UNIVERSITY OF SOUTHAMPTON

FACULTY OF MATHEMATICAL STUDIES

OPERATIONAL RESEARCH

OPERATIONAL MODELLING FOR THE PLANNING AND MANAGEMENT OF HEALTHCARE RESOURCES

by

Paul Robert Harper

Doctor of Philosophy

February 2002

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF MATHEMATICAL STUDIES
OPERATIONAL RESEARCH

Doctor of Philosophy

OPERATIONAL MODELLING FOR THE PLANNING AND MANAGEMENT OF HEALTHCARE RESOURCES

By Paul Robert Harper

The provision of healthcare services is perhaps one of the largest and most complex industries worldwide. As one of the essential necessities to sustain life, it faces the consequences of increasing demand in times of limited financial resources and competing social needs. Providing the appropriate medical care involves decision-making in terms of planning and management of healthcare resources.

There is currently a great need to evolve a framework in which necessarily detailed, stochastic, flexible and user-friendly operational models, to aid both the planning and management of hospital resources, can be developed. Such a framework is considered and created within this research. Furthermore, detailed integrated simulation models for the planning and management of hospital beds, operating theatres, workforce needs and critical care services have been designed and built. An evolutionary development methodology has been adopted whereby the research work and model development has been guided by a number of steering groups within participating NHS Trusts.

The derived framework incorporates the need for sophisticated patient classification techniques to be adopted. In order to capture the uncertainty and variability amongst the patient population, a number of classification techniques have been considered and evaluated for their relative performances and practical usefulness. Healthcare issues representing both challenges and opportunities are explored in order to provide a basis for tentative conclusions about the current state of operational modelling for healthcare. A framework for the successful design and implementation of operational models in a healthcare environment is proposed.

Contents

List of Figures

List of Tables

Acknowledgments

This thesis has benefited enormously from the confluence of two highly regarded and respected academics, my supervisors Dr Arjan Shahani and Professor Russell Cheng. I am extremely grateful to both for passing on their knowledge, ideas and for generally motivating me throughout my research. I am also most grateful to my advisor Mrs Susan Meacock for patiently reading through the entire manuscript and for making numerous helpful suggestions, and in particular for her help with the statistical elements of my research on classification techniques.

This research would not have been possible without the valuable input, time and effort afforded by many dedicated NHS staff. I would like to take this opportunity to thank those at both The Royal Berkshire and Battle Hospitals NHS Trust and within the Portsmouth Hospitals NHS Trust. I am particularly indebted to Jana Dale, Heather Bunce, Sharon Kearnes, Peter Howlett and Ginette Alexander.

My time at the University of Southampton, throughout both my M.Sc and Ph.D, has been made that much more enjoyable as a consequence of the day-to-day support, assistance and humour of the clerical staff. Many thanks and warm wishes to them all on level 5. I have also benefited from an invasion of Brazilians into the OR group. My gratitude to Professor Valter de Senna, Israel Vieira and Andre Costa for their computer modelling assistance and attempts to get me speaking Portuguese.

My love and thanks to all my family for their continued love and support.

To my wife Julia, God's greatest blessing to me, who has brought so much excitement and happiness to my life since we met. For all of her boundless love, motivation and understanding during the writing of this thesis, which has made the completion of this work possible.

This thesis is also dedicated to my Lord Jesus Christ who guides my path through life and who has provided me with contentment and everlasting joy.

Chapter 1 – Introduction

1.1 Chapter Introduction

The purpose of this chapter is to establish the general context of the work presented in this thesis and provide a perspective into which the individual elements can be integrated. Firstly there is a clear statement of the project objectives followed by the general methodological approach adopted and research context. Details of the research participants and project management arrangements are then presented. The chapter concludes with an outline of the overall thesis structure, providing an overview of the document as a whole.

1.2 Research Objectives

This thesis concerns the planning and management of healthcare resources. The research work however may be broadly divided into the following distinct but integrated objectives:

- To research the use of the simulation methodology in support of management decision-making processes within the healthcare environment, with particular emphasis on hospital beds, human resources and operating theatre capacities.

- To explore the use of various classification techniques with particular emphasis on creating healthcare groupings. This will require an examination and comparison of existing methodologies and the need to improve the current theoretical knowledge.

- To derive a generic framework for modelling of healthcare resources. This should involve the linking of patient groupings from classification analyses with operational models.

- To develop necessarily detailed but flexible operational models to aid the chosen study areas and demonstrate the value of such models through various case studies.

- To evaluate the effectiveness of the adopted methodology and developed applications.

1.3 Research Context, Rationale and Methodology

As the NHS enters the new millennium, it does so as it faced its birth in 1948; an astonishing vision of comprehensive healthcare for an entire population. Now, as then, it is beset with controversies, both at local and national level. The nation it serves has changed considerably from those post war years. The balance of population living in large conurbations has changed, the age structure is different and people's expectations of the service they think it should provide change continuously.

As each year passes, the proportion of the population that can remember what healthcare was like before the establishment of the NHS declines. For those born in the fifties and afterwards, it is simply part of the way things are. Perhaps this, more than anything else, explains why attitudes to it have changed.

It will face many new challenges in this third millennium. The need to balance a supply driven service with finite funding, against demand fuelled by legitimate expectations will not go away. Advances in our understanding of the very generic chemistry of life will plunge the service into moral and ethical debates, the like of which have never been seen. Expensive treatment for "self inflicted" illnesses related to smoking and alcohol abuse will come increasingly into question. Some countries have already introduced voluntary euthanasia into law and in Britain the debates will continue to be heard.

Against this backcloth it would be absurd to assume that the delivery, planning and management of healthcare will remain static. The state of the NHS has seen many structural reforms by various governments since its birth and the continuing emphasis towards improved planning and management of resources within the NHS is clearly visible.

The Conservative government under John Major in 1991 produced "The Health of The Nation" reform document outlining their policies for the future role of the NHS (Department of Health, 1991). Mr Waldergrave, then Secretary of State for Health states, "A key feature of the reforms has been the establishment of a clear strategic role for health authorities". Later in the document he acknowledges the need to plan and manage resources carefully as there is only a finite amount available and a key objective is to make the best possible use of these resources. More recently, the Labour government lead by Tony Blair defined what it saw as "The *new* NHS; a Modern and Dependable Service" (Department of Health, 1997). Once again a key objective within the *new national performance framework*, as outlined in the document, is for "an evidence-based service in which change will be driven by improving the performance and efficiency of NHS Trusts; a combined approach of quality and efficiency to build a modern and dependable health service fit for the twenty first century and fit for the people of this nation".

The research topic concerns modelling for the planning and management of healthcare resources. This thesis aims to demonstrate that healthcare planning and management issues can be greatly benefited through an Operational Research approach. Together with the research participants, it was decided to concentrate on the planning and management of hospital capacities, which is currently a highly topical and important issue. More specifically, this research will focus on hospital beds, human resources and operating theatres capacities within a hospital environment. There is currently a great need to research these three highly critical, complex and interlinking elements of the hospital system. A generic framework for modelling of the hospital system is required.

Capacity planning in hospitals is largely a strategic decision. For example the total number of beds in a new hospital and the number of beds in various specialties are very major concerns; here the planning horizon could be about ten years. Management of available capacities could be from day to day or over longer periods such as winter months and summer months. An example would be a planned transfer of surgical beds to elderly medical patients in winter. Appropriate detailed models that can evaluate a variety of scenarios could be powerful tools for good planning and management decisions.

A common current practice is to plan and manage hospital capacities through a simple deterministic approach using average patient flows, average needs, average length-of-stay, average duration of surgical operations etc. Patient flows, patient needs, and utilisation of hospital capacities involve complexity, uncertainty, variability, constraints, and scarce resources. Mathematically speaking, a hospital corresponds to a complex stochastic system so that the common deterministic approach for planning and managing the system can be expected to be inadequate. Typically the deterministic approach will underestimate hospital requirements. The mathematical modelling approach of Operational Research is ideal for dealing with complexity, uncertainty, variability, constraints, and scarce resources and appropriate models can avoid the dangers of planning on the basis of average values only. This research is concerned with the development, solution, and validation of sufficiently detailed stochastic models for planning and managing hospital capacities.

Appropriate classification methods, linked to routine databases that are easy to use are needed. Classification methods such as Classification and Regression Tree (CART) that use binary splits have a great practical appeal. There is a need to improve the current theoretical knowledge about criteria for goodness of groupings and the effects of transformations. There is a need to test the accuracy and practical usefulness of a variety of classification methods for a number of variables. Examples include lengths of stay, theatre operations and nursing needs. The necessary theoretical and practical research for classification is an important element of this research. Various classification techniques will be compared using illustrative hospital datasets.

A combination of classification and simulation modelling can be used to describe, control, and monitor the flow of patients in a hospital. Patient flows correspond to queues in networks and it is not surprising to find that queuing models have been used extensively in the literature. Integer programming, forecasting, and simulation are other commonly used techniques. Although often viewed with success, much of the work to date has been project specific or has made highly simplifying assumptions.

There is a great need for flexible and sufficiently detailed models that can be used in a hospital at various levels. Such models will need to be stochastic in nature and take uncertainty, variability, complexity and use of resources into account. Ease of use including links with databases that are in common use is an important consideration. The models will be too complex for analytical solutions and simulation will be needed for solving the models. Validation and verification of complex models are clearly important issues and careful attention is given to these matters.

Historically within Operational Research, simulation methods have been successfully proven in manufacturing domains where processes can *relatively* easily be quantified and assessed. Ease of measurement and data collection are key factors in the construction and validation of the models used in such applications. In contrast to manufacturing systems, the service sector presents particular challenges to the modeller (Checkland, 1981). The potential benefit of the simulation methodology within the service industries is considerable. Arguably the challenges of service sector modelling are most acute in healthcare where issues of complexity, diversity and lack of quantitative data create particular problems (Davies, 1985). Such rewards for successful implementation of simulation within a healthcare environment provide ample incentive for this research. Furthermore, the development of a generic framework for simulation of hospital resources is paramount.

1.4 The Research Participants

The author worked closely with a number of healthcare organisations. A strong relationship was forged between all of the participants throughout the duration of the research work and permitted the evaluation of the research methodology and developed applications in a real-life environment. Ultimately this resulted in the research work to be of great practical use and benefit.

1.4.1 The Author

The author completed a B.Sc. (Hons) honours degree in Statistics at the University of Bath in 1995. He then went on to obtain a M.Sc. (with distinction) in Operational Research at the University of Southampton during 1996. Since this time he has worked for Cap Gemini Management Consultancy in London, before returning to Southampton as a Research Assistant within the Operational Research Department.

1.4.2 Institute of Modelling for Healthcare

The Institute of Modelling for Healthcare is a group within the Operational Research Department in the Faculty of Mathematical Studies at the University of Southampton. The Institute was established in the early 1990's, under the directorship of Dr. A. K. Shahani, due to the expansion of the healthcare modelling work following successful efforts in collaborative work and increases in research funding.
The Institute is pursuing and co-ordinating research on modelling for a number of healthcare areas. These areas include:

- Hospital capacity.
- Prevention, early detection and treatment of a particular disease.
- HIV/AIDS patient care.
- Intensive care capacity.
- Maternity care.
- Interventions for infectious diseases.
- Organisational issues such as interactions between primary and secondary care.

One of the aspects emphasised in the IMH approach is that developed models should account for the variability found in the real world. This distinguishes them from other more simplistic models that consider only average conditions, which often represents an unrealistic simplification.

1.4.3 The Royal Berkshire and Battle Hospitals NHS Trust

The Royal Berkshire and Battle Hospitals NHS Trust is an acute district general hospital based approximately 40 miles west of London in the Thames Valley. It serves a population of around 500,000 in a mixture of urban and semi-rural setting. The catchment population is on the whole relatively affluent and in good health, with patches of deprivation within the two main towns in the area.

The Trust covers the usual range of acute specialties including maternity services, with the exception of heart surgery and neurosurgery. The Trust treats approximately 260,000 outpatients, 65,000 inpatients and day cases and about 80,000 casualty attendees annually. The beds complement is 800 and the Trust employs over 3,500 people.

The hospital was founded on its London Road site in central Reading in 1839. The Battle Hospital, on Portman Road in west Reading, began life in 1891. The two hospitals have worked closely together since the time of the First World War. In 1987 they were brought together as one management unit. The Royal Berkshire and Battle NHS Trust was formed in April 1993.

During the middle of the 1990s the Trust, as did the whole of the NHS, came under a considerable pressure to treat ever increasing numbers of patients through a diminishing bed pool. The demand for certain surgical specialties was showing large increases. At same time numbers of medical emergency admissions were on the rise.

Internally the Trust launched a programme of consolidation of services from the current two locations within Reading to a single site. The building constraints meant that the pressure on space would become even greater, so that the number of beds that would be provided demanded higher bed utilisation and more accurate scheduling of other associated services.

The Trust had to make some fundamental changes to its way of working in order to meet these new challenges. An analysis of cost drivers confirmed what everybody intuitively felt; that any significant efficiencies could only be achieved through radical systems review and redesign of the whole system. This recognition led to the launch of a re-engineering programme within the Trust.

The Trust already had some limited experience of process oriented improvement projects, and now took the bold step to run a Trust wide programme that would approach the redesign from a strategic, whole system point of view. Very few constraints were placed on the staff working on the project and a much simplified organisational model was born (see Figure 1.1).

