variations – is this significant?	

Appendix B

The information store of the simulation model holds the following information, all of the references relate to the list given in 5.4.2.1. The list is given in alphabetical order (with those starting with capitals followed by those starting with lowercase letters).

- Avg Emergency wait – used for returning the average emergency wait as calculated in End Run Logic (see 17a ii)
- Avg VR Wait – used for returning the average wait for VR patients as calculated in End Run Logic (see 17a ii above)
- AvgUtilizationThur – used for returning the average utilization on Thursdays as calculated in End Run Logic (see 17b vi)
- AvgUtilizationTue – used for returning the average utilization on Tuesdays as calculated in End Run Logic (see 17b vi)
- Catduein10avgwait – used for returning the average waiting time for cataract patients whose initial time to target is 10 weeks as calculated in End Run Logic (see 17b vii)
- Catduein10maxwait – used for returning the maximum waiting time for cataract patients whose initial time to target is 10 weeks as calculated in End Run Logic (see 17b vii)
- Catduein10minwait – used for returning the minimum waiting time for cataract patients whose initial time to target is 10 weeks as calculated in End Run Logic (see 17b vii)
- Catduein10sumsqu – used for returning the sum of squared waiting times for cataract patients whose initial time to target is 10 weeks as calculated in End Run Logic (see 17b vii), this is used to calculate the standard deviation
- Catduein16avgwait – used for returning the average waiting time for cataract patients whose initial time to target is 16 weeks as calculated in End Run Logic (see 17b vii)
- Catduein16maxwait – used for returning the maximum waiting time for cataract patients whose initial time to target is 16 weeks as calculated in End Run Logic (see 17b vii)
- Catduein16minwait – used for returning the minimum waiting time for cataract patients whose initial time to target is 16 weeks as calculated in End Run Logic (see 17b vii)

- Catduein16sumsqu – used for returning the sum of squared waiting times for cataract patients whose initial time to target is 16 weeks as calculated in End Run Logic (see 17b vii), this is used to calculate the standard deviation

- Catduein4avgwait – used for returning the average waiting time for cataract patients whose initial time to target is 4 weeks as calculated in End Run Logic (see 17b vii)

- Catduein4maxwait – used for returning the maximum waiting time for cataract patients whose initial time to target is 4 weeks as calculated in End Run Logic (see 17b vii)

- Catduein4minwait – used for returning the minimum waiting time for cataract patients whose initial time to target is 4 weeks as calculated in End Run Logic (see 17b vii)

- Catduein4sumsqu – used for returning the sum of squared waiting times for cataract patients whose initial time to target is 4 weeks as calculated in End Run Logic (see 17b vii), this is used to calculate the standard deviation

- Count Cat in Target – used to count the number of cataract patients treated within their due date (see 8b and 17a i)

- Count CatTreated – used to keep track of the number of cataract patients treated to enable calculations (for example in 17a ii).

- Count Emergency Treated – used to keep track of the number of emergency patients treated to enable calculations (for example in 17a ii).

- Count Emergency in Target – used to count the number of emergency patients treated within their due date (see 8d and 17a i)

- Count VR Treated – used to keep track of the number of VR patients treated to enable calculations (for example in 17a ii).

- Count VR in Target – used to count the number of VR patients treated within their due date (see 8c and 17a i)

- Count bookings – used to keep track of the number of patients booked.

- Count routine treated – used to keep track of the number of routine patients treated to enable calculations (for example in 17a ii).

- Count urgent treated – used to keep track of the number of urgent patients treated to enable calculations (for example in 17a ii).

- CountDays – used in the booking algorithms to look through the slots available on Tuesdays
- CountDays2 – used in the booking algorithms to look through the slots available on Thursdays
- CountThurExactFits – used in the booking algorithms to keep track of the number of patients who have been assigned to slots which they fill exactly on Thursdays
- CountThurtoolong – used to keep track of the number of the number of slots that over run by more than acceptable overtime on Thursdays (see 17b v)
- CountTueExactFits – used in the booking algorithms to keep track of the number of patients who have been assigned to slots which they fill exactly on Tuesdays
- CountTuetoolong – used to keep track of the number of the number of slots that over run by more than acceptable overtime on Tuesdays (see 17b v)
- Daysfullybooked – used in some booking algorithms to keep track of the number of the last consecutive day from the current simulation time this is fully booked (to allow such days to be ignored when making future bookings).
- ExpTurn – the expected turnaround time between patients (used in the booking algorithms to allow time for turnarounds between patients).
- MaxUtilizationThur – used to return the value of the maximum utilization for a Thursday (see 17b iii)
- MaxUtilzationTue – used to return the value of the maximum utilization for a Tuesday (see 17b iii)
- MaxVRwait – used to return the value of the maximum waiting time for a VR patient (see 8c)
- Maxcatwait– used to return the value of the maximum waiting time for a cataract patient (see 8b)
- Maxemergwait – used to return the value of the maximum waiting time for an emergency patient (see 8a)
- Maxroutinewait – used to return the value of the maximum waiting time for a routine patient (see 8d)
- Maxurgentwait– used to return the value of the maximum waiting time for an urgent patient (see 8d)

- MinUtilizationThur – used to return the value of the minimum utilization for a Thursday (see 17b iv)
- MinUtilizationTue – used to return the value of the minimum utilization for a Tuesday (see 17b iv)
- Most Empty2 – used in some booking algorithms to keep track of which day has the least bookings in it, out of a set checked.
- Overtakes – used to keep track of the number of overtakes (as described in 17b ix above).
- Per cent Cat in Target – used to return the percentage of cataract patients treated before their due dates (see 17a i)
- Per cent Emergency in Target – used to return the percentage of emergency patients treated before their due dates (see 17a i)
- Per cent VR in Target – used to return the percentage of VR patients treated before their due dates (see 17a i)
- Per cent routine in target – used to return the percentage of routine patients treated before their due dates (see 17a i)
- Per cent urgent in target – used to return the percentage of urgent patients treated before their due dates (see 17a i)
- Results Collection Period – the number of weeks for which the simulation runs, set at 300 (see Section 6.1.3).
- Simulation Time – the clock in Simul8, gives the current day
- Sum Cat Wait Time – used to calculate the average waiting time for cataract patients (see 16d and 17a ii)
- Sum Emergency wait time – used to calculate the average waiting time for emergency patients (see 16d and 17a ii)
- Sum VR Wait Time – used to calculate the average waiting time for VR patients (see 16d and 17a ii)
- Sum routine wait time – used to calculate the average waiting time for routine patients (see 16d and 17a ii)
- Sum urgent wait time – used to calculate the average waiting time for urgent patients (see 16d and 17a ii)
- Test – used in some algorithms
- Testorig – used in some algorithms

- Thur slot length – the maximum booking limit for Thursdays usually set at 0.21, where 0.01 is 10 minutes. This can be adjusted to allow more or less time for emergencies

- Thurexactfit – used in some of the booking algorithms which look for slots which the patients expected theatre time will exactly fill to the booking limit

- TotalUtilzationThur – used to calculate the average utilization on Thursdays (see 17b vi)

- TotalUtilzationThurSumsqu – used to generate the values required to calculate the standard deviation of utilization on Thursdays (see 17b vii)

- TotalUtilzationThuravSumsqu – used to generate the values required to calculate the standard deviation of utilization on Thursdays (see 17b vii)

- TotalUtilzationTue – used to calculate the average utilization on Tuesdays (see 17b vi)

- TotalUtilzationTueSumsqu – used to generate the values required to calculate the standard deviation of utilization on Tuesdays (see 17b vii)

- TotalUtilzationTueavSumsqu – used to generate the values required to calculate the standard deviation of utilization on Tuesdays (see 17b vii)

- Tue slot length – the maximum booking limit for Tuesdays usually set at 0.215, where 0.01 is 10 minutes. This can be adjusted to allow more or less time for emergencies

- Tueexactfit – used in some of the booking algorithms which look for slots which the patients expected theatre time will exactly fill to the booking limit

- Unique No Counter – used to count the unique numbers that have been assigned so that none is assigned twice (see 1a, 2)

- VRduein10avgwait - used to return the average waiting time for VR patients due to be treated 10 weeks from their arrival, for considering the fairness of the algorithms (see 17b viii)

- VRduein10maxwait - used to return the maximum waiting time for VR patients due to be treated 10 weeks from their arrival, for considering the fairness of the algorithms (see 17b viii)

- VRduein10minwait- used to return the minimum waiting time for patients VR due to be treated 10 weeks from their arrival, for considering the fairness of the algorithms (see 17b viii)

- VRduein10sumsqu - used to return the values needed to calculate the standard deviation of the waiting time for VR patients due to be treated 10 weeks from their arrival, for considering the fairness of the algorithms (see 17b viii)

- VRduein16avgwait - used to return the average waiting time for VR patients due to be treated 16 weeks from their arrival, for considering the fairness of the algorithms (see 17b viii)

- VRduein16maxwait - used to return the maximum waiting time for VR patients due to be treated 16 weeks from their arrival, for considering the fairness of the algorithms (see 17b viii)

- VRduein16minwait - used to return the minimum waiting time for VR patients due to be treated 16 weeks from their arrival, for considering the fairness of the algorithms (see 17b viii)

- VRduein16sumsqu - used to return the values needed to calculate the standard deviation of the waiting time for patients due to be treated 16 weeks from their arrival, for considering the fairness of the algorithms (see 17b viii)

- VRduein4avgwait - used to return the average waiting time for VR patients due to be treated 4 weeks from their arrival, for considering the fairness of the algorithms (see 17b viii)

- VRduein4maxwait - used to return the maximum waiting time for VR patients due to be treated 4 weeks from their arrival, for considering the fairness of the algorithms (see 17b viii)

- VRduein4minwait - used to return the minimum waiting time for VR patients due to be treated 4 weeks from their arrival, for considering the fairness of the algorithms (see 17b viii)

- VRduein4sumsqu - used to return the values needed to calculate the standard deviation of the waiting time for VR patients due to be treated 4 weeks from their arrival, for considering the fairness of the algorithms (see 17b viii)

- Valuemostempty – used in some algorithms to record the most empty slot considered so far when searching through slots on Tuesdays, initially set to contain a high value which will be replaced

- Valuemostempty2 – used in some algorithms to record the most empty slot considered so far when searching through slots on Thursdays, initially set to contain a high value which will be replaced

- Warm Up Period – contains the length of time during which results will not be collected, to ensure the results reflect the effects of the algorithm when running with patients already booked rather than when starting with an empty system (see Section 6.1.3).

- avgusedThur – used to calculate the average length of a Thursday session (see 17b vi)

- avgusedTue – used to calculate the average length of a Tuesday session (see 17b vi)

- booking limit Thur – stores the limit on the amount of time that can be booked on Thursdays, set to 3 and a half hours unless specified otherwise in the algorithms used

- booking limit Tue – stores the limit on the amount of time that can be booked on Thursdays, set to just over 3 and a half hours unless specified otherwise in the algorithms used

- booking limitThuremerg – allows a limit on the amount of time that can be booked including emergency patients on Thursdays

- booking limitTueemerg– allows a limit on the amount of time that can be booked including emergency patients on Tuesdays

- count routine in target – used to count the number of routine patients treated within their due date (see 8d and 17a i)

- count urgent in target – used to count the number of urgent patients treated within their due date (see 8d and 17a i)

- countThur – used to keep track of the number of Thursdays on which operations have occurred, used in calculations like the average utilization and slot length on Thursdays (see 17b vi)

- countTue – used to keep track of the number of Thursdays on which operations have occurred, used in calculations like the average utilization and slot length on Thursdays (see 17b vi)

- countdayscalc – used in some of the algorithms to store data relating to the number of days considered

- i – used as a dummy to keep track of slots considered in some algorithms

- maxusedThur – used to keep track of and return the maximum length of a slot on a Thursday (see 17b i)

- maxusedTue – used to keep track of and return the maximum length of a slot on a Tuesday (see 17b i)

- minusedThur – used to keep track of and return the minimum length of a slot on a Thursday (see 17b ii)

- minusedTue – used to keep track of and return the minimum length of a slot on a Tuesday (see 17b ii)

- totalusedThur – used to calculate the average length of a theatre slot used on a Thursday (see 17b vii)

- totalusedThurSumsqu– used to calculate the standard deviation of the length of a theatre slot used on a Thursday (see 17b vii)

- totalusedThuravSumsqu– used to calculate the standard deviation of the length of a theatre slot used on a Thursday (see 17b vii)

- totalusedTue– used to calculate the average length of a theatre slot used on a Tuesday (see 17b vii)

- totalusedTueSumsqu– used to calculate the standard deviation of the length of a theatre slot used on a Tuesday (see 17b vii)

- totalusedTueavSumsqu– used to calculate the standard deviation of the length of a theatre slot used on a Tuesday (see 17b vii)

Appendix C

This appendix contains additional results from the tests discussed in Section 6.8.

Table 16: Other results for the best algorithms, when applied to the variation of the fitted data.

	Algorithms									
	FittedFIFOp	FittedFIFO 2w	FittedFIFO 6w	FittedFIFO 10w	FittedFIFO 18w	fittedDDor next1w	fittedDDor next2w	fittedDDor next6w	fittedDDor next10w	fittedDDor next18w
Mean slot length Tuesday %	90.2	90.7	90.2	90.2	90.2	90.2	90.7	90.7	90.3	89.8
Mean slot length Thursday %	97.9	97.2	98.0	98.0	98.0	97.9	97.2	97.2	97.7	98.5
% of Tuesdays overrunning	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
% of Thursdays overrunning	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
Mean Utilization on Tues %	74.5	74.9	74.5	74.5	74.5	74.5	74.9	75.0	74.6	74.2
Mean Utilization on Thurs %	80.4	79.8	80.5	80.5	80.5	80.4	79.8	79.6	80.2	80.9
Number of Overtakes	1565	704	816	1213	1422	1553	702	680	1865	3589

Table 17: Other results for the best algorithms, when applied to the variation of the fitted data, continued.