The model defined two categories of patients:

- **Planned** patients; and
- **Unplanned** patients.

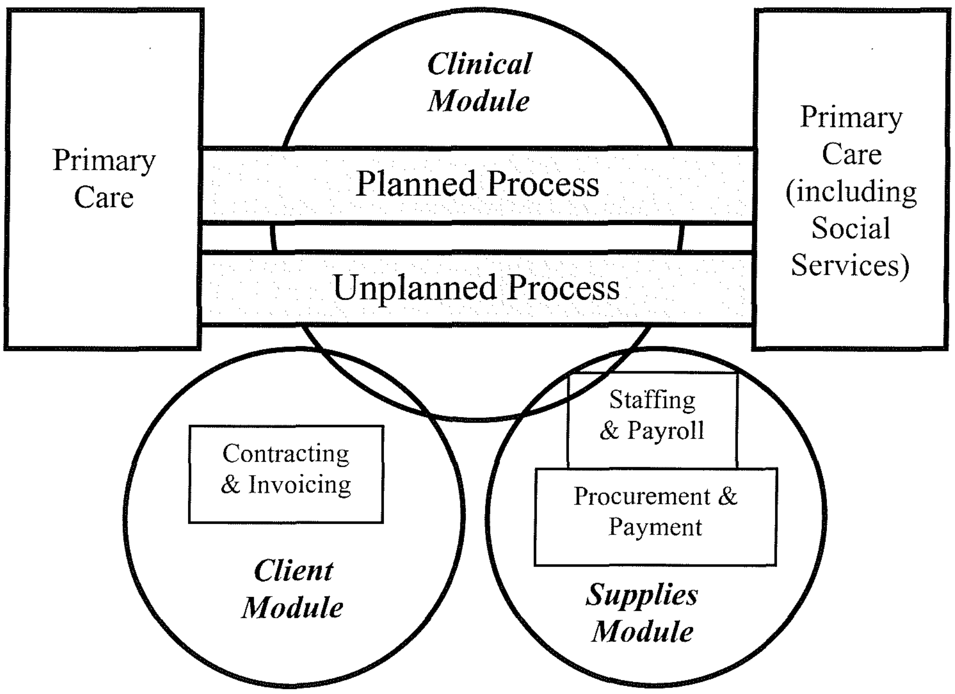


Figure1.1: Royal Berkshire and Battle NHS Trust process model

This simplicity was a departure from the established groupings and at first was not universally accepted. The logic behind the grouping, however, is straightforward. For 'planned' patients we know in advance that they are coming in and why. We can therefore plan, book, and schedule etc. what we are going to do. In fact the process itself is not too dissimilar for the 'unplanned' patients. Although they present unexpectedly, they still pass through more or less standard processes of diagnosis and treatment.

Having developed a new process model the Trust wanted to test its validity using established statistical and mathematical methods. The Trust also wanted to model and evaluate detailed options within the new design. Apart from this 'one off' application, the Trust was also looking to develop a forecasting and planning tool that could be used in long term planning and during the annual business planning cycle.
The Trust evaluated a number of approaches and decided to work jointly with the Institute of Modelling for Healthcare at Southampton University. The main reasons for choosing this particular establishment was their experience in building highly detailed and realistic models both at global and detailed local levels.

1.4.4 Portsmouth Hospitals NHS Trust

PORTSMOUTH HOSPITALS NHS TRUST

Portsmouth Hospitals NHS Trust is one of the largest in England, with an annual income of around £240 million (2001/02). It provides acute healthcare services for almost a million people; 550,000 covered by Portsmouth and South East Hampshire and many more from the surrounding counties.

The Trust has a regional specialty in its Renal and Transplant Unit and is a designated Cancer Centre. It is involved in the training of nurses and doctors at undergraduate level from the University of Southampton and postgraduate teaching.

Based at two locations in the City of Portsmouth four miles apart, the Trust effectively operates out of three main sites, as St Mary's Hospital (SMH) is divided in two by a main road. St Mary's Hospital West Wing is an historic 'workhouse' hospital and St Mary's East Wing is a former infectious diseases hospital, both dating from the nineteenth century. Queen Alexandra Hospital (QAH) is located in the north of the city in Cosham. It largely comprises two tower blocks of seven floors opened in 1978.

This is connected to the original 1908 military hospital, which is leased by Portsmouth Healthcare NHS Trust to provide acute elderly services.

Separately, neither Queen Alexandra Hospital nor St Mary's Hospital represents a full District General Hospital (DGH), but together they provide the full range of services expected of a DGH, albeit with services scattered around three sites. This leads to duplication of services in some instances, and the current cross-site dislocation of such services has a heavy cost premium of several million pounds per annum. The use of inappropriately configured, ageing building stock contributes to these problems and generates a backlog maintenance requirement estimated at £34 million (1994), without providing better facilities for patients. This historically generated pattern of split site working damages service delivery.

Internally the Trust launched a consolidation and reconfiguration programme of facilities. The following key strategic objectives were identified:

- To re-engineer the Trust's acute services to ensure the effective delivery of patient care on a single acute hospital at Queen Alexandra Hospital.
- To rationalise existing sites and create a Community Hospital Plus on the St Mary's Hospital site (a community hospital with extended facilities).
- To secure investment for the redevelopment of facilities to support the Service Development Strategy through the Private Finance Initiative (PFI).
- To provide specified services delivered by the Trust in local primary care and community settings.

The Private Finance Initiative (PFI) programme is a government-run initiative to attract private sector investment into health to permit the building of new hospitals and the development of new services. The Trust, supported by the local health authority, submitted a Strategic Outline Case (SOC), which demonstrated the benefits to the local community of the PFI. Portsmouth's application was successful, which represented a major success to the Trust. Portsmouth were then asked to submit a fully costed Outline Business Case (OBC), to appraise the options for delivering such benefits and recommend a preferred option.

A PFI proposal has to demonstrate value-for-money against the range of options considered. In the case of Portsmouth, the preferred option is the development of a single acute hospital on the Queen Alexandra Hospital site and the development of a community hospital on the St Mary's Hospital West Wing site. The prospects for a successful PFI partnership and development are excellent. Adjacent to the new build site is an additional 5.7 acres of land obtained from Portsmouth City Council as part of a land exchange deal. The outline planning permission includes multi-level car parks for patients and visitors and facilities for management of non-clinical services, such as laundry, maintenance and clinical waste incineration.

The significant planned investment, through the PFI process, is intended to develop Queen Alexandra Hospital as one of the most modern in this country and to deliver the highest possible standards of care to patients. The PFI gives the Trust an excellent opportunity to reconfigure the current structure of hospital processes and design the layout of the new building development. The Trust launched a consolidation and reconfiguration programme of facilities known as the Processes to Improve Care (PIC) project. PIC initially concentrated on ways of improving the process of patient care to deliver quality benefits, increase the number of patients treated, or reduce costs, or a combination of all three. In order to submit a successful and realistic OBC, as required under the PFI scheme, the Trust needed to quantify the benefit of re-configured processes as identified during PIC. There was a vital need to calculate the likely bed compliment for the entire hospital.

The Trust required an operational model to capture the flow of patients through the re-configured hospital system. Various scenarios needed to be examined and numbers of beds had to be quantified and costed. The Trust acknowledged that internally it did not have the necessary personnel with the skills and expertise to undertake this critical component of the OBC. Traditionally it had used simple spreadsheet calculations but had grave misgivings about the accuracy of these in the past. They appreciated that a more sophisticated and mathematically correct methodology was required in the form of operational models to capture the complex and variable flows of patients through Portsmouth Hospitals NHS Trust.

1.4.5 Critical Care Units

A number of critical care units also participated in the work. Involvement ranged from the transfer of knowledge regarding rules governing admissions and flows of patients through a critical care unit, through to the supply of data for validation and use of the developed model (see Chapter 8). For the purposes of this thesis, the participating units wish to remain anonymous.

1.5 Supervisory Arrangements

The supervisory framework for the Ph.D. was provided from a number of sources. The primary academic axes of supervision came from internal supervisors within the Operational Research Department at the University of Southampton. Such supervisors oversaw the development and advancement of methodologies as outlined in the general research context. Alongside this academic source, a key motivation of the Ph.D. was to test the efficacy of the developed methodologies by their application in the development of realistic and practically useful applications for use by hospital managers themselves. To meet this need, a more managerial supervisory role was adopted by the participating hospitals themselves, namely the Royal Berkshire and Battle NHS Trust (section 1.4.3), the Portsmouth Hospitals NHS Trust (section 1.4.4) and participating critical care units (section 1.4.5). Considerable time was spent working within the Trusts, enabling the author to fully understand local issues, processes and to build strong relations with a number of key personnel within the hospitals. Steering groups were established (Appendix A) which consisted of a number of healthcare managers from different departments within each Trust, as well as the academic supervisors from the University. Steering group meetings were held at regular intervals throughout the duration of the research and a range of reports and other documents were produced for the participating Trusts.

It is believed that the combination of academic and hospital supervisors enabled this research to be of great practical use and benefit to the healthcare profession. As a consequence of the generic framework and flexibility of the developed operational models as discussed in this thesis, they have since been successfully used by a variety of Trusts across the UK and have been presented at a number of domestic and international conferences.

1.6 Thesis Structure

The main body of the thesis is divided into ten chapters which are designed to follow a logical sequence through the subject matter. Thus Chapters 1, 2 and 3 are used to set the context for the work; Chapters 4 through to 8 describe the research work itself; and Chapters 9 and 10 consist of the discussion of the research findings, conclusions and recommended further research.

In general the first section of each chapter is used to introduce the content and the final section gives a synopsis of the ground covered with salient points. Technical detail and description relating to specific aspects of the research work are included in appendices at the end.

Figure 1.2 below graphically depicts the thesis structure in terms of its components and illustrates how the elements are logically integrated within the thesis as a whole.

1.7 Chapter Summary

The general research context and methodology has been outlined; that of the need for the development, solution, and validation of sufficiently detailed stochastic models for planning and managing healthcare resources. This research will focus on hospital beds, human resources and operating theatre capacities within a hospital environment. There is currently a great need to research these three highly critical, complex and interlinking elements of the hospital system. A generic framework for modelling the hospital system is required. Participating NHS Trusts provide the real-life context and needs for developing operational models.

Basic Information	
Chapter 1 *Introduction*	Outline of general context, parties involved and research objectives.

Background Context	
Chapter 2 *Planning and Management of Healthcare*	History and structure of the NHS and an overview of the issues surrounding the planning and management of resources.
Chapter 3 *Operational Modelling for Healthcare*	Survey and review of existing literature and application.

Research Work	
Chapter 4 *Requirements, Methodology and Generic Framework*	Description of the research approach and general methods adopted.
Chapter 5 *Patient Classification Techniques*	Discussion on creating patient groupings with descriptions and comparisons between a number of classification techniques.
Chapter 6 *A Simulation Model for Hospital Resources*	Detailed description and validation of a hospital resources simulation developed within the evolved generic framework.
Chapter 7 *Hospital Case Studies*	Illustrative case studies of the hospital resources simulation model in use.
Chapter 8 *Simulation Models for Critical Care Services*	Description of the developed models for Critical Care Services with model validation and illustrative applications.

Discussion and Conclusions	
Chapter 9 *Moving Forward: Challenges and Opportunities*	Analysis on the current state and future of operational modelling within healthcare settings.
Chapter 10 *Conclusions and Further Research*	Summary of ground covered with thoughts and recommendations.

Figure 1.2: Overview of thesis structure

Chapter 2 – Planning and Management of Healthcare

2.1 Chapter Introduction

This chapter provides a review of the history and structure of the NHS. It will identify the many structural reforms that the NHS has witnessed since its birth. Furthermore, this chapter will establish the current issues that have resulted in the need for detailed operational models for the planning and management of healthcare resources.

2.2 The Health of a Nation

The National Health Service (NHS) was founded in 1948, and since then has become one of the institutions of the State. It is one of the best loved in principle, most vilified in debate and least understood parts of the welfare provision of this country. We take the NHS for granted now, but it is only just over 50 years ago that healthcare was a luxury not everyone could afford. It is difficult today for us to imagine what life must have been like without free healthcare and the difference that the arrival of the NHS made to people's lives. Many of the references provided in the following sections may be found in Rivett (1998), which gives a more detailed description of the health of the nation and the birth of the NHS.

2.2.1 Life before the NHS

Just before the creation of the NHS, the services available were, as you might expect, the same as after; no new hospitals were built nor hundreds of new doctors employed. What was different was that poor people often went without medical treatment, relying instead on dubious, and sometimes dangerous, home remedies or on the charity of doctors who gave their services free to their poorest patients. Access to a doctor was free to workers, who were on lower pay, but this didn't necessarily cover their wives or children, nor did it cover other workers or those with a better standard of living. Hospitals charged for services, though sometimes poorer people would be reimbursed. Even so, it meant paying for the service in the first place which not everyone could afford.

The need for free healthcare was widely recognised, but was impossible to achieve without the support or resources of the State. Throughout the nineteenth century, philanthropists and social reformers working alone had tried to provide free medical care for the poor. One such man was William Marsden, a young surgeon, who in 1828 opened a dispensary for advice and medicines. His grandly named London General Institution for the Gratuitous Cure of Malignant Diseases, a simple four-storey house in one of the poorest parts of the city, was conceived as a hospital to which the only passport should be poverty and disease and where treatment was provided free of charge to any destitute or sick person who asked for it.