	Algorithms							
	fittednexactF orDD2w	fittednexactF orDD6w	fittednexactF orDD10w	fittednexactF orDD18w	fittednexactFe mptyF2w	fittednexactFe mptyF6w	fittednexactFe mptyF10w	fittednexactFe mptyF18w
Mean slot length Tuesday %	89.8	90.6	90.6	90.2	89.5	89.3	90.6	90.6
Mean slot length Thursday %	98.6	97.5	97.6	98.3	99.5	99.8	97.5	97.4
% of Tuesdays overrunning	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
% of Thursdays overrunning	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
Mean Utilization on Tues %	74.7	74.9	74.9	74.5	73.9	74.1	74.9	74.9
Mean Utilization on Thurs %	79.7	80.0	79.9	80.7	81.7	81.2	80.0	79.8
Number of Overtakes	3176	697	731	2119	3698	3654	697	887

References

Abouleish, A.E., Dexter, F., Epstein, R.H., Lubarsky, D.A., Whitten, C.W. and Prough, D.S., (2003). Labor costs incurred by anaesthesiology groups because of operating rooms not being allocated and cases not being scheduled to maximize operating room efficiency. *Anesthesia and Analgesia 96(4)*, 1109-1113

Ackermann, F., Eden, C. and Williams, T., (1997). Modelling for litigation: mixing qualitative and quantitative approaches. *Interfaces 27(2)*, 48–65

Adan, I. and Vissers, J., (2002). Patient mix optimisation in hospital admission planning: a case study. *International Journal of Operations and Production Management 22(4)*, 445-461

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

Ajtai, M., Aspnes, J., Rabani, Y., Schulman, L.J. and Waarts, O., (1998). Fairness in scheduling. *Journal of Algorithms 29(2)*, 306-357

Anderson, D.R., Sweeney, D.J. and Williams, A.T., (2003). *An Introduction to Management Science: Quantitative Approaches to Decision Making*. Cincinnati, Ohio: South-Western College Publishers

Anderson, E.J. and Potts, C.N., (2004). Online scheduling of a single machine to minimize total weighted completion time. *Mathematics of Operational Research 29(3)*, 686-697

Angle, E. and Bampis, E., (2005). A multi-start dynasearch algorithm for the time dependent single-machine total weighted tardiness scheduling problem. *European Journal of Operational Research 162(1)*, 281-289

Arenas, M., Bilbao, A., Caballero, R., Gomez, T., Rodriguez, M.V. and Ruiz, F., (2002). Analysis via goal programming of the minimum achievable stay in surgical waiting lists. *Journal of the Operational Research Society 53(4)*, 387-396

Atienza N., García-Heras J., Munoz-Pichardo, J.M. and Villa R., (2008). An application of mixture distributions in modelization of length of hospital stay. *Statistics in Medicine 27(9)*, 1403-1420

Augusto, V., Xie, X. and Perdomo, V., (2010). Operating theatre scheduling with patient recovery in both operating rooms and recovery beds. *Computers and Industrial Engineering 58(2)*, 231-238

Avi-Itzhak, B., Brosh, E. and Levy, H., (2007). SQF: A slowdown queuing fairness measure. *Performance Evaluation 64(9-12)*, 1121-1136

Ballard, S.M. and Kuhl, M.E., (2006). *The Use of Simulation to Determine Maximum Capacity in the Surgical Suite Operating Room*. In. Perrone, L. F., Wieland, F. P., Liu, J., Lawson, B. G., Nicol, D. M., and Fujimoto, R. M., Piscataway (eds.), (2006). *Proceedings of the 2006 Winter Simulation Conference*. New Jersey: Institute of Electrical and Electronics Engineers, 433-438.

Banks, J., Carson II, J.S., Nelson, B.L., (1999). *Discrete- Event System Simulation (Second Edition)*. New Jersey: Prentice Hall

Barnoon, S. and Wolfe, H., (1968). Scheduling a multiple operating room system: a simulation approach. *Health Services Research 3(4)*, 272-285

Basson, M.D. and Butler, T., (2006). Evaluation of operating room suite efficiency in the Veterans Health Administration system by using data-envelopment analysis. *The American Journal of Surgery 192(5)*, 649-656

Basson, M.D., Butler, T. and Verma, H., (2006). Predicting patient nonappearence for surgery as a scheduling strategy to optimise operating room utilization in Veterans' Administration Hospital. *Anesthesiology 104(4)*, 826-834

Batun, S., Denton, B.T., Huschka, T.R. and Schaefer, A.J., (2010). The benefit of pooling surgery processing under uncertainty operating rooms and parallel. *INFORMS Journal on Computing,* DOI: doi:10.1287/ijoc.1100.0396

Beasley, J.E., (1996). *Advances in Linear and Integer Programming.* Oxford: Oxford University Press

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

Belien, J. and Demeulemeester, E., (2007b). On the trade-off between staff-decomposed and activity-decomposed column generation for the staff scheduling problem. *Annals of Operations Research 155(1)*, 143-166

Belien, J. and Demeulemeester, E., (2008). A branch-and-price approach for integrating nurse and surgery scheduling. *European Journal of Operational Research 189(3)*, 652-668

Belien, J., Demeulemeester, E. and Cardoen, B., (2006). 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

Belien, J., Demeulemeester, E. and Cardoen, B., (2009). A decision support system for cyclic master surgery scheduling with multiple objectives. *Journal of Scheduling 12(2)*, 147-161

Bellamy, M., (2002). Anaesthesia induction rooms - sheer luxury. *Anaesthesia 57(2)*, 187-188

Bekker, R. and Koeleman P.M., (2011). Scheduling admissions and reducing variability in bed demand. *Health Care Management Science 14(3)*, 237-249

Bhattacharayya, T., Vrahas, M., Morrison, S.M., Kim, E., Wiklund, R.A., Smith, R.M. and Rubash, H.E., (2006). The value of the dedicated orthopaedic trauma operating room. *The Journal of Trauma-Injury, Infection and Critical Care 60(6)*, 1336-1341

Blake J.T. and Donald J., (2002). Mount Sinai Hospital uses integer programming to allocate operating room time. *Interfaces 32(2)*, 63-73

Blake, J.T. and Carter, M.W., (2002). A goal programming approach to strategic resource allocation in acute care hospitals. *European Journal of Operational Research 140(3)*, 541-561

Blake, J.T., Dexter, F. and Donald, J., (2002). Operating room manager's use of integer programming for assigning block time to surgical groups: a case study. *Anesthesia and Analgesia 94(1)*, 143-148

Bowers, J. and Mould, G., (2002). Concentration and variability of orthopaedic demand. *Journal of the Operational Research Society 53(2)*, 203-210

Bowers, J. and Mould, G., (2004). Managing uncertainty in orthopaedic trauma theatres. *European Journal of Operational Research 154(3)*, 599-608

Bowers, J. and Mould, G., (2005). Ambulatory care and orthopaedic capacity planning. *Health Care Management Science 8(1)*, 41-47

Bowers, J., (2010). Waiting list behaviour and the consequences for NHS targets. *Journal of the Operational Research Society 61(2)*, 246-254

Brahimi, M. and Worthington, D.J., (1991). Queueing models for out-patient appointment systems – a case study. *Journal of the Operational Research Society 42(9)*, 733-746

Brailsford, S.C., Lattimer, V.A., Tarnaras, P. and Turnbull, J.C., (2004). Emergency and on-demand health care: modelling a large complex system. *Journal of the Operational Research Society 55(1)*, 34-42

Broadway, J.W., Smith, M.G. and Archer, T.J., (2001). Anaesthesia without induction rooms. *Anaesthesia 56(1)*, 82-97

Bromhead, H.J. and Jones, N.A., (2002). The use of anaesthetic rooms for induction of anaesthesia: a postal survey of current practice and attitudes in Great Britain and Northern Ireland. *Anaesthesia 57(9)*, 850-854

Brown, J., Cooper, C. and Pidd, M., (2006). A taxing problem: the complementary use of hard and soft OR in the public sector. *European Journal of Operational Research 172(2)*, 666-679