The demand for Marsden's free services was overwhelming. By 1844 his dispensary, now called the Royal Free Hospital, was treating 30,000 patients a year. With consultant medical staff giving their services free of charge and money from legacies, donations, subscriptions and fund-raising events, the Royal Free, now re-housed in larger premises, struggled to fulfil Marsden's vision until 1920 when, on the brink of bankruptcy, it was forced to ask patients to pay whatever they could towards their treatment.

As well as the charitable and voluntary hospitals, which tended to deal mainly with serious illnesses, the local authorities of large towns provided municipal hospitals, maternity hospitals, hospitals for infectious diseases like smallpox and tuberculosis, as well as hospitals for the elderly, mentally ill and mentally handicapped. Mentally ill and mentally handicapped people were locked away in large forbidding institutions, not always for their own benefit but to save other people from embarrassment. Conditions were often so bad that many patients became worse, not better. Older people who were no longer able to look after themselves also fared badly. Many ended their lives in the workhouse, a Victorian institution feared by everyone, where paupers did unpaid work in return for food and shelter. Workhouses changed their names to Public Assistance Institutions in 1929, but their character, and the stigma attached to them, remained.

2.2.2 Towards a public health service

The foundations for a public health service could be said to have been laid in the nineteenth century when in 1834 the Poor Law Amendment Act called for the provision of sick wards in parish workhouses. Although intended for the people in the workhouses, the wards soon became full with sick poor people from the parish in general, prompting the State to assess how best this situation could be dealt with. In 1848, the Public Health Act created the General Board of Health, a centralised body intended to review and reform provision for public health, but which, in fact, had little power to do so.

By mid-century, voluntary hospitals, which tended to be more exclusive since they had been created by the wealthy, charities and religious bodies and paid for by donations or subscriptions, began accepting some of the more complicated cases from the workhouses, indicating a further step towards public healthcare provision.
Another milestone during this time was the provision of separate institutional care for smallpox, fevers, insanity and tuberculosis; first in London and later in the provinces.

While the skeleton of public access to health services was beginning to emerge, with workhouses, voluntary hospitals, asylums and isolation hospitals, the level and conditions of care were poor. At the beginning of the twentieth century, preventative healthcare measures became a focus of attention since both the Boer and Crimean wars had highlighted the poor health of soldiers; more had died from fevers and typhoid than through actual warfare.

Any treatment received by wage earners tended to be paid for by their subscriptions to trade unions or friendly societies who, in turn, paid the doctors. This system, however, only covered the worker and not the family. Those who couldn't afford to pay relied on outpatient departments and dispensaries at local voluntary hospitals or simply did not receive treatment. Towards the end of the nineteenth century, voluntary hospitals, unable to provide services on charitable donations alone, began charging for hospital costs.

In 1905, the Minority Report of the Poor Law Commission pointed out the differences in standards of healthcare services provided across the country and urged the Government to make amends. It responded with benefits for the unemployed and pensions for the elderly rather than a direct approach. The National Health Insurance Act in 1911 ensured that workers at the bottom of the wage scale received free treatment with their GP, but did little to improve the situation for the rest of the population.

One of the biggest steps towards organising a National Health Service came in 1920 with the publishing of the Dawson Report. It recommended a comprehensive system, from establishing a single authority to look after all medical and allied services to providing standardised clinical records. The next landmark was made in 1926 when the Royal Commission on National Health Insurance suggested separating the medical service from the insurance system and setting up instead a service, which encompassed all public health activities, paid for by public funds. Then the Second World War brought more change. Although there were no statutory changes, the effects of war brought about significant developments in healthcare provision.

The creation of the Emergency Medical Service gave central Government control over both the voluntary and local authority hospitals as well as taking responsibility for funding. This was the first time healthcare had not been paid for by local authority rates, patient's contributions or voluntary hospitals funds. In 1941 the Government commissioned an independent inquiry to look at the discrepancies in provision of hospital services across the whole country. It concluded that there were vastly differing standards which would remain the case unless a comprehensive overhaul was to take place.

2.2.3 The birth of the NHS

With the voluntary hospitals permanently on the verge of financial collapse and the municipal hospitals almost universally loathed, there was no shortage of pressure for change. The first call for a National Health Service is usually attributed to Beatrice Webb, who argued the case for a state medical service in a submission to the Royal Commission on the Poor Law in 1909. Over the next 30 years the case for reform was taken up and developed in a succession of reports from the Ministry of Health, the British Medical Association and others, culminating in the groundbreaking Beveridge report of 1942.

Sir William Beveridge had been appointed by the Government to chair an inter-departmental committee to look into the existing National Insurance schemes. He made no detailed recommendations about how a National Health Service should be run, but by identifying healthcare as one of the three basic prerequisites for a viable social security system, he laid the foundations for the NHS as we know it today. The Beveridge report was followed by a White Paper, "A National Health Service", published in 1944, which stated: "everybody, irrespective of means, age, sex or occupation shall have equal opportunity to benefit from the best and most up-to-date medical and allied services available". It added, "The services should be comprehensive and free of charge and should promote good health as well as treating sickness and disease".

In 1945 came a second White Paper. The NHS Bill of March 1946 proposed the nationalisation of all the voluntary and municipal hospitals and the creation of 14 regional hospital boards to control them. The National Health Service Act, steered through Parliament by Aneurin Bevan, the then Minister of Health, became law on November 6 1946. It laid the ground rules for the modern NHS. It was to be a comprehensive health service designed to secure improvements in the physical and mental health of the people of England and Wales and the prevention, diagnosis and treatment of illness, funded through general taxation rather than National Insurance contributions.

Many compromises from many parties were still needed before the NHS could come into being on the appointed day - 5 July 1948. The key to the success of the plan was in winning over the various parties involved and it this with which Bevan is credited. The NHS was born.

2.3 Healthcare Reforms

2.3.1 The need for structural reform

The purpose of the NHS is to secure through the resources available the greatest possible improvement in the physical and mental health of the nation by: promoting health, preventing ill-health, diagnosing and treating injury and disease and caring for those with long term illness and disability who require the services of the NHS. The field however in which the NHS is operating is in constant turmoil. The population that it seeks to serve is always changing. Compared with its inception in 1948, there are now far more elderly in the population, greater diversity of ethnic groups, and fewer people living in the inner cities. Most noticeable is that healthcare has a political dimension. Funded largely from national taxes, central Government has an intense interest in its well being, the resources it consumes, and the service it provides. Public accountability is desirable in any public service, but it does not bring stability.

Because of the political and socio-economic dimensions, the NHS has been subject to many structural reforms by various Governments since its birth. There has been a growing need for Governments to control the amount of spending on the NHS and to be seen to be making a firm stance on providing an effieicient and effective service to the nation in light of people's changing expectations through the ages. It is an organisation that constantly must change to keep pace with the demands placed upon it.

Initial estimates for the cost of the NHS in 1948 were greatly underestimated and as such, resources were scarce. The first change to the NHS was made in 1951 when Winston Churchill's Government introduced prescription charges as a way to subsidise the service. Other more radical changes to the NHS have been implemented since, notably in 1974 and 1982. It is the last decade however that has witnessed by far the greatest number of overhauls to the system. Major structural reforms were made throughout the last decade of the twentieth century by the Conservative Government and since by the present Labour Government.

2.3.2 Reforms under the Conservative Government

The Conservative Governments of 1987 and 1992 oversaw far-reaching reforms of the National Health Service, which created much controversy. Supporters claimed the reforms brought increased efficiency and effectiveness, but opponents said they undermined the founding principles of the health service. Under Margaret Thatcher, the Government encouraged people to use private medical services (The Health Service Act 1980 being the first step). However, the public remained committed to the NHS and grew concerned when waiting lists increased and wards closed.

There was also concern about the level of spending on the NHS. With an ageing population and increasing use of expensive new technology, experts said that the NHS needed above inflation increases of at least 1% a year simply to stand still.

The Government, influenced by the Griffiths Report (1983), blamed inefficient management and structures within the NHS for the cash problems. The National Health Service and Community Act of 1990 was the proposed solution. It reformed both management and patient care by introducing the concept of an *internal market* (i.e. the division between purchasers and providers of care - Harrison, 1991). The reforms represented a movement away from the principles of paternalism, collegiality and egalitarianism, and towards the concepts of autonomy, entrepreneurialism and differentiation within healthcare. Such reforms typified the nature of the Conservative Government at this time. The creation of an internal market was designed to:

- Reduce inefficiencies in the UK state hospital system.
- Increase cost-effectiveness of health provision so that a greater "health output" could be achieved with a given budget.

- Improve responsiveness to consumers/patients (healthcare to be demand-led rather than supply-led).
- Introduction of competition on the supply-side to keep healthcare costs under more control.

The reforms concerned the mode of provision and not the source of funding of healthcare. In this respect, the reforms could not be classified as traditional privatisation.

Under the reforms, 'self governing' NHS Hospital Trusts were created but remained in the public sector. Trusts are run by a board of non-executive directors who report directly to the Secretary of State, bypassing the district or regional Health Authorities[1]. Each Trust received a block grant to cover expenditures. Other Trust income was secured through competition for GP contracts. Trusts were permitted to set their own pay levels, manage their assets and specialise in certain forms of treatments through:

- Acute services
- Community services
- Mental health and mental handicap services
- Ambulance services
- Specialist hospitals

A further reform, as part of the internal market system, was the introduction of GP fundholders. GP fundholders are a collection of GP practices and allow individual GPs to control their own budgets and to provide and buy a limited range of healthcare for registered patients using contracts. GPs were allowed to shop around in the internal market for standard and inexpensive treatments for their patients.

The central aim of the reforms was to produce a more cost-effective NHS. As well as the internal market, contracting-out was introduced. This forced the NHS to put in-house services out to tender and award contracts to the lowest bidder. The Private

[1] A Health Authority is an organisation at either a local or regional level designed to care for those patients living within its boundary by ensuring that all the parts of the local NHS work together to plan and deliver health service improvements for local people.

Finance Initiative (PFI) (1992) involved private firms or consortia putting up the capital for major NHS projects. Private firms could pay for the design, construction and operation of buildings and support services. Health unions complained that PFI was privatisation by the back door.

Under John Major, the Conservatives put forward policies which they claimed would make the health service more accountable to patients. They introduced hospital performance tables, the Patient's Charter which aimed to clarify health organisations' duties towards patients (Department of Health, 1992) and A Code of Practice on Openness in the NHS (Department of Health, 1995). They also extended the jurisdiction of Health Service Commissioners (Department of Health, 1996). But, at the same time, the Conservatives were blamed for reducing accountability by allowing hospitals to bypass Health Authorities.

Other structural reforms saw the eight English Regional Health Authorities abolished from April 1996 and replaced by eight regional offices of a new NHS Executive[2], based in Leeds. Likewise, 100 new Health Authorities (HAs) replaced the previous structure of District Health Authorities and Family Health Service Authorities, the aim being to reduce bureaucracy and improve services. With no regional structure in Scotland, Wales and Northern Ireland, responsibility was left with health departments at national and local trust level.

The Community Healthcare Reforms introduced from April 1993 changed the way society cared for the elderly, the mentally ill, the physically disabled and people with learning difficulties. The stated aim was to release people from long-stay institutions and house them in the community where they could be more independent and have a greater say in how they lived and the services they used. But there was criticism from health and social care experts that the changes were not properly funded. They believe care in the community is more expensive than hospital care and that the extra funding has not been forthcoming.

[2] The NHS Executive was formally established in 1989 and is made up of a headquarters and eight regional offices. It is responsible for advising Ministers and for formulating and insuring implementation of policy on healthcare. It has a strategic rather than an operational role.

Despite all the changes, the Conservatives argued that the NHS was "safe in their hands" and that the basic principles of a universal and largely free healthcare service had been preserved. Their critics, however, claimed they had created a two-tier health system, whereby patients of some GP fundholders got faster access to healthcare and where the NHS was starved of the resources it needed, with the result that those who could afford to were encouraged to purchase private healthcare. The following lists highlight some of the perceived positive aspects of the internal market reforms of the nineties, and some of the criticisms levelled at the Conservatives:

Positive aspects of the internal market

- Competition among providers drove down real prices.
- Improved information about costs of health treatments.
- Decline in long term waiting lists.
- Increase in the numbers of patients treated.
- Decrease in average time spent by patients in hospital.
- Fall in public dissatisfaction with the NHS.
- GP fundholders invested heavily in new technology.
- Competition for patients among GPs may have led to improved standards of service for patients.
- Emphasis on outcomes and appropriateness of treatment.

Criticisms of the internal market

- Resource crises - claims of "health rationing" and "priority setting".
- Closure of some hospital wards and banning of some non-essential operations.
- Claims that a two-tier system of treatment emerged.
- Preferential "fast-track" treatment for patients of GP fundholders.
- Hybrid market - competition for routine treatment but little for non-routine, expensive treatments.
- GP fundholders preferred taking on younger, healthier, low cost patients.
- Growth in bureaucracy and organisational turbulence.
- Some Health Authorities merged to give monopoly power.
- Concern about eligibility for care of those with terminal illnesses.
- Concern over quality of psychiatric patient care in the community.

2.3.3 Post 1997 Labour reform

The fundamental changes of the nineties to the structure of the system of healthcare in Britain, together with the rapid speed at which change took place, meant that many people within the health service felt that many processes were irreversible. The Labour Government, which came into power during 1997, had different ideas. They immediately began to address the concerns in the system and launched their programme of changes in "The *new* NHS; a Modern and Dependable Service" (Department of Health, 1997). The main theme of the changes, as outlined in the document, is to replace the competition engendered by the internal market with a new ethos of co-operation. Ministers, like their predecessors in the Conservative administration, were keen that GPs should drive reform in the NHS and decide where resources should go.

Labour have scrapped the controversial GP fundholding scheme under which individual GP practices were given direct control of large sections of their healthcare budget. The Government is to replace fundholding with a system of primary care groups (PCGs) designed to ensure that all patients are treated equally. GPs, together with other health and social services professionals, will join forces in PCGs covering approximately 100,000 patients to decide together how to purchase hospital services.

PCGs will have to work to a three-year *Health Improvement Programme* drawn up by the local health authority to ensure a consistent approach across a locality. A National Institute for Clinical Effectiveness will promote high quality guidelines for treatment based on scientific research, and a Commission for Health Improvement will intervene where local standards are failing.

The Labour Government has also stressed that hospitals, which under the Tories were encouraged to compete for business, must now co-operate to ensure that patients get the best care possible. Many hospital managers and GPs have welcomed the introduction of PCGs and the scrapping of the internal market as way to reduce inequality in the NHS. But they also fear that the new system could be a way for politicians to shirk their responsibility for underfunding in the NHS. With finite resources available, and demand seemingly infinite, managers warn that it is inevitable that some treatments will have to be rationed, particularly as the new system will inherit the debts run up in previous years.

The new NHS will mean new roles and responsibilities for Health Authorities and NHS Trusts and the Department of Health. Figure 2.1 summarises the new financing and accountability arrangements compared with those of the internal market.

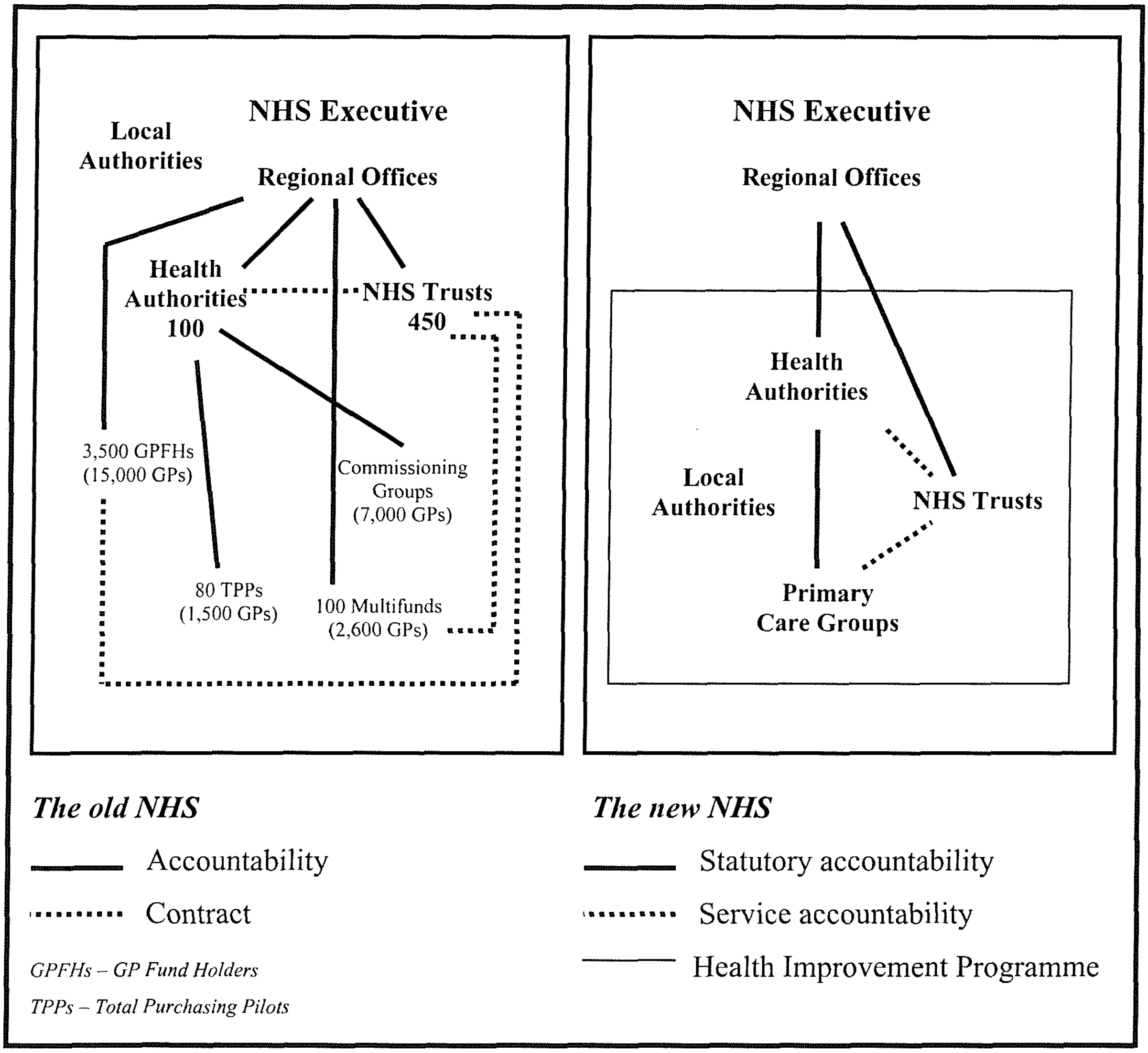


Figure 2.1: Financing and accountability arrangements in the old and new NHS

Under the new Labour proposals, Health Authorities will be leaner bodies with stronger powers to improve the health of their residents and oversee the effectiveness of the NHS locally. Over time, they will relinquish direct commissioning responsibility. Working with local authorities, NHS Trusts and Primary Care Groups, they will take the lead in drawing up three-year Health Improvement Programmes which will provide the framework within which all local NHS bodies will operate.

The Department of Health, and within it, the NHS Executive, will shoulder responsibility for action genuinely needed at a national level. It will work with the clinical professions to develop new *National Service Frameworks.* National Service Frameworks are a series of programmes designed to deliver an evidence-based health service in which change will be driven by improving the performance and efficiency of NHS Trusts.

2.3.4 Future challenges

The Prime Minister, Tony Blair, has set out five challenges that face the NHS as it prepares to make best use of the four-year package of funding as announced in the 2000 Budget:

- **Partnership** - working together across the NHS to ensure the best possible care.
- **Performance** - taking action to review and deliver higher standards in the NHS.
- **Professions** and the wider NHS workforce - getting the right people to deliver the right services for patients. Breaking down traditional barriers between health-care professionals.
- **Patient care** - speed of access, and empowerment
    - ➢ delivering fast and convenient care for patients, and
    - ➢ listening to patients' needs and letting them know their rights.
- **Prevention** - promoting healthy living across all sections of society and tackling variations in care.

The current Government has announced the biggest sustained increase in the amount of money available for the NHS since it was founded. This investment in health is equivalent to an increase over the next four years from £1,800 per household to £2,800 (Creating a 21st Century NHS, Department of Health, 2000). The next step is to agree how to use the money for the greatest benefit. To meet this need, six Modernisation Action Teams have been established. Each team draws together expertise from frontline health professionals, patients and user groups, academics, policy makers, and healthcare managers. A team is led by a Government minister and will come up with suggestions for improving performance and standards across the whole of the NHS. Staff in every NHS and Social Care organisation will be given the opportunity to provide their views. Leaflets giving people the chance to have their say have been distributed to supermarkets and high street pharmacists as well as places where NHS services are available such as GP surgeries, hospitals and dentists.

The Secretary of State for Health, Alan Milburn stated in May 2000: "This is quite simply the most important stage in providing excellent public healthcare since the NHS was set up more than 50 years ago. We are building a new NHS upon the sure foundation of NHS values of high quality public healthcare available to anyone regardless of their wealth or status. The new NHS will be fast and convenient, delivering healthcare that consistently matches the standards of the best in the world. The NHS already delivers high quality care and as different parts of the system work together more and more effectively the quality is becoming more consistent. The unprecedented investment that the Government is making in public healthcare gives the NHS the opportunity to respond by providing the kind of care that people have a right to expect, a modern and dependable health service. This inclusive process will help the NHS to come up with a plan that will re-establish it as the pride of the nation."

Under the Labour Government, the NHS has been set a tough and challenging programme for the future. The Government claims that the programme will involve evolutionary change rather than structural upheaval. The result, it hopes, is a NHS responding to a changed and changing world where patients can expect services quickly and of high quality; an NHS that is accessible and responsive which gets better every year.

2.4 Planning and Management of Resources

2.4.1 The enormity of today's NHS

The scale of the National Health Service is awesome. As Europe's biggest organisation, it has a workforce of around one million people who provide care and treatment for many millions more every year. The NHS spends in excess of £42 billion each year, representing the largest item of central Government expenditure after social security. It occupies about 17,000 hectares of land, an area roughly equivalent to the size of Liverpool (statistics from General and Personal Medical Service Statistics, Department of Health, 1999).

During the financial year 1998/99, the NHS treated a staggering 11,983,983 finished consultant episodes[3] (FCE), 3,420,795 day-cases[4] and delivered 598,805 babies (Hospital Episode Statistics, England 1998-1999, Department of Health, 2000). A more detailed analysis by age of inpatient is shown in Figure 2.2.

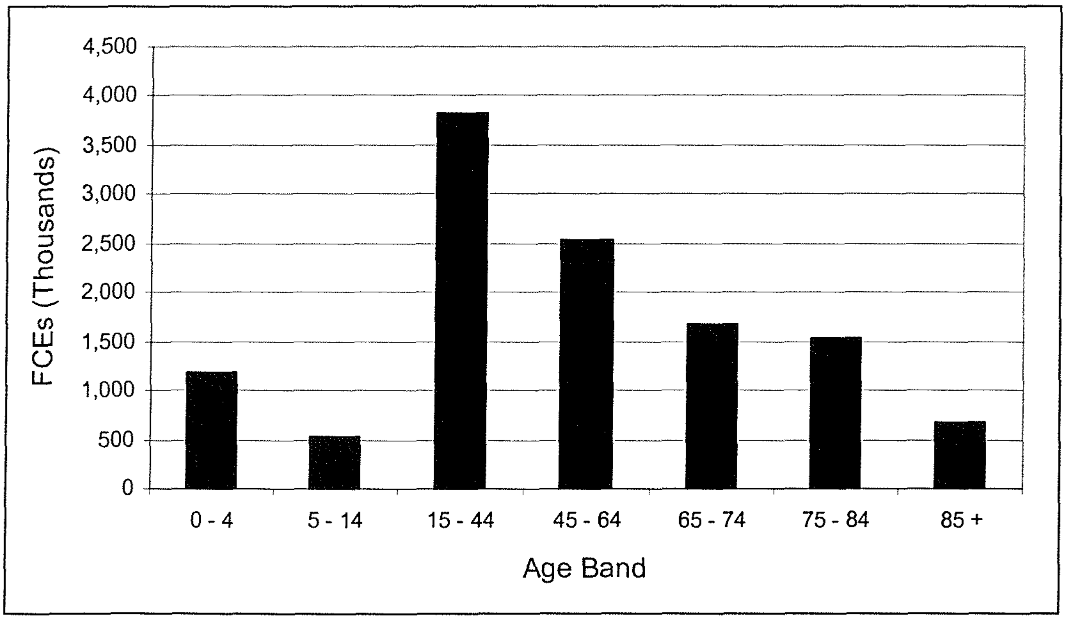


Figure 2.2: Hospital episodes by age of patient

[3] A period of continuous inpatient treatment under the care of a specific consultant. If, during a spell of treatment, a patient is transferred from one consultant to another, a new Consultant Episode commences.

[4] A day case is an elective admission where the patient was treated during the course of a single day. Most day cases are episodes involving minor surgical procedures.

Of the total admissions to hospital, approximately five million were emergency patients and seven million planned patients were admitted from waiting lists.

Table 2.1 presents health service activity over a number of years from 1983 to 1999 (Department of Health, 2000). This table helps to show the general increase in admissions treated by the NHS. In particular, the rise in day-case surgery has been dramatic.

Table 2.1: Health Service activity (1983 to 1999)

Hospital and Community Health Services					
				(Thousands)	% Change
	1983/84	1988/89	1993/94	1998/99	(1983 to 1998)
Ordinary Admissions					
General and Acute	5,113	5,572	6,125	7,595	49
Geriatric	320	412	554	606	89
Maternity (births)	716	650	600	599	- 16
All Specialties					
FCEs (Inpatient)	5,900	6,577	7,984	11,984	103
Day Cases	787	1,005	2,080	3,421	335
New Outpatients	8,311	8,389	9,685	11,100	34
Average Length of Stay (Days)					
General and Acute	11.1	8.8	7.0	5.5	- 50
Geriatrics	58.0	38.5	23.5	22.2	- 62

The NHS is one of the largest employers in the world, employing over one million people, including 350,000 nursing staff and 140,000 administrative and clerical staff. The NHS needs to continually add to its complement of workers to meet the needs of increasing admissions and to keep the service running 24 hours a day, 365 days a year. The total number of hospital medical staff alone has risen from 43,957 in 1987 to 63,548 in 1999 (Department of Health, 2000). Figure 2.3 shows this trend over time.