Bu, T., and Ramjee, R., (2006). Generalized proportional fair scheduling in third generation wireless data networks. *Proceedings of IEEE Infocom'06,* Apr 2006

Cantlay, K.L., Baker, S., Parry, A. and Danjoux, G., (2006). The impact of a consultant anaesthetist led pre-operative assessment clinic on patients undergoing major vascular surgery. *Anaesthesia 61(3)*, 234-239

Cardoen, B., Demeulemeester, E. and Belien, J., (2009a). Sequencing surgical cases in a day-care environment: an exact branch-and-price approach. *Computers and Operations Research 36(9)*, 2660 -2669

Cardoen, B., Demeulemeester, E. and Belien, J., (2009b). Optimizing a multiple objective surgical case sequencing problem. *International Journal of Production Economics 119(2)*, 354-366

Cardoen, B., Demeulemeester, E. and Belien, J., (2010a). Operating room planning and scheduling: a literature review. *European Journal of Operational Research 201(3)*, 921-932

Cardoen, B., Demeulemeester, E. and Van de Hoeven, J., (2010b). On the use of planning models in the operating theatre: results of a survey in Flanders. *Journal of Health Planning and Management 25(4)*, 400-414

Castille, K., Gowland, B. and Wally, P., (2002). Booked inpatient admissions and hospital capacity. Variability must be managed to reduce waiting times and improve care. *British Medical Journal 324(7349)*, 1336

Chen, B., Potts, C.N., and Woeginger, G.J., (1998). *A Review of Machine Scheduling: Complexity, Algorithms and Approximability*. In Du, D.-Z. and Pardalos, P.M (eds), (1998). *Handbook of Optimization Volume 3*. Kluwer Academic Publishers: Dordrecht, Netherlands. 21-171

Cheng, T.C.E., Ng, C.T., Yuan, Z.H. and Liu, Z.H., (2005). Single machine scheduling to minimize total weighted tardiness. *European Journal of Operational Research 165(2)*, 423-443

Chinneck, J.W., (1997). Finding a useful subset of constraints for analysis in an infeasible linear program. *Informs 9(2)*, 164-174

Chinneck, J.W., (2001). Fast heuristics for the maximum feasible subsystem problem. *Informs Journal on Computing 13(3)*, 210-223

Cleary P.D., Greenfield S. and Mullay A.G., (1991). Variations in length of stay and outcomes for six medical and surgical conditions in Massachusetts and California. *Journal of the American Medical Association 266(1)*, 73-79

Combes, C., Meskens, N., Rivat, C. and Vandamme, J.P., (2008). Using KDD process to forecast the duration of surgery. *Int. J. Production Economics 112(1)*, 279-293

Coyle, G., (2000). Qualitative and quantitative modelling in system dynamics: some research questions. *System Dynamics Review 16(3)*, 225-244

Creemers, S. and Lambrecht, M., (2009). An advanced queueing model to analyze appointment-driven service systems. *Computers and Operations Research 36(10),* 2773-2785

Daellenbach, H.G. and McNickle, D.C., (2005). *Management Science, Decision Making Through Systems Thinking*. Palgrave Macmillan: Hampshire

Dekhici, L. and Belkadi, K., (2010). Operating theatre scheduling under constraints. *Journal of Applied Sciences 10(14)*, 1380-1388

Denton, B., Viapiano, J. and Vogl, A., (2007). Optimization of surgery sequencing and scheduling decisions under uncertainty. *Health Care Management Science 10(1)*, 13-24

Department of Health, (2012a). *A simple guide to Payment by Results*. Department of Health: Leeds. Accessed via: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/213150/PbR-Simple-Guide-FINAL.pdf

Department of Health, (2012b). 2013/14 Choice Framework. Department of Health: Online publication only at: http://www.nhs.uk/choiceintheNHS/Rightsandpledges/NHSConstitution/Documents/2013/choice-framework-2013-14.pdf

Desai, M.S., Penn, M.L., Brailsford, S. and Chipulu, M., (2008). Modelling of Hampshire Adult Services – gearing up for future demands. *Health Care Management Science 11(2)*, 167-176

Dexter, F. and Epstein, R.H., (2006). Holiday and weekend operating room on-call staffing requirements. *Anesthesia and Analgesia 103(6)*, 1494-1498

Dexter, F. and Ledolter, J., (2003). Managing risk and expected financial return from selective expansion of operating room capacity: mean-variance analysis of a hospital's portfolio, *Anesthesia and Analgesia 97(1)*, 190-195

Dexter, F. and Macario, A., (1999). Decrease in case duration required to complete an additional case during regular scheduled hours in an operating room suite: a computer simulation study. *Anesthesia and Analgesia 88(1)*, 72-76

Dexter, F. and Macario, A., (2002). 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 and Analgesia 94(5)*, 1272-1279

Dexter, F. and Macario, A., (2004). When to release allocated operating room time to increase operating room efficiency. *Anesthesia and Analgesia 98(3)*, 758-762

Dexter, F. and Traub, R.D., (2002). How to schedule elective surgical cases into specific operating rooms to maximise the efficiency of use of operating room time. *Anesthesia and Analgesia 94(4)*, 933-942

Dexter, F., (2000). A strategy to decide whether to move the last case of the day in an operating room to another empty operating room to decrease overtime labour costs. *Anesthesia and Analgesia 91(4)*, 925-928

Dexter, F., Abouleish, A.E., Epstein, R.H., Whitten, C.W. and Lubarsky, D.A., (2003). Use of operating room information system data to predict the impact of reducing turnover times on staffing costs. *Anesthesia and Analgesia 97(4)*, 1119-1126

Dexter, F., Blake, J.T., Penning, D.H., and Lubarsky, D.A., (2002c). 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 and Analgesia 94(1)*, 138-42

Dexter, F., Blake, J.T., Penning, D.H., Sloan, B., Chung, P. and Lubarsky, D.A., (2002a). Use of linear programming to estimate impact of changes in a hospital's operating room time allocation on perioperative variable costs. *Anesthesiology 96(3)*, 718-724

Dexter, F., Ledolter, J. and Wachtel, R., (2005). Tactical decision making for selective expansion of operating room resources incorporating financial criteria and uncertainty in subspecialties' future workloads. *Anesthesia and Analgesia 100(5)*, 1425-1432

Dexter, F., Lubarsky, M.D. and Blake, J.T., (2002b). Sampling error can significantly affect measured hospital financial performance of surgeons and resulting operating room time allocations. *Anesthesia and Analgesia 95(1)*, 184-188

Dexter, F., Macario, A. and Lubarsky, D.A., (2001a). The impact on revenue of increasing patient volume at surgical suites with relatively high operating room utilization. *Anesthesia & Analgesia 92(5)*, 1215-1221

Dexter, F., Macario, A. and O'Neill, L., (2000a). Scheduling surgical cases into overflow block time - computer simulation of the effects of scheduling strategies on operating room labour costs. *Anesthesia and Analgesia 90(4)*, 980-988

Dexter, F., Macario, A. and Traub, R.D., (1999b). Which algorithm for scheduling add-on elective cases maximizes operating room utilization? *Anesthesiology 91(5)*, 1491-1500

Dexter, F., Macario, A. and Traub, R.D., (2000b). Enterprise-wide patient scheduling information systems to coordinate surgical clinic and operating room scheduling can impair operating room efficiency. *Anesthesia & Analgesia 91(3)*, 617-626

Dexter, F., Macario, A., Lubarsky, D.A. and Burns, D.D., (1999c). Statistical Method to evaluate management strategies to decrease variability in operating room utilization. *Anesthesiology 91(1)*, 262-274