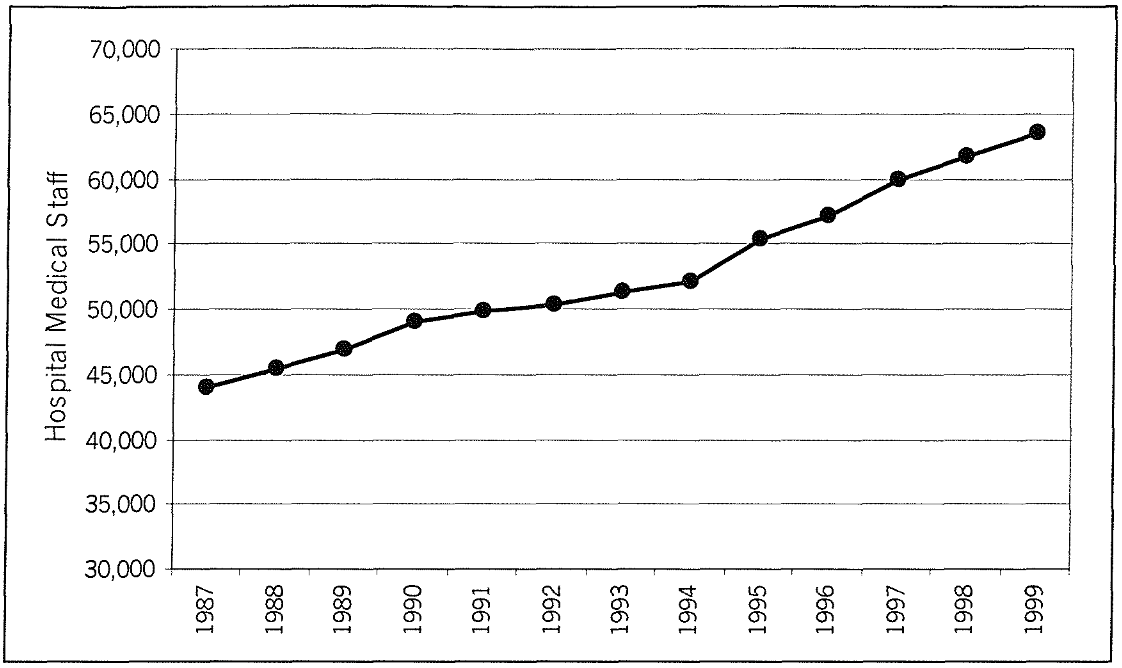


Figure 2.3: Number of hospital medical staff employed in the NHS

Over a quarter of administrative and clerical staff work in direct support to clinicians, so allowing the medical professionals more time to concentrate their skills and experience on direct patient care. General and senior managers account for only 2.6% of the total NHS workforce and 3.6% of total NHS expenditure on salaries and wages. Women make up approximately 80% of the total workforce.

2.4.2 The critical role of the NHS Trust

NHS Trusts continue to provide the focal point of healthcare in Britain. They offer a wide range of hospital and community based services ranging from accident and emergency (A&E), to delivering babies, and to providing care for people with long-term illness or disability. People usually access non-emergency services from NHS Trusts following a referral from their own general practitioner. The care and treatment provided by Trusts remains free to patients. Hospital Trusts are found in most large towns and cities, offering a general range of services to meet most people's needs. Some Trusts also act as regional or national centres of expertise for more specialised care, whilst some are attached to universities and meet teaching commitments. Trusts also provide services in the community, for example through health centres, clinics or in people's homes. Together, NHS Trusts employ the majority of the NHS workforce including nurses, doctors, dentists, pharmacists, midwives, health visitors and staff from the professions allied to medicine (PAMs) such as physiotherapists, radiographers, podiatrists, speech and language therapists, counsellors, occupational therapists and psychologists. Many other staff work to keep the NHS running continuously; receptionists, porters, cleaners, IT specialists, engineers, caterers, domestic and security staff.

Under the new government proposals, NHS Trusts have been given devolved operational responsibility, but are also party to the local Health Improvement Programme. They will agree long-term service agreements with Primary Care Groups. These services agreements will generally be organised around a particular care group (such as children) or disease area (such as heart disease) linked to the new National Service Frameworks. In this way, hospital clinicians will be able to make a more significant contribution to service planning. NHS Trusts will also have new statutory duties of quality and partnership and have to be more accountable to the public by publishing details of their performance and their future strategic plans. They need to demonstrate the development and involvement of staff.

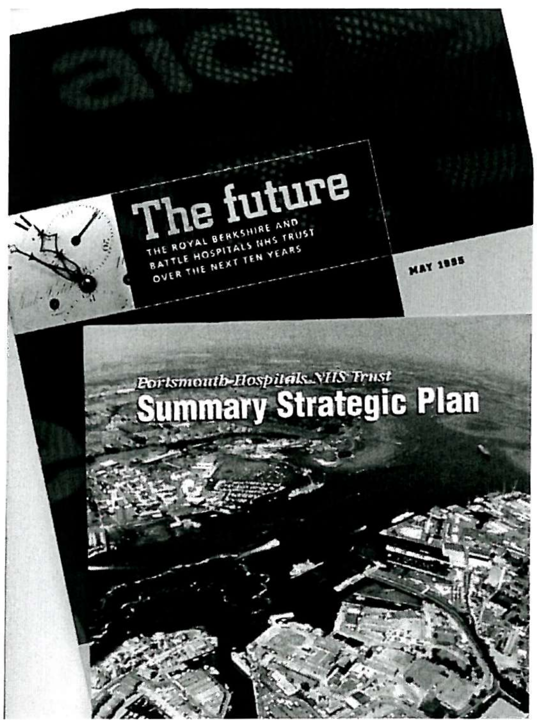


Figure 2.4: Royal Berkshire and Portsmouth NHS Trusts strategic plans

The Conservatives' stated aim in creating NHS Trusts was to raise standards of care for patients through devolving power and responsibility away from health authorities and towards managers and medical staff. By April 1997, all NHS hospitals in Britain (other than in the Scottish Islands) had become Trusts. However, concerns about the accountability of Trusts led the new Labour Government to require all trusts to hold their board meetings in public from June 1997. It is also seeking to ensure Trust boards are more representative of the local community.

2.4.3 The need to plan and manage resources - a system in crisis

The performance of the health system in this country is more important than it has ever been in the past. This is partly because of the massive use of resources needed to meet the growing demands, but also because of the more sophisticated way we now look at healthcare, taking greater note of health outcomes.

We now perceive the NHS more as a service and we assess it accordingly. As a result, the management of the NHS needs to become more sophisticated with a far greater emphasis on value for money and performance.

The Government has a big interest in what goes on in public services. With so many other priorities that need to be addressed, and only a limited bag of resources, the Government needs to be assured that its allocation of money is being used wisely and in line with the polices it has laid out. The 1983 Griffiths Report led to the introduction of general management in the NHS. Since this time, successive Governments have placed more and more emphasis on the need for effective and efficient use of resources consumed within the system. Instead of profit, the NHS's "bottom line" is that it must provide for the healthcare needs of the population. Increasingly within this remit, Hospital Trusts have been given devolved operational responsibility and greater local flexibility in their own planning and management issues. It is the Trust's own responsibility to plan accordingly and manage key hospital resources, such as numbers of inpatient beds, numbers of operating theatres and the size of its workforce.

In recent years, the system has been stretched to its very limits. Problems of provision and use of hospital resources and health outcomes have clearly damaged the public's confidence in the NHS. Critics claim that the new NHS reforms have provided a way for politicians to shirk their responsibility for underfunding in the NHS. With finite resources available, and demand now seemingly infinite, managers must now, more than ever, plan and manage their resources in the best way possible.

The flu outbreaks during the winters of 1999 and 2000 illustrated the fragility of the current system. Although the outbreaks did not officially reached epidemic proportions, the surge in demand, combined with factors including bed shortages and staff problems, caused chaos.

FORTY YEARS OF DECLINE MUST BE REVERSED, REPORT WARNS MINISTERS

Our packed hospitals 'need 4,000 more beds'

ACUTE HOSPITAL BEDS, 1996

THE TIMES

Hospitals don't have beds for new year

By Helen Rumbelow and Emily Davies

INTENSIVE-CARE BEDS AVAILABLE AROUND ENGLAND at midday yesterday

'An increase in community services'

THE INDEPENDENT

NHS inefficiency condemned by watchdog as one hospital in three goes short of beds

Advertiser

WAIT-LIST MISERY

By Advertiser reporter

The Guardian

NHS boost evaporates

Time for a winter emergency package

Figure 2.5: A system in crisis – newspaper headlines from Winter 1999/2000

During the winter of 1999/2000, the impact of the crisis was clearly evident. The state of the NHS made the news headlines on a regular basis (Figure 2.5). The urgent situation was widespread and the repercussions were felt cross the country. Various stories made their way into the newspapers and onto the television. The public confidence in the system was shattered.

The Midlands was badly hit, with two hospitals having to use refrigerated lorries as mortuaries because of lack of space and some doctors working double shifts. One crematorium in Nottingham was forced to use floodlights to fit in more services and another held Saturday services. Death rates over two weeks of the 1999 Christmas break were at 520, compared with 200 for the same period in 1998. In Portsmouth, relatives and friends of hospital patients were asked to wash, shave and feed them to relieve pressure on staff. And in Torbay, hospital managers issued an urgent appeal for people with nursing qualifications to come forward because of staff shortages. In the North, most routine surgery was cancelled.

The NHS Confederation, which represents NHS managers, concluded in a statement to the press, "There are many factors that led to the crises. These include nursing recruitment problems and a ten year rise in emergency admissions, but the main problem is a shortage of beds". In January 1999, the Emergency Bed Service issued a warning about bed shortages for the first time in its two-year history. The warning was repeated a year later. It reported that there were only a handful of additional beds available in England for hospitals that had reached capacity levels. "Bed numbers have been cut around the country as part of efficiency savings by hospitals. The cuts began under the Conservatives. In some cases, hospitals have mothballed entire wards. Many health authorities and hospitals have gone over budget in recent years and each year the problem gets worse as the budget deficits accumulate, leading to more cuts".

Bed shortages have taken the brunt of the blame for the problems facing hospitals over the winter periods. But the debate over bed numbers is a complex one, with some arguing that a modern NHS will not need so many general medical beds because of advances in technology. In the 1960s, there were more than 3,000 hospitals with 550,000 beds between them. By 1995, there were only 250,000 beds and now there are some 194,000, of which 108,000 are for acute patients.

Some health workers argued that the Government's concentration on getting waiting lists down caused the problem. But a survey by the NHS Confederation concluded that hospitals were suspending non-urgent operations to cope with the extra emergency admissions.

Private finance initiatives (PFI) are likely to reduce bed numbers even further, although patient numbers are increasing. The Government has heavily promoted its hospital building programme, which includes an emphasis on PFI. A £2.5bn PFI programme will build 30 new hospitals over the next few years, with the first PFI hospital expected to be finished in the year 2001.

In the aftermath of the winter crisis, the Government published the long-awaited report from National Beds Inquiry; an inquiry charged with looking into the future needs of the health service. It warned that action is vital because hospitals have reached bursting point. The decline in number of acute hospital beds per head of population had put Britain towards the bottom of the league table in Europe. Only four other comparable nations have lower bed numbers per head. It concluded that at least 24,000 extra NHS beds must be provided over the next 20 years, with 4,000 needed almost immediately. "The days when health authorities could count on a continuing decline in the length of stay to allow for the number of beds being cut are over," commented the report's author Clive Smee, chief economic adviser to the Department of Health. He claimed that occupancy levels have reached maximum limits with most hospitals operating at up to 86% capacity. "The safety valve has been reached and occupancy rates cannot go up any further".

Intensive care beds, a critical part of the system, were also badly hit. The turn of the new Millennium heralded the worst-yet crisis in emergency critical care provision, with a severe shortage of intensive care beds and overwhelmed casualty departments. A survey of available beds at midday on 5th January 2000 revealed that there were no intensive care beds for critically ill patients in the South East or West Midlands. In London, doctors were told that the nearest bed was 60 miles away, in Eastbourne. On Christmas Day, doctors were told to send patients recovering from serious operations

200 miles away in an ambulance to Doncaster. An anaesthetist in a North London hospital, who anonymously spoke to the press, said that the crisis was killing patients. "I had someone who died who probably would have lived but for the fact that there was not an intensive care bed available," he said.

2.5 The Need for OR in Healthcare Management

The National Health Service in the United Kingdom has over the last decade undergone major changes in its organisation and delivery, and this experience has to a large degree been mirrored in most major western nations (World Bank, 1993 and Ham, 1992). Increasingly large amounts of resources are being directed to support a service which is strained sometimes to its limit under growing demands. The flu outbreaks of 1999 and 2000 bear witness to a system in crisis. Changes in technology and medical practice generally have led to shifting patterns of care which are often difficult to predict. In addition, demographic shifts (for example, an increasingly ageing population) also impact on healthcare demand (Saltman and von Otter, 1992). In this context there is a growing need to tightly manage healthcare resources. Bed usage in hospitals, for example, is a specific area where capacity management techniques can be implemented more widely to even out monthly variations and other fluctuations in demand (Yates, 1982).

In the UK there is little doubt that the complexity of healthcare management has been compounded by Government led changes. Strategic initiatives, often instigated by central Government, have been pursued by health authorities and need careful management if they are to succeed (Estes and Swan, 1993). Such initiatives have a radical impact on the shape of hospital services. Vetter (1995) for example, predicts a scenario of secondary care very different from current practice. He identifies the following four factors as central to shaping future service:

- Demographic pressures – changes in population patterns and corresponding health needs.
- Technological pressures – changes in the way medicine is practised due to technology.

- Economic pressures – the increasing need for efficiency in the face of growing demand.
- Patient pressures – the ever growing level of expectation present within the population at large.

The management of healthcare at all levels has acknowledged the need to more precisely monitor and control the use of expensive resources (Appleby, 1992). Old tolerances for surplus capacity are increasingly questioned as the trend towards smaller more efficiently run units is pursued. The political element of healthcare emphasises the need for objective methods and tools to inform the debate and provide a better foundation for decision-making.

The current Government has promised further cash injections into the healthcare system to avoid a repetition of the winter pressures of 1999 and 2000. NHS Trusts must now decide wisely how and where to best use the extra money in their planning and management of the services they offer. This task should not be taken lightly nor its complexity underestimated. NHS Trusts need to make fundamental changes to their way of working in order to meet these new Government challenges. They are under considerable pressure to treat ever increasing numbers of patients through a diminishing bed pool. The demands for certain surgical specialties are showing large increases. At the same time numbers of medical emergency admissions are on the rise.

Capacity planning in hospitals is largely a strategic decision. For example the total number of beds in a hospital and the number of beds in various specialities are very major concerns; here the planning horizon could cover a number of years. Management of available capacities is from day to day or over longer periods such as winter months and summer months. An example would be a planned transfer of surgical beds to elderly medical patients in winter. A common current practice is to plan and manage hospital capacities through a simple deterministic spreadsheet calculation approach using average patient flows, average needs, average length-of-stay, average duration of surgical operations etc.

Mathematically speaking, the internal dynamics of a hospital corresponds to a complex non-linear stochastic structure. Hospital managers must deal with:

- Thousands of patients
- Limited resources
- Variable patient-needs
- Uncertain demands

The common deterministic approach for planning and managing the system can be expected to be inadequate in such a system. Typically the deterministic approach will underestimate hospital requirements, the truth of which is witnessed during the regular crises that hit the healthcare system. In the current climate of healthcare provision there is a growing need for operational tools which can support management. The mathematical modelling approach of Operational Research (OR) is ideal for dealing with complexity, uncertainty, variability, constraints and scarce resources and appropriate models can avoid the dangers of planning on the basis of average values only. This research is concerned with the development, solution and validation of sufficiently detailed stochastic models for planning and managing hospital capacities.

The participating NHS Trusts (see sections 1.4.3 and 1.4.4) expressed a great need for evolving operational models to evaluate detailed options within proposed re-structuring processes. Apart from ad-hoc queries regarding particular re-design projects and management of resources, the Trusts were looking to develop a forecasting and planning tool that could be used in mid-term planning and during the annual business planning cycle. It was felt that such a model would greatly aid clinical and specialty managers, allowing them to fully appreciate their local needs and for both at a local level, and more globally at Trust level, to examine a number of "What if?" scenarios.

The developed capacity-planning tools must account for the complex internal dynamics of a hospital. In particular, the Trusts identified the need to model in detail the use of key hospital resources: hospital beds (including critical care beds), hospital operating theatres and the hospital's workforce. Theses three elements are intricately linked together in real-life.

The potential benefit of OR within the public services is considerable. *Modernising Government*, the White Paper published in early 2000, sets out a programme for reforming the way in which Government works. It gives some key aims and commitments, notably:

- Ensuring that policy making is more joined-up and forward-looking.
- Delivering responsive, efficient and high quality public services.
- Using new technology to provide information age Government.

The benefits of an OR approach within a healthcare environment may be summarised under the following three headings (Royston, 2000):

- *Sharpening foresight* – e.g. scenario planning to identify trends and discontinuities, and assess the policy consequences of possible futures.
- *Improving hindsight* – e.g. designing and conducting studies to assess the effectiveness of a policy or programme.
- *Generating insight* – e.g. developing dynamic simulation models to better understand how systems work and the likely effect of changes to them.

The Modernising Government programme of public services, and within this remit that of the NHS, calls for a more integrated approach to issues, working as necessary across institutional boundaries and in partnership with a wide range of stakeholders. This requires thinking and working with whole systems. Such an approach is not just about ensuring that all key elements are recognised. Systems, especially human systems, are more than the sum of their parts: their behaviour emerges from a dynamic interplay between their components. This interplay is the source of much of the counter-intuitive and surprising behaviour of complex systems, such as that of a hospital.

Modernising Government also calls for more forward-looking policy making. Hospitals need to have the appropriate tools. They need tools to let them scan the horizon ahead to spot the early weak signals that may be portents of things to come: to build a coherent and plausible picture of possible futures, and to test and develop policies and programmes that will be robust in an uncertain world. It is evident that the proposed operational modelling work, as described in this thesis, falls within the objectives of the Modernising Government White Paper.

2.6 Chapter Summary

The scale of the NHS is awesome. From its inception in 1948, it has grown to become Europe's largest organisation. It is absurd to assume that the delivery and planning of healthcare will remain static in this country. The political and socio-economic elements of the healthcare system give rise to the need for structural reform. The NHS has witnessed many such reforms since its birth.

Hospitals are increasingly under considerable pressure to treat ever increasing numbers of patients through a diminishing bed pool. The system has witnessed a number of crises in recent years surrounding the use and provision of resources. With devolved operational responsibility and greater local freedom, NHS Trusts must now place greater emphasis on planning and management of key hospital resources. The methods of Operational Research are ideal for this purpose and the potential benefits of modelling work considerable.

Chapter 3 – Operational Modelling for Healthcare

3.1 Chapter Introduction

In the current climate of healthcare provision, there is a growing need for operational tools to support management. This chapter reviews the Operational Research (OR) modelling approaches that have been applied to the planning and management of healthcare resources. Initially a general overview of OR healthcare approaches, by technique, will be reviewed before a more detailed study of developed OR models by hospital application. This is followed by a summary and discussion of the literature review. It will be shown that a number of key healthcare issues have not been fully addressed by previously published work. Specifically, healthcare resource tools should account for complexity, uncertainty, variability and limited resources whilst being designed for utilisation at both the planning and management levels. Deterministic models are too simplistic and stochastic models are required to meet these needs. Computer simulations are particularly useful for solving the complexity of patient-flows through a hospital.

3.2 OR Approaches to Modelling for Healthcare Planning and Management

The perceived need to create and utilise models for a wide range of scenarios in healthcare has spawned a vast array of different approaches. A survey of the literature (as given below) reveals a wide range of OR modelling techniques which have at some stage been used for the planning and management of healthcare resources. The following summary gives example of some of the major modelling approaches adopted in the healthcare domain. The subsequent section explores in more detail different OR models classified according to areas of planning and management application.

3.2.1 Queueing Theory

The principles of Queueing Theory have been available for a relatively long period, and it is therefore not surprising to find its application to healthcare dating back many years. Bailey (1952), for example, uses Queueing Theory to analyse the behaviour of outpatient waiting lists under a range of operational conditions. A more recent study along similar lines (Brahami and Worthington, 1991) uses Queueing Theory to analyse the trade-off between outpatient waiting times and doctor idle times in the management of outpatient lists. Further examples can be found in studies by Kao and Tung (1981) who apply a Queueing Theory approach to bed allocation in healthcare delivery; Weiss and McLain (1987) for acute care facilities; Worthington (1987, 1991) for hospital waiting lists; Young (1962) for the control of hospital inpatient monitoring; and Keller and Laughman (1973) who analyse outpatient queues and congestion.

3.2.2 Markov Models

Markov models have been extensively used in healthcare modelling. Hannan (1984), for example, demonstrates the use of Markov models to examine the strategic allocation of costs in hospitals. Liu *et al.* (1991) use a Markov chain model in medical storage. The main limitation of Markov models (as reported in Davies, 1985 and Shahani *et al.*, 1994) is that the transitional probability of a patient changing status is taken to be independent of previous events, thus the likelihood of a patient remaining in a specific state remains the same from one time unit to the next. Several models have introduced subdivisions within treatment states (Davies *et al.*, 1975 and Farrow *et al.*, 1971). A much better approach is the use of semi-Markov processes (Shahani *et al.*, 1994, Brailsford *et al.*, 1996 and Ridge *et al.*, 1998). A detailed description of semi-Markov processes is given in Appendix B.

3.2.3 System Dynamics

System Dynamics (Forrester 1961, 1968) is currently experiencing a growing interest in healthcare planning and management applications, largely attributable to its ability to deal with operational feedback (see Roberts *et al.*, 1994 and Pidd, 1992). Wolstenholme (1993) develops a revised framework called *Systems Thinking*, which integrates information acquisition, quantitative and qualitative elements, and *archetypes* and *microworlds* as a basis for facilitating discussion and knowledge exchange between the disparate groups involved in the modelling process in healthcare. More recently System Dynamics has benefited from the availability of modern software (Richamond, 1990) and group processes (Lane, 1992), which are both more technically representative and more persuasive to their users. As a consequence, more recent successful applications include Lane (1999) on patient-flows through hospitals and Lane *et al.* (2000) on modelling of patient-flows through an accident and emergency department.

3.2.4 Mathematical Programming

There has been a limited use of Mathematical Programming techniques for healthcare management. One of the major disadvantages of those models published is little recognition of the stochastic nature of healthcare. Typically, models have been built to aid managers as a higher level, more strategic, decision-making tool, such as nurse rostering or regional planning. One early example is Ruth (1981) who developed a Mathematical Programming model to aid regional planning of hospital inpatient services. Other models include nurse requirement planning by Kao and Tung (1981) and Miller *et al.* (1976) for nurse scheduling.

3.2.5 Simulation

The culture of the NHS leads to a focus on individual patients passing through the healthcare system. It is not surprising therefore to find that many models have tended to use a discrete event simulation (DES) approach. Simulation enables the modeller to handle patients with stochastically generated attributes. The simulation approach has been used in a variety of healthcare applications for a number of years, with early models including Goldman *et al.* (1968) on the evaluation of bed allocation. More recent models include Dumas (1984 and 1985) on hospital bed utilisation and planning; Gove *et al.* (1995) on hospital caseload support; and Kalton *et al.* (1997) for operational planning in a multi-disciplinary clinic.

3.2.6 Data Envelopment Analysis (DEA)

DEA (Charnes *et al.*, 1978, 1995) has attracted a growing interest in the OR community and examples of its application in healthcare are beginning to appear with more frequency, especially in relation to cost and performance modelling. Hollingsworth and Parkin (1995) use DEA to measure efficiency in acute hospitals in Scotland; Ozcan and McCue (1996) have adapted a DEA model to provide a financial performance index for hospitals; Kleinsorge and Karney (1992) show how DEA is used in the management of nursing homes; Ozcan *et al.* (1998) examine the relationship between healthcare provider and technical efficiency for stroke patients. A more general review of DEA in the public sector can be found in Ball and Roberts (1998). DEA has been extensively used to measure efficiency of healthcare systems, however the technique does not appear to be well suited to the application of day-to-day management tools where managers need to track individual patients through time. Instead DEA has traditionally been used as a cost and performance tool with an emphasis on comparing the efficiency of one hospital (or healthcare system) to another.

3.2.7 'Soft OR' methods

Soft Systems Methodology (SSM), the most prominent of the 'Soft-OR' methods (Checkland, 1981 and 1984), has been both advocated and adopted as a primary approach in healthcare modelling (Checkland and Scholes, 1990). Lehaney and Paul (1994, 1996) demonstrate the use of SSM in the specification and acceptance of a computer model of outpatient care and conclude that the linkage between SSM and simulation is worthy of further investigation. Other research (Roginski, 1995) discusses the successful use of Strategic Option Development Analysis (SODA), an alternative to SSM within the NHS management context.

3.2.8 Other techniques

In addition to those listed above, many authors have utilised other OR approaches in the study of healthcare systems. In some cases these entail the use of novel methods and in other cases more established methods in combination. One such example of a novel approach is by Eliasz *et al.* (1993) that firstly combines Structured Analysis and Design Techniques (Marca and McGowan, 1988) and then Coloured Petri nets (Eliasz, 1992) as a basis for the development of dynamic simulations in community care. Other techniques include Structured Systems and Design Methodology (SSADM), a long time favoured approach of system analysts (Herbert and Willis, 1992, Ashwood and Woodland, 1990); and Object Orientated Analysis (OOA) and Design (OOD) (Meyer, 1993, Kay, 1993 and Graeber, 1995).

3.3 OR Applications for the Planning and Management of Hospital Resources

Any attempt to classify the disparate studies and application of OR methodologies in healthcare planning and management is inevitably met with problems. A major issue, in this context, is the selection of an appropriate dimension as a basis for classification. For example, one could distinguish between strategic versus operational dimensions of the developed models. Still another is the divide between research projects and commercial applications. Invariably any classification is flawed by the numerous examples of studies that cross category boundaries.