Dexter, F., Macario, A., Traub, R.D. and Lubarsky, D.A., (2003). Operating room utilization alone is not an accurate metric for the allocation of operating room block time to individual surgeons with low caseloads. *Anesthesiology 98(5)*, 1243-1249

Dexter, F., Macario, A., Traub, R.D., Hopwood, M. and Lubarsky, D.A., (1999a). An operating room scheduling strategy to maximize the use of operating room block time: computer simulation of patient scheduling and survey of patients' preferences for surgical waiting times. *Anesthesia and Analgesia 89(1)*, 7-20

223

Dexter, F., Traub, R.D. and Lebowitz, P., (2001b). Scheduling a delay between different surgeons' cases in the same operating room on the same day using upper prediction bounds for case durations. *Anesthesia and Analgesia 92(4)*, 943-946

Dexter, F., Traub, R.D., Fleisher, L.A. and Rock, P., (2002d). What sample sizes are required for pooling surgical case durations among facilities to decrease the incidence of procedures with little historical data? *Anesthesiology 96(5)*, 1239-1236

Dexter, F., Willemsen- Dunlap, A. and Lee, J.D., (2007). Operating room managerial decision-making on the day of surgery with and without computer recommendations and status displays. *Anesthesia and Analgesia 105(2)*, 419-429

Dowsland, K.A., (1995), *Simulated Annealing*. In Reeves, C.R., (1995). *Modern Heuristic Techniques for Combinatorial Problems*. Guildford: McGraw-Hill. pp. 20-63

Eden, C. and Ackermann F., (2001). *SODA – The Principles*. In Rosenhead, J. and Mingers, J. (eds.), (2001), *Rational Analysis for a Problematic World*, Wiley: Chichester. pp. 21-42

Eden, C., Ackermann, F., Bryson, J.M., Richardson, G.P., Andersen, D.F. and Finn, C.B., (2009). Integrating modes of policy analysis and strategic management practice: requisite elements and dilemmas. *Journal of the Operational Research Society 60(1),* 2-13

Eldabi, T., Paul, R.J. and Young, T., (2007). Simulation modelling in healthcare: reviewing the legacies and investigating the futures. *Journal of the Operational Research Society 58(2)*, 262-270

Epstein, R.H. and Dexter, F., (2002). Uncertainty in knowing the operating rooms in which cases were performed has little effect on operating room allocations or efficiency. *Anesthesia and Analgesia 95(6)*, 1726-1730

Ernst, E.A., Hoppel, C.L., Lorig, J.L. and Danielson, A.R., (1977). Operating room scheduling by computer. *Anesthesia and Analgesia 56(6)*, 831-835

Everett, J.E., (2002). A decision support simulation model for the management of an elective surgery waiting system. *Health Care Management Science 5(2)*, 89-95

Fei, H., Chu, C. and Meskens, N., (2009b). Solving a tactical operating room planning problem by a column-generation-based heuristic procedure with four criteria. *Annals of Operations Research 166(1)*, 91-108

Fei, H., Chu, C., Meskens, N. and Artiba, A., (2008). Solving surgical cases assignment problem by a branch-and-price approach. *International Journal of Production Economics 112(1)*, 96-108

Fei, H., Meskens, N. and Chu, C., (2010). A planning and scheduling problem for an operating theatre using an open scheduling strategy. *Computers and Industrial Engineering 58(2)*, 221-230

Fei, H., Meskens, N., Combes, C. and Chu, C., (2009a). The endoscopy scheduling problem: A case study with two specialised operating rooms. *International Journal of Production Economics 120(2)*, 452-462

Franco, L.A. and Lord, E., (2011). Understanding multi-methodology: evaluating the perceived impact of mixing methods for group budgetary decisions. *Omega 39(3)*, 362-372

Fu, M.C., (2002). Optimization for simulation: theory vs. practice. *INFORMS Journal on Computing 14(3),* 192-215

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

Gallivan, S., Utley, M., Treasure, T. and Valencia, O., (2002a). Booked inpatient admissions and hospital capacity. *British Medical Journal 324(7349)*, 280-282

Gallivan, S., Utley, M., Treasure, T. and Valencia, O., (2002b). Authors respond to criticisms of booked inpatient admissions and hospital capacity study. *British Medical Journal* 325(7361), 443

Gemmel, P., Van Dierdonck, R., (1999). Admission scheduling in acute care hospitals: does the practice fit with the theory? *International Journal of Operations and Production Management 19(9)*, 863-878

Gerchak, Y., Gupta, D. and Henig, M., (1996). Reservation planning for elective surgery under uncertain demand for emergency surgery. *Management Science 42(3)*, 321-334

Goldberg, D.E., (1989) Genetic Algorithms in Search, Optimization, and Machine Learning, Addison-Wesley: Reading

Gorunescu, F., McClean, S.I. and Millard, P.H., (2002). A queueing model for bed-occupancy management and planning of hospitals. *Journal of the Operational Research Society 53(1)*, 19-24

Greenberg, A.G., (1992). How fair is queuing? *Journal of the Association for Computing Machinery 39(3*), 568-598

Greenberg, H.J., (1988). Diagnosing infeasibility for min-cost network flow models. ii. primary infeasibility. *IMA Journal of Mathematics Applied to Business and Industry 2(1)*, 1-12

Greenberg, H.J. and Murphy, F.H., (1991). Approaches to diagnosing infeasible linear programs. *ORSA Journal on Computing 3(3)*, 253-261

Guerriero, F. and Guido, R., (2011). Operational research in the management of the operating theatre: a survey. *Health Care Management Science 14(1)*, 89-114

Guieu, O. and Chinneck, J.W., (1999). Analysing infeasible mixed-integer and integer linear programs. *Informs Journal on Computing 11(1)*, 63-77

Guinet, A. and Chaabane, S., (2003). Operating theatre planning. *Int. J. Production Economics 85(1)*, 69-81

Gupta, D. and Denton, B., (2008). Appointment scheduling in health care: challenges and opportunities. *IIE Transactions 40(9)*, 800-819

Gustafson, D.H., (1968). Length of stay: prediction and explanation. *Health Services Research 3(1)*, 12-34

Hans, E.W., Nieberg, T. and van Oostrum, J.M., (2007). Optimization in surgery planning. *Medium Econometrische Toepassingen 15(1)*, 20-28

Hans, E.W., Wullink, G., van Houdenhoven, M. and Kazemier, G., (2008). Robust surgery loading. *European Journal of Operational Research 185(3)*, 1038-1050

Harper, P.A. and Pitt, M.A., (2004). On the challenges of healthcare modelling and a proposed project life cycle for successful implementation. *Journal of the Operational Research Society 55(6)*, 657-661

Harper, P.R. and Shahani, A.K., (2002). Modelling for the planning and management of bed capacities in hospitals. *Journal of the Operational Research Society 53(1)*, 11-18

Harwood, S.A, (2011). Mixing methodologies and paradigmatic commensurability. *Journal of the Operational Research Society 62(4)*, 806-809

Hindle, G.A. and Franco, L.A., (2009). Combining problem structuring methods to conduct applied research: a mixed methods approach to studying fitness-to-drive in the UK. *Journal of the Operational Research Society 60(12)*, 1637-1648

Ho, C.-J. and Lau, H.-S., (1999). Evaluating the impact of operating conditions on the performance of appointment scheduling rules in service systems. *European Journal of Operational Research 112(3)*, 542-553