Despite these concerns, a survey of research literature of hospital resource models is provided below. This is classified according to area of hospital application: inpatient beds, operating theatres, workforce planning, critical care, outpatient services, emergency services, waiting list management and hospital ancillary services. It concludes with a discussion of the published work and the future direction of this research in response to the literature study and the needs of the participating hospital NHS Trusts. More general reviews on OR hospital management may be found in Shuman *et al.* (1975), Fries (1976, 1979) and Smith-Daniels (1989).

3.3.1 Hospital bed planning

The National Health Service in the United Kingdom has over the last decade undergone major changes in its organisation and delivery. Central Government strategic initiatives, pursued by health authorities, have had a radical impact on the shape of hospital services. In this context, bed management is a particularly important area of concern. A hospital must accommodate thousands of patients each year, coupled with a complex case-mix, uncertain patients demands and variable patient needs. It is not surprising to find many published papers concerning the planning and management of hospital beds.

Interest in the provision and allocation of hospital beds dates back many years. An early example can be seen in the recommendations of the Commission of Hospital Care (1947), which concluded that the number of beds in a hospital should be calculated as $X + 3(\sqrt{X})$, where X denotes the expected average occupancy of the hospital. No explanation is given for the formula. However it was widely recognised at the time that the underlying assumptions included a Poisson distribution of census (people in beds), random arrivals and a duty to accept all arrivals. This was the first known attempt to mathematically derive the number of hospital beds required. It is clearly far too simplistic to represent the complexity and diversity of real-life.

Goldman *et al.* (1968) appears to be the first paper to utilise computer simulation for modelling hospital resources. The authors employ the technique, although in a primitive way, to evaluate bed allocation policies. This paper, whilst acknowledging the stochastic nature of healthcare, fails to account for differences in patient needs and complex patient inter-arrival times, which are simply generated from the same negative exponential distribution and take no account of the time-dependent nature of arrivals. Once again, this model assumes that hospital departments can accept all arrivals, the concept of which is flawed in practice.

Ruth (1981) applied a mathematical programming approach to develop inpatient services among hospitals within a region. The problem was formulated as a mixed integer programme. The objective function was subject to accessibility (to ensure distribution of beds throughout the region meets demand) and acceptability (to ensure that increases in level or size of services are feasible). The model was used to examine various population distributions across a region and the resulting needs. There is no recognition of the stochastic nature of care provided by each level within each hospital. Its purpose seems only to provide information for strategic management of care at a global level. It lacks robustness at a more individual level of care and provides no information on how to manage or indeed plan beds over time.

The first paper seemingly to recognise and identify an integrated decision making process within hospital bed planning was by Butler *et al.* (1992). It tries to encapsulate the idea of linking bed planning decisions between the different levels of infrastructure within the hospital. The authors build a model which, whilst finding optimal decisions at each level, uses results from previous levels as part of an iterative procedure for finding overall optimal and consistent solutions. For example, the authors state that it is useless to find an optimal policy for providing extra beds for a department when there is no room to accommodate them. Figure 3.1 shows the integrated stages of the model.

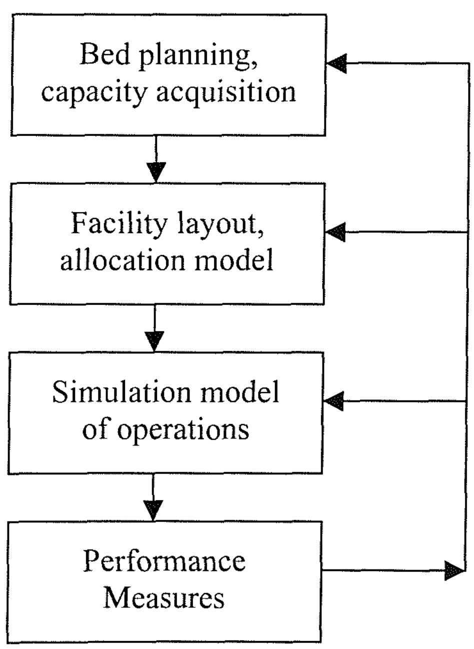


Figure 3.1: Multi-level based hospital bed planning approach (Butler *et al.*, 1992)

Long-term decisions on bed planning and capacity acquisition are incorporated into the planning horizon. Information on the physical constraints of the layout (maximum number of beds by hospital department) is used to generate various scenarios for the distribution of beds across the hospital. There is very little information given on the simulation model itself, raising doubts over the ability of the model to capture the necessary details. As too often found in literature on this topic, the proposed model seemingly fails to incorporate the necessary mathematical and operational details governing the dynamics of individual patient-flows through the hospital system.

Other attempts to produce various hospital inpatient bed allocation tools have been proposed by a variety of authors, including Blewett *et al.* (1972), Shonick and Jackson (1973), Kao and Tung (1981), Cohen *et al.* (1980), Dumas (1984 and 1985), Vassilacopoulos (1985), Dundas and Meechan (1986) and Wright (1987). All of these papers examine how best to allocate beds between various hospital departments. Again, most fail to recognise the many real-life complications inherent within the workings of a hospital. For example, none recognise or incorporate daily arrival patterns, nor do they sample length of stay from appropriate statistical distributions. Indeed many use only a deterministic calculation for length of stay. All of the models are hospital or department specific and correspondingly the methods and results are not readily transferable to a wider hospital bed-planning context. In most cases, there is little or no evidence that hospital managers have actually used any of these proposed academic models themselves.

More recent papers (Gove and Hewitt, 1995, Bagust *et al.*, 1999 and Lane *et* al., 2000) have utilised the advancement of computing power to build more realistic hospital simulation models. Lane *et al* have examined the impact of hour-by-hour demand on an accident and emergency department and resulting inpatient bed needs, with particular emphasis on patient waiting times. There is however a need to study the implications of the stochastic nature of demand on a daily basis as it affects the use of bed stock. Bagust *et al.* go some way to addressing this need by developing a discrete-event stochastic spreadsheet model which generates new arrivals each day as random variations around a long-term trend line. The model is used to find the probability of no beds being available in the hospital on any given day of the year. It is the first such paper to critically examine the relationship between refusal rate (probability of a patient being refused admission because of no available bed) and occupancy rate (indicator of how busy the hospital is). They conclude that when occupancy rates exceed 85%, an acute hospital can expect regular bed shortages, and periodic bed crises if bed occupancy rises to 90% or more. It is commonly known, by hospital managers and consultants, that bed occupancies naturally vary between different hospital departments. Occupancy itself is multi-dimensional, depending on the distribution of emergency and elective (planned) patients, length of stay and case-mix. This paper, although well received and publicised amongst the medical profession, takes no account of these factors. Instead it has treated the hospital as a single entity and has drawn generalised conclusion at the hospital level. Although daily variations have been considered, a mean length of stay for all hospital admissions has been used. In a complex stochastic system, this common deterministic approach can be expected to be inadequate (Shahani, 1981). This is a major oversight of the work.

3.3.2 Operating theatres

An operating theatre is an expensive hospital resource, and therefore hospitals need to maximise their utilisation. This equates to maximising the number of operations that can be fitted in to each theatre whilst avoiding over-run in the theatre schedule. High utilisation brings the benefits of reduced waiting lists for operations (a benefit for the patient) and a reduction in overtime and under-utilised resources (a benefit for the hospital). Many of the published papers on this topic have focussed on issues of scheduling of patients into theatre.

A simple definition of theatre utilisation is the time that the theatre is occupied divided by the time that the theatre is available. Intuitive though this might seem, it is interesting to find various assumed definitions throughout the literature. For example, McQuarrie (1981) uses the time used, excluding overtime and emergencies, divided by the time available. The author suggests a normal utilisation figure for an efficiently run hospital as above 60% but with peaks around 75%. However O'Donnell (1976) indicates that a figure in excess of 50%would require extreme effort, and a level of 65% would be impossible to maintain without causing staff fatigue. Here the utilisation is calculated in a different way. Comparisons of utilisation can only be made with the same method. Clearly the reader should be aware of the different definitions as used by different authors.

Regardless of the method used to measure utilisation, it has been shown that increased utilisation can be obtained through efficient scheduling of operations (Wright *et al.*, 1996 and Hackney *et al.*, 1984). A study by Gordon *et al.* (1998) showed that refinement of scheduling policies, including the use of computer scheduling, could lead to much better utilisation of the available time. Improved scheduling of operations at the hospital led to a reported increase in theatre throughput of 4.5%.

Hanson (1982) and Sier *et al.* (1997) detail some of the constraints involved when scheduling patients. These include the available surgeons, types of operations that need to be done, and how much free theatre time is available. The scheduling process generally involves fitting a number of elective patients into time slots in the available operating theatres. Ideally, the objective is to schedule the patients in such a manner that results in shorter patient waiting times, no overrunning, increased utilisation of the theatres, less wasted time for surgeons, and a more even workload for the staff at the hospital. This is a complex task, complicated further by the need to consider emergency operations, which can severely disrupt schedules, as these need to be inserted into the pre-planned schedule as soon as possible. Gerchak *et al.* (1996) examines such problems in more detail.

Scheduling is often done with paper systems, with a schedule on which the operating rooms and time slots allocated are displayed. Surgeons and anaesthetists play a key role in estimating operation times. In this way, the accuracy of the schedule depends on the skill and knowledge of the staff. Computer scheduling packages are now widely used in hospitals (e.g. *Surgiserver* and *Orbit Surgical Services Management Software*) although these are not always able to improve upon staff estimates (Wright *et al.*, 1996).

According to a survey carried out amongst a selection of operating room directors in the US by Hamilton and Breslawski (1994), the five most important factors in scheduling are:

1. Number of theatres – the more theatres the greater flexibility in scheduling patients.
2. Equipment limitations – if an operation requires specialist equipment, but another theatre is using it, the operation will be delayed for a significant amount of time.
3. Estimated surgery duration – if operation lengths are not estimated accurately, then problems with either idle time or overtime will occur.
4. Hospital scheduling policy – depending on the system used, different effects will be felt in the cancellation rate, the recovery unit utilisation, and the working practices of the doctors.
5. Block time available – if the blocks are all allocated, then surgeons may have problems obtaining a theatre to perform an operation.

Given these concerns, there are surprisingly relatively few published papers concerning numbers and allocation of theatres. Instead, the vast majority of literature in this area focuses on the prediction of operation times and scheduling of individual theatres. Papers that attempt to address some of the above issues include Gibson (1998), Lowery (1992) and Kwak *et al.* (1975) who develop computer simulations to model patient-flows through surgical suite areas.

There have been a number of studies that have investigated the estimation of operation times, often as part of a wider examination of theatre scheduling. Very few however seem to be of practical use. This could be attributed to the large data handling requirements needed to produce estimates, the lack of recorded data, or high variability in the estimates. Many studies, such as that by Magerlein and Martin (1978) involve small sample sizes (212 cases in this example). Any detailed conclusions drawn from these can be expected to lack robustness and reliability. Linear regression techniques are employed by Magerlein and Martin to predict operation times. The paper reports that the developed model accounts for only 44% of the total variation within operation times. Another model for predicting surgical durations (Rose and Davies, 1984) uses the Beta distribution combined with ideas from Critical Path Analysis. This formula was used for a year at Morrison Hospital, Swansea, where it was found to produce statistically significant reductions in the variance of operation lengths. However only urology data was used and it is well known within the medical profession that such operations generally exhibit low levels of variation.

Other OR applications in this area include Carter *et al.* (1992) and Murphy *et al.* (1985) who adopt a computer simulation approach for patient scheduling. Luck *et al.* (1971) and Luckman (1971) present a simulation model of surgical resource use to assist in the scheduling of patients through a unit in the Wessex region. As with most papers on theatre scheduling, little or no account is taken of the stochastic nature of operation times and the complex relationship of the hospital operating theatre to the planning and management of hospital inpatient beds.

3.3.3 Hospital workforce planning

The NHS is one of the largest employers in the world, employing over one million people, including 350,000 nursing staff and 140,000 administrative and clerical staff (Department of Health, 2000). The NHS needs to continually add to its complement of workers to meet the needs of increasing admissions and to keep the service running 24 hours a day, 365 days a year. Managers need to quantify the number and type of workforce required in order to successfully staff a hospital. Shortages in necessary nurses, for example, can lead to the temporary closure of hospital wards. With beds in constant demand, managers can ill-afford to have to take such action.

Wolfe and Young (1965) were early pioneers in developing OR tools to predict future nursing requirements. They divided nursing activities into two categories: *direct-care* (such as bedside care) and *indirect-care* (such as administrative duties). In order to ascertain the number of nurses required, they evolved a classification scheme for rating the severity of illness of a patient. This rating indicated the level of care required. Many subsequent papers have utilised this early work to great effect leading to the concept of patient dependency scores (Buist, 1994). The rating system of Wolf and Young consisted of only three levels of care. No mention is made of how these levels were derived. Indices were subsequently evaluated for each level of care to predict the amount of care per eight-hour shift. These indices were simply estimated mean amount of care required per patient level. The aggregated nursing requirement per shift, I, was then simply found by $I = 0.5N_1 + 1.0N_2 + 2.5N_3$, where N_1, N_2 and N_3 represent the number of patients requiring care levels 1, 2 and 3 respectively. To discover the total amount of nursing time required, including that of indirect-care, a survey by the authors found that for each eight-hour shift, twenty hours of total indirect nursing time was required on the ward. Clearly the proposed model was only valid for the hospital in question. It makes no attempt to address the different types of nurses required but provided the first known application of OR, albeit on a simplistic scale, to quantify hospital workforce needs.