Horlick-Jones, T., and Rosenhead, J., (2006). The uses of observation: combining problem structuring methods and ethnography. *Journal of the Operational Research Society 58(5)*, 588-601

Hollocks, B.W., (2005). Forty years of discrete-event simulation – a personal reflection. *Journal of the Operational Research Society 57(12)*, 1383-1399

Huegler, P.A. and Vasko, F.J, (1997). A performance comparison of heuristics for the total weighted tardiness problem. *Computers in Industrial Engineering 32(4)*, 753-767

Humphreys, H., Stacey, A.R. and Taylor E. W., (1995). Survey of operating theatres in Great Britain and Ireland. *Journal of Hospital Infection 30(4)*, 245-252

Isken, M.W., Ward, T.J. and Littig, S.J., (2011). An open source software project for obstetrical procedure scheduling and occupancy analysis. *Health Care Management Science 14(1)*, 56-73

Jackson, M.J., (2011). The multi-methodology debate: a response to Harwood. *Journal of the Operational Research Society 62(4)*, 811-813

Jackson, R.R.P., (1964). Appointments systems in hospitals and general practice. *Operational Research Quarterly 15(3)*, 219-237

Jebali, A., Hadj Alouane, A.B. and Ladet, P., (2006). Operating rooms scheduling. *International Journal of Production Economics 99(1-2)*, 52-62

Jerbi, B. and Kamoun, H., (2011). Multiobjective study to implement outpatient appointment system at Hedi Chaker Hospital. *Simulation Modelling Practice and Theory 19(5)*, 1363-1370

Jones, S.A., Joy, M.P. and Pearson, J., (2002). Forecasting demand of emergency care. *Health Care Management Science 5(4)*, 297-305

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

Kelly, F.P., Maulloo, A.K. and Tan, D.K.H., (1998). Rate control for communication networks: shadow prices, proportional fairness and stability. *Journal of the Operational Research Society 49(3)*, 237-252

Khammuang, K., Abdekhodaee, A. and Wirth, A., (2007). On-line scheduling with forbidden zones. *Journal of the Operational Research Society 58(1)*, 80-90

Kim, H. and Han, Y., (2005). A proportional fair scheduling for multicarrier transmission systems. *IEEE Communications Letters 9(3),* 210-212

Klassen, K.J. and Rohleder, T.R., (1996). Scheduling outpatient appointments in a dynamic environment. *Journal of Operations Management 14(2)*, 83-101

Klassen, K.J. and Rohleder, T.R., (2000). Using client-variance information to improve dynamic appointment scheduling performance. *Omega International Journal of Management Science 28(3)*, 293-302

Klassen, K.J. and Rohleder, T.R., (2002). Rolling horizon appointment scheduling: a simulation study. *Health Care Management Science 5(3)*, 201-209

Kolliopoulos, S.G. and Steiner, G., (2007). Approximation algorithms for scheduling problems with a modified total weighted tardiness objective. *Operational Research Letters 35(5)*, 685-692

Kotiadis, K. and Mingers, J., (2006). Combining PSMs with hard OR methods: the philosophical and practical challenges. *Journal of the Operational Research Society 57(7)*, 856-867

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

Lagergren, M., (1998).What is the role and contribution of models to management and research in the health services? A view from Europe. *European Journal of Operational Research 105(2)*, 257-266

Lane, D.C., Monefeld, C. and Husemann, E., (2003). Client involvement in simulation model building: hints and insights from a case study in a London hospital. *Health Care Management Science 6(2)*, 105-116

Lamiri, M., Xie, X., Dolgui, A. and Grimaud, F., (2008). A stochastic model for operating rooms planning with elective and emergency surgery demands. *European Journal of Operational Research 185(3)*, 1026-1037

Lamiri, M., Grimaud, F. and Xie, X., (2009). Optimization methods for a stochastic surgery planning problem. *International Journal of Production Economics 120(2)*, 400-410

Lee A.H., Gracey M., Wang K. and Yau K.K.W., (2005). A robustified modelling approach to analyze paediatric length of stay. *Annals of Epidemiology 15(9)*, 673-677

Lee, A.H., Ng, A.S.K. and Yan, K.K.W., (2001). Determinants of maternity length of stay: a gamma mixture risk-adjusted model. *Health Care Management Science 4(4)*, 249-255

Lin, B.M.T., (2007). Complexity results for single-machine scheduling with positional learning effects. *Journal of the Operational Research Society 58(8)*, 1099-1102

Liu, M., Xu, Y., Chu, C., and Zheng, F., (2009). Online scheduling to minimize modified total tardiness with an availability constraint. *Theoretical Computer Science 410(47-49)*, 5039-5046

Liu, M., Zheng, F., Chu, C. and Xu, Y., (2011). Optimal algorithms for online scheduling on parallel machines to minimize the makespan with a periodic availability constraint. *Theoretical Computer Science 412(39)*, 5225-5231

Liu, Y., Chu, C. and Wang, K., (2011). A new heuristic for the operating room scheduling problem. *Computers and Industrial Engineering 61(3)*, 865-871

Longo, M. and Masella, C., (2002). Organisation of operating theatres: an Italian benchmarking study. *Journal of Operations and Production Management 22(4)*, 425-444

Lu, X., Sitters, R.A. and Stougie, L., (2003). A class of on-line scheduling algorithms to minimize total completion time. *Operations Research Letters 31(3)*, 232-236

MacCormick, A.D., Collecut, W.G. and Parry, B.R., (2003). Prioritizing patients for elective surgery: a systematic review. *Australia and New Zealand Journal of Surgery 73(8)*, 633-642

Magerlein, J.M. and Martin, J.B., (1978). Surgical demand scheduling: a review. *Health Services Research 13(4)*, 418-433

Marcon, E. and Dexter, F., (2007). An observational study of surgeons' sequencing of cases and its impact on postanesthesia care unit and holding area staffing requirements at hospitals. *Anesthesia and Analgesia 105(1)*, 119-126

Marcon, E., Kharraja, S. and Simonnet, G., (2003b). The operating theatre planning by the follow-up of the risk of no realization. *International Journal of Production Economics 85(1)*, 83-90

Marcon, E., Kharraja, S., Smolski, N., Luquett, B. and Viale, J.P., (2003a). Determining the Number of Beds in the Postanesthesia Care Unit: A Computer Simulation Flow Approach. *Anesthesia and Analgesia 96(5)*, 1415-2423

Maros, I. and Mitra G., (1996). *Simplex Algorithms*. In Beasley J.E. (ed.) (1996) *Advances in Linear and Integer Programming*, Oxford: Oxford University Press, 1-46

Mazzei, J.W., (1999). Maximizing operating room utilization: a landmark study. *Anesthesia and Analgesia 89(1)*, 1-2

McIntosh, C., Dexter, F. and Epstein, R.H., (2006). The impact of service-specific staffing, case scheduling, turnovers, and first-case starts on anaesthesia group and operating room productivity: a tutorial using data from an Australian hospital. *Anesthesia and Analgesia 103(6)*, 1499-1516

Min, D. and Yin, Y., (2010). An elective surgery scheduling problem considering patient priority. *Computers and Operations Research 37(6)*, 1091-1099

Minitab (version 16), software, (2010), USA:Minitab Inc.

Mingers, J., (2001). *Multimethodology – Mixing and Matching Methods*. In Rosenhead, J. and Mingers, J., (eds.), (2001). *Rational Analysis for a Problematic World*, Wiley: Chichester. pp. 289-310

Mingers, J. and Brocklesby, J., (1997). Multimethodology: towards a framework for mixing methodologies. *Omega International Journal of Management Science 25(5)*, 489-509

Mingers, J., (2011). Clarification or confusion: response to Harwood. *Journal of the Operational Research Society 62(4)*, 809-811

Moore, J. M., (1968). An *n* job, one machine sequencing algorithm for minimizing the number of late jobs. *Management Science 15(1)*, 102-109

Mullen, P.M., (2003). Prioritising waiting lists: how and why? *European Journal of Operational Research 150(1)*, 32-45

Munro, I. and Mingers, J., (2002). The use of multimethodology in practice – results of a survey of practitioners. *Journal of the Operational Research Society 53(4)*, 369-378

Murat, A. and Nepal, B., (2010). The effect of scheduling policies on operating room overtime performance. *International Journal of Services and Operations Management 7(2)*, 231-251

Ogulata, S.N. and Erol, R., (2003). A hierarchical multiple criteria mathematical programming approach for scheduling general surgery operations in large hospitals. *Journal or Medical Systems 27(3)*, 259-270

Olivares, M., Terwiesch, C. and Cassarola, L., (2008). Structural estimation of the newsvendor model: an application to reserving operating room time. *Management Science 54(1)*, 41-55

O'Neill, L. and Dexter, F., (2007). Tactical increases in operating room block time based on financial data and market growth estimates from data envelopment analysis. *Anesthesia and Analgesia 104(2)*, 355-358

Ozkarahan, I., (2000). Allocation of surgeries to operating rooms by goal programing. *Journal or Medical Systems 25(6)*, 339-378

Pandit, J.J. and Carey, A., (2006). Estimating the duration of common elective operations: implications for operating list management. Anaesthesia 61(8), 768-776

Pandit, J.J., Westbury, S and Pandit, M., (2007). Operating room efficiency in the national health service. *Anesthesia and Analgesia 105(1)*, 290-291

Patrick, J. and Puterman, M.L., (2007). Improving resource utilization for diagnostic services through flexible impatient scheduling: a method for improving resource utilisation. *Journal of the Operational Research Society 58(2)*, 235-245

Patrick, J., Puterman, M.L. and Queyranne, M., (2008). Dynamic multipriority patient scheduling for a diagnostic resource. *Operations Research 56(6)*, 1507-1525

Peltokorpi, A., (2011). How do strategic decisions and operative practices affect operating room productivity? *Health Care Management Science 14(4),* 370-382

Penn, M.L., (2006). *Optimisation of Hampshire Adult Services Gearing up for Future Demands*, Dissertation (MSc), University of Southampton

Persson, M.J. and Persson J.A., (2010). Analysing management policies for operating room planning using simulation. *Health Care Management Science 13(2)*, 182-191

Pham, D-N. and Klinkert, A., (2008). Surgical case scheduling as a generalized job shop scheduling problem. *European Journal of Operational Research 185(3)*, 1011-1025

Pidd, M., (1998). *Computer Simulation in Management Science*. 4th ed. Chichester: Wiley

Pidd, M., (2003). *Tools for Thinking*. Wiley: Chichester

Pidd, M., (2004). *Bringing it all together*. In: Pidd, M. (ed.), *Systems Modelling: Theory and Practice*, Wiley: Chichester. pp. 197-297

Pollack, J., (2009). Multimethodology in series and parallel: strategic planning using hard and soft OR. *Journal of the Operational Research Society 60(2)*, 156-167

Potts, C.N. and Strusevich, V.A., (2009). Fifty years of scheduling: a survey of milestones. *Journal of the Operational Research Society 60(S1)*, S41-S68

Pruhs, K., Sgall, J. and Torng, E., (2004). *Online Scheduling*. In. Leung, J. Yl-T., (2004). *Handbook of Scheduling: Algorithms, Models and Performance Analysis*. Chapman and Hall/CRC: London. pp. 15-1 to 15-43

Raz, H., Levy, H. and Avi-Itzhak, B., (2004). A resource allocation queueing fairness measure. *Performance Evaluation Review 32(1)*, 130-141

Reeves, C.R and Beasley, J.E., (1995). *Introduction,* In: Reeves, C.R. (ed.), (1995). *Modern Heuristic Techniques for Combinatorial Problems*, Guildford: McGraw-Hill. pp. 1-14

Ridley, S., Jones, S., Shahani, A., Brampton, W., Nielsen, M. and Rowan, K., (1998). Classification trees. *Anaesthesia 53(9)*, 833-840

Riihimaki, J., Sund, R. and Vehtari, A., (2010). Analysing the length of care episode after hip fracture: a nonparametric and parametric Bayesian approach, *Health Care Management Science 13(2)*, 170-181

Riise, A. and Burke, E.K., (2011). Local search for the surgery admission planning problem. *Journal of Heuristics 17(4)*, 389-414

Robinson, S., (2001). Soft with a hard centre: discrete-event simulation in facilitation. *Journal of the Operational Research Society 52(8)*, 905-915

Robinson, S., (2005). Discrete-event simulation: from the pioneers to the present, what next? *Journal of the Operational Research Society 56(6)*, 619-629

Rogers, H., Warner, J., Steyn, R., Silvester, K., Pepperman, M. and Nash, R., (2002), .Mathematical model misses the point. *British Medical Journal 324(7349)*, 1336

Rohleder, T. R., Bischak, D.P. and Baskin, L.B., (2007). Modelling patient service centers with simulation and system dynamics. *Health Care Management Science* 10(1), 1-12

Roland, B., Di Martinelly, C., Riane, F., and Pochet, Y., (2009). Scheduling an operating theatre under human resource constraints. *Computers and Industrial Engineering 58(2)*, 212-220

Sachdeva, R., Williams, T. and Quigley, J., (2006). Mixing methodologies to enhance the implementation of healthcare operational research. *Journal of the Operational Research Society 58(2)*, 159-167

Sandmann, W., (2011). Quantitative fairness for assessing perceived service quality in queues. *Operational Research,* doi:10.1007/s12351-011-0111-9

Santibanez, P., Begen, M. and Atkins, D., (2007). 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

Schmidt, G., (1988). Scheduling independent tasks with deadlines on semi-identical processors. *Journal of the Operational Research Society 39(3)*, 271-277

Schwiegelshohn, U. and Yahyapour, R., (2000). Fairness in parallel job scheduling. *Journal of Scheduling 3(5)*, 297-320

Shi, Y. and Ye, D., (2008). Online bin packing with arbitrary release times. *Theoretical Computer Science 390(1)*, 110-119

Sciomachen, A., Tanfani, E. and Testi, A., (2005). Simulation models for optimal schedules of operating theatres. *International Journal of Simulation 6(12-13)*, 26-34

Sier, D., Tobin, P. and McGurk, C., (1997). Scheduling surgical procedures. *Journal of the Operational Research Society 48(9)*, 884-891

Simul8 Help Centre, (2012b). Simulation a guide to the technique: Warm up time. http://www.simul8.com/support/help/doku.php?id=gettingstarted:techguide:warmup accessed: 15/11/12

Simul8 Help Centre, (2012a). Simulation a guide to the technique: Accuracy of results. http://www.simul8.com/support/help/doku.php?id=gettingstarted:techguide:accuracy accessed: 15/11/12

Singh C.H. and Ladusingh L., (2010). Inpatient length of stay, a finite mixture modelling analysis. *The European Journal of Health Economics 11(2)*, 119-126

Smyth, E.T.M., Humphreys, H., Stacey, A., Taylor, E.W., Hoffman, P. and Bannister, G., (2005). Survey of operating theatre ventilation facilities for minimally invasive

surgery in Great Britain and Northern Ireland: current practice and considerations for the future. *Journal of Hospital Infection 61(2)*, 112-122

Sobolev, B., Harel, D., Vasilakis, C. and Levy, A., (2008). Using the statecharts paradigm for simulation of patient flow in surgical care. *Health Care Management Science 11(1)*, 79-86

Strum, D.P., Sampson, A.R., May, J.H. and Vargas, L.G., (2000). Surgeon and type of anesthesia predict variability in surgical procedure times. *Anesthesiology 92(5)*, 1454-66

Strum, D.P., Vargas, L.G. and May, J.H., (1999). Surgical subspecialty block utilization and capacity planning: a minimal cost analysis model. *Anesthesiology 90(4)*, 1176-1185

Strum, D.P., Vargas, L.G., May, J.H. and Bashein, G., (1997). Surgical suit utilization and capacity planning: a minimal cost analysis model. *Journal of Medical Systems 21(5)*, 309-322

Su, M.-C., Lai, S.-C., Wang, P.-C., Hsieh, Y.-Z. and Lin, S.-C., (2011). A SOMO-based approach to the operating room scheduling problem. *Expert Systems with Applications 38(12)*, 15447-15454

Su, S. and Shih, C.-L., (2003). Managing a mixed-registration-type appointment system in outpatients clinics. *International Journal of Medical Informatics 70(1)*, 31-40

Tai, G. and Williams, P., (2011). Optimization of scheduling patient appointments in clinics using a novel modelling technique of patient arrival. *Computer Methods and Programs in Biomedicine*, doi. 10.1016/j.cmpb.2011.02.010

Tao, J., Chao, Z. and Tao, Y., (2010). An optimal semi-online algorithm for a single machine scheduling problem with bounded processing time. *Information Processing Letters 110(8-9)*, 325-330

Tanfani, E. and Testi, A., (2010). Improving surgical department performances via simulation and optimisation. *Proceedings of the IEEE Workshop on Health Care Management*, Venezia, Italy.

Testi, A., Tanfani, E. and Torre, G., (2007). A three-phase approach for operating theatre schedules. *Health Care Management Science 10(2)*, 163-172

Tyler, D.C., Pasquariello, C.A. and Chen, C.H., (2003). Determining optimum operating room utilization. *Anesthesia and Analgesia 96(4)*, 1114-1121

Utley, M., Gallivan, S., Treasure, T. and Valencia, O., (2003). Analytical methods for calculating the capacity required to operate an effective booked admissions policy for elective inpatient services. *Health Care Management Science 6(2)*, 97-104

Vanberkel, P.T., Bousherie R.J., Hans E.W., Hurlink, J.L., van Lent, W.A.M and Harten, W.H., (2011a). An exact approach for relating recovering surgical patient workload to the master surgical schedule. *Journal of the Operational Research Society 62(10)*, 1851-1860

Vanberkel, P.T., Bousherie R.J., Hans E.W., Hurlink, J.L., van Lent, W.A.M and Harten, W.H., (2011b). Accounting for inpatient wards when developing master surgical schedules. *Anesthesia and Analgesia 112(6)*, 1472-1479

Van Houdenhoven, M., Hans, E.W., Klein, J., Wullink, G. and Kazemier, G., (2007). A norm utilisation for scarce hospital resources: evidence from operating rooms in a Dutch university hospital. *Journal of Medical Systems 31(4)*, 231-236

Van Houdenhoven, M., van Oostrum, J.M., Wullink, G., Hans, E., Hurink, J.L., Bakker, J. and Kazemier, G., (2008). Fewer intensive care unit refusals and a higher capacity utilization by using a cyclic surgical case schedule. *Journal of Critical Care 23(2)*, 222-226

Van Houdenhoven, M., van Oostrum, J.M., Wullink, G., Hans, E., Wullink, G. and Kazemier, G., (2007). Improving operating room efficiency by applying bin-packing and

portfolio techniques to surgical case scheduling. *Anesthesia and Analgesia 105(3)*, 707-714

van Oostrum, J.M., Van Houdenhoven, M., Hurink, J.L., Hans, E.W., Wullink, G. and Kazemier, G., (2008). A master surgical scheduling approach for cyclic scheduling in operating room departments. *OR Spectrum 30(2)*, 355-374

Varaiya, P.P., (1972). *Notes on Optimisation*. New York: Van Nostrand Reinhold Company

Vasilakis, C., Sobolev, B.G., Kuramoto, L. and Levy, A.V., (2010). A simulation study of scheduling clinic appointments in surgical care: individual surgeon versus pooled lists. *Journal of the Operational Research Society 58(2)*, 1-10

Vissers, J.M.H., Adan, I.J.B.F. and Bekkers J.A., (2005). Patient mix optimization in tactical cardiothoracic surgery planning: a case study. *IMA Journal of Management mathematics 16(3)*, 281-304

Wachtel, R.E. and Dexter, F., (2008). Tactical increases in operating room block time for capacity planning should not be based on utilization. *Anesthesia and Analgesia 106(1)*, 215-26

Welch, J.D., (1964). Appointment systems in hospital outpatient departments, *Operational Research Quarterly 15(3)*, 224-232

Wiermann, A., (2011). Fairness and scheduling in single server queues. *Surveys in OR and Management Science 16(1)*, 39-48

Wiers, V.C.S., (1997). A review of the applicability of OR and AI scheduling techniques in practice. *Journal of Management Science 25(2)*, 145-153

Winston, W.L., (1994). *Operations Research, Applications and Algorithms 3rd Ed.*, California: Wadsworth

Wolstenholme, E.F., (1993). A case study in community care using systems thinking. *Journal of the Operational Research Society 44 (9)*, 925-934

Wong, L. and Lee, L.S., (2009). Heuristic placement routines for two-dimensional bin packing problem. *Journal of Mathematics and Statistics 5(4)* 334-341

Wullink, G., Van Houdenhoven, M., Hans, E.W., van Oostrum, J.M., van der Lans, M. and Kazemier, G., (2007). Closing Emergency operating rooms improves efficiency. *Journal of Medical Systems 31(6)*, 543-546

Zaharia, M., Borthakur, D., Sarma, J.S., Elmeleegy, K., Shenker, S. and Stoica, I., (2010). Delay scheduling: a simple technique for achieving locality and fairness in cluster scheduling. *EuroSys 2010*, April 2010.

Zhang, B., Murali, P., Dessouky, M.M. and Belson, D., (2008). A mixed integer programming approach for allocating operating room capacity. *Journal of the Operational Research Society 60(5)*, 663-673

Zhang, C.Y., Li, P., Rao, Y. and Guan, Z., (2008). A very fast TS/SA algorithm for the job shop scheduling problem. *Computers and Operations Research 35(1)*, 282-294

Zhu, Z., (2011). After the paradigm: why mixing-methodology theorising fails and how to make it work again. *Journal of the Operational Research Society 62(4)*, 784-798

Zonderland, M.E., Boucherie, R.J., Litvak, N. and Vleggeert-Lamkamp, C.L.A.M., (2010). Planning and scheduling of semi-urgent surgeries. *Health Care Management Science 13(3)*, 256-267