Abernathy *et al.* (1973) proposed a hierarchical structure of hospital manpower planning and scheduling. Linear Programming and Monte-Carlo techniques were employed to solve the different tiers of the problem (see Figure 3.2).

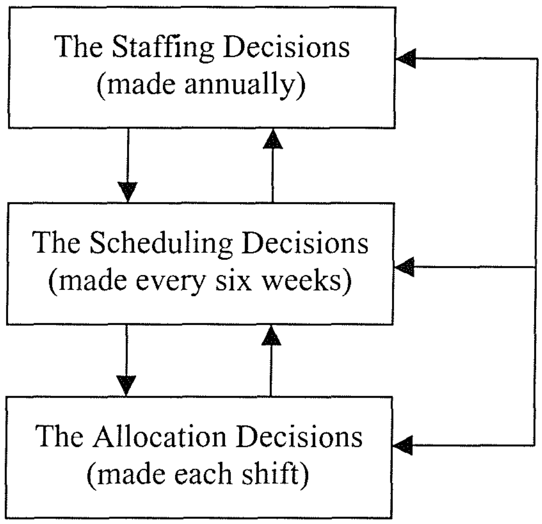


Figure 3.2: Hierarchical structure for nurse scheduling (Abernathy *et al.*, 1973)

The model presented in Abernathy's paper requires high levels of data and computing power to have any benefit (such power was not readily available in 1973). Essentially the developed model was too complex to be understood by the hospital managers themselves and consequently does not appear to have been used in any hospital setting. The mathematical complexity and lack of user-friendliness has proved to be the downfall of a model that was initially designed to be used by hospital workforce planners themselves.

Many of the other published papers in this area suffer from similar shortcomings. Liebman *et al.* (1972), Maier-Rothe and Wolfe (1973), Miller *et al.* (1976), Kao and

Tung (1981), Rossenbloom and Goertzen (1987), and Sifred and Benton (1992, 1994) all fail to acknowledge and incorporate the stochastic nature of workforce planning. Many are hospital or department specific. Few of the proposed models in the literature review can realistically be expected to be of use to hospital managers.

More recent attempts include Dowsland (1998) and Dowsland and Thompson (2000) who schedule nurse rosters with a combination of knapsack, networks and tabu search. Weekly rosters may be produced which the authors report has minimised complaints of unfair treatment by individual nurses. The paper however does not recognise the fundamental aspect of patient demand that will undoubtedly impact on nursing needs across different hospital wards over time. A prerequisite surely must be to consider what the likely nursing needs will be, by grade of nurse over the scheduling period of time, before employing a valuable methodology as described in this paper to schedule the available nurses appropriately.

3.3.4 Critical care units

Critical care concerns the provision within a hospital of both intensive care and high dependency care services. Intensive Care Units (ICU) provide a service for patients who have "potentially recoverable conditions, who can benefit from more detailed observation and invasive treatment than can be provided safely in an ordinary ward or high dependency area" (Intensive Care Society, 1997). It is usually reserved for patients with threatened or established failure of one or more organs, particularly respiratory, cardiovascular or renal systems. Such failure normally arises as a result or complication of an acute illness or trauma (*emergency*), or as a predictable phase in a planned treatment programme (*planned*).

Patients often require technological support including mechanical ventilation and/or invasive monitoring (Intensive Care Society, 1998). Intensive care is consequently very expensive with a cost of between £1,000 and £1,800 per bed per day in the UK (Sachdeva and Guntupalli, 1999). High Dependency Units (HDU) have been introduced as a step between intensive care and ward care. They reflect a need for more suitable levels of care for patients and as a means for reducing some of the costs of an ICU.

The provision of critical care has to meet the challenges of considerable uncertainty and variability in the needs of the patients, high costs and scarce resources. The demand for critical care beds arises from many sources as emergency or planned admissions. The vast majority of the demand for intensive care is experienced as emergencies. Patient's lengths of stay, and the high cost of treating patients, are very variable.

A report on intensive care by the Department of Health (Metcalfe and McPherson 1994) highlighted the uneven spread of intensive care beds between hospitals, and showed that patient refusal rates were strongly linked to local bed allocation. Crosby and Rees' (1994) survey of eight UK acute general hospitals indicated that a significant number of surgical patients need high dependency care but that the number of high dependency beds is insufficient, with the result that many patients are unnecessarily staying on more expensive intensive care units.

There is a great need to better plan and manage critical care beds at both a local and regional level. The majority of published papers on this topic have concentrated on the number of beds within a single intensive care unit. More recent examples include Knighton *et al.* (1994) and Ridge *et al.* (1998). Many of the proposed models fail to reflect the complexity of the critical care unit and are not sufficiently detailed for providing the necessary information for critical care managers. Furthermore, none have examined the relationship of high dependency and intensive care units, and the allocation of beds between them, nor have they researched the allocation of beds or the efficient use of shared capacities amongst a number of regional units.

In April 1999, the Department of Health established a review of adult critical care services. The invited expert group outlined a far-reaching modernisation programme and identified the necessary in-depth work required in many areas, including a need for detailed capacity planning. The summary of findings, as given in the Department of Health's *Comprehensive Critical Care* document states "The current provision of critical care in the UK is characterised by considerable variation in organisation and delivery, quality, funding and effectiveness. This situation is largely the product of historic legacy and ad hoc development. The expert group believes that, while the development of additional beds and services is crucial, the final shape and size can only be determined through evaluation of the impact of the proposed changes, supported by the assessment of need" (Department of Health, 2000 and Day, 2000). Clearly there is a current need for appropriate operational models for capacity planning of critical care units in the UK.

3.3.5 Outpatient services

A literature review of hospital outpatient services reveals that many papers have been written concerning waiting times and appointment systems. Vissers (1979) describes any appointment system as having three characteristics:

1 The number of patients given the same appointment time at the start of the clinic (initial block).
2 The number of patients given the same appointment time during the clinic (blocksize).
3 The number of minutes between two successive appointment times (appointment level).

A simulation model was used to investigate relationship between the idle time of physician and the waiting time of the patient.

A simulation by Fetter and Thompson (1966) examines the effects of varying different factors on the clinic. Factors include appointment level, service time, number of patients not arriving and interruptions to the clinic schedule. Other papers include Arnold (1992) on modelling of an Orthopaedic clinic; Birchall *et al.* (1983), Freeman (1992) and Gamlin (1997) for scheduling of ENT clinics; Aharonson-Daniel, Paul and Hedley (1996) on the management of queues; and Klassen and Rohleder (1996) on scheduling appointments in a dynamic environment.

A Soft Systems modelling approach is described by Lehaney and Paul (1994, 1996), and Lehaney, Clarke and Paul (1999) who discuss the benefit of linking the soft systems methodology and discrete-event simulation approach.

Paul (1995) discusses the use of *CLINSIM*, a simulation package designed for generalised use in outpatient clinic management. Taylor and Kuljis (1998) use this package to model the outpatient services offered at Leeds General Infirmary. Other examples of the simulation modelling include Levy *et al.* (1989), Huebner and Miller (1996) and Williams *et al.* (1967), all of who describe the re-organisation of outpatient clinics.

Taylor and Kuljis (1998) state the importance of clinical staff and hospital administrator involvement in any study of an outpatient system, so that operational modelling "can make complex, chaotic systems comprehensible".

3.3.6 Emergency services

Accident and Emergency (A&E) services represent, to a certain extent, a self-contained unit within the hospital system. This has led to a large number of papers on the topic. A&E models tend to focus on individual patients passing through the system and it is not surprising therefore to find that many models have adopted a simulation approach. Kirtland *et al.* (1995) discuss the issues surrounding the modelling of emergency services.

Examples of the models themselves are numerous, and include Huang *et* al. (1995) who simulate an A&E department at Plymouth NHS Trust; McGuire (1997) uses simulation to reduce patient time in A&E; Clark *et al.* (1994) simulate the location of an emergency helicopter for emergency services in rural areas; Kilmer *et al.* (1997) incorporate simulation and neural networks in modelling an emergency department; Schroyer (1997) redesigns an ambulatory surgery facility; and Muller and Muller (1998) use a simulation model as a decision support system for the dispatch of ambulances in Belgium.

3.3.7 Waiting list management

The increasing focus of central Government on numbers of patients on waiting lists for surgery and length of time spent waiting has raised concerns amongst the profession that the clinical priorities of patients for their surgery is in danger of being forgotten in the search for shorter waiting lists (BMA, 1998).

Recent announcements of additional funding to decrease the numbers on the waiting list is likely to mean that the easier, shorter cases will be brought in for surgery. This will achieve the political wish for fewer people on the list (in the short term) but those remaining are likely to represent those with more complex conditions who are quite likely to have waited longer. There is a need for a waiting list prioritisation scoring system and improved management, at local hospital level, of waiting lists. Some authors have tackled these concerns by adopting various OR techniques. These include Worthington (1991, 1995) on the management of outpatient and inpatient waiting lists; Croft *et al.* (1997) which outlines the *CliniQueue* model – a generic system to support management of referrals to outpatient waiting lists; and Benneyan (1994, 1997) which demonstrates a simulation model in the management and reduction of inpatient waiting lists in hospitals.

New Zealand has achieved considerable success with a sophisticated national scoring system for elective surgery (Handorn and Holmes, 1997, Dennett *et al.*, 1998). The system identifies "which patients are likely to derive substantial health benefit from those services, bearing in mind competing claims on resources". The established cost-effectiveness plays a central part in the system. Other successful scoring systems are currently in place in Canada and Sweden. The United Kingdom, as yet, has no national-level system in place, although a number of waiting list priority systems have been experimented with in various hospitals including Guy's in London (Gudex *et al.*, 1990) and in Birmingham (Health Service Management Centre Birmingham Seminar, October, 1998). A vast array of proposed prioritisation formulae may be viewed in Mullen (1998). The author concludes that most attempts to produce priority-scoring systems are too simplistic and in reality waiting list management requires both a micro and macro-level approach to succeed.

3.3.8 Hospital ancillary services

A number of other studies have focussed on auxiliary services provided in the hospital. Moores (1987) tackles hospital car parking problems at the University Hospital of Wales. Berchtold *et al.* (1994) use simulation to model different configurations of clinical laboratories. Dankbar *et al.* (1992) evaluate automated path lab equipment and Mukhejee (1991) models the management of operations in the hospital pharmacy. Ceric (1990) uses simulation to assist in the planning of a hospital automatic guided vehicle system. Studies have even focussed on operations of a hospital cafeteria (Stout, 1995).

3.4 Discussion of Literature Review and Future Research Needs

The provision of hospital resources, such as beds, operating theatres and nurses, is a matter of considerable public and political concern and has been the subject of widespread debate (Capewell, 1996, Blatchford *et al.,* 1997 and Bagust *et al.*, 1999). For several years hospital managers have been under considerable pressure to reduce hospital capacities and increase patient throughput in the name of operational efficiency. More recently, public disquiet has arisen in cases where patients could not gain ready access to their local hospital or were subjected to extended delays while vacant beds were identified. This has pointed to the need for improved management of resources. Recent flu outbreaks have highlighted a system in crisis (see section 2.4.3).

An appreciation of the dynamics of the hospital system in its ability to respond to fluctuating demand, variable patient needs and complex case-mixes is important in framing policy, determining appropriate levels of resource provision, and establishing realistic performance monitoring criteria. The management of healthcare at all levels has recognised the need to more precisely monitor and control the use of expensive resources. Old tolerances for surplus capacity are increasingly questioned as the trend towards smaller more efficiently run units is pursued. The political element of healthcare emphasises the need for objective methods and tools to inform the debate and provide a better foundation for decision-making.

There is considerable scope for Operational Research methods to be widely used for this purpose. There is a great need for necessarily detailed and realistic operational tools to aid with the planning and management of hospital resources. The literature study has revealed a vast array of papers on these topics, ranging from the allocation of beds between hospital departments, through to the scheduling of patients for theatre.

The dynamics governing a hospital, and the flow of patients through it, means that the necessary models should reflect the complexity, uncertainty, variability and limited resources. Examples of these conditions are listed below. These key issues have not been fully addressed in the existing literature.

- *Complexity*
    - Rules governing patient admissions into hospital e.g. keep some beds free for emergency patients only; elective patients may only be deferred so many times before increasing their priority.
    - Patient-flows through the hospital e.g. when there is no available bed, we try to admit the patient into another suitable, and available, hospital bed albeit on a different ward; intensive care patients may be discharged early to high dependency care if, and only if, various complex criteria are satisfied.
    - Constraints imposed by other hospital services e.g. patients cannot go to theatre if there is no available inpatient bed for them in the first place; operations themselves are subject to theatre space and surgeon's hours.

- *Uncertainty*
    - Demand is likely to be a function of time e.g. elective (planned) patient arrivals can be controlled and are often therefore highly correlated with the month of the year, day of the week, *and* hour of the day; Planned scheduled admissions to hospital though must also account for emergency patients who arrive at random, often in quick succession, and who must be admitted with the minimum of delay.

- *Variability*
  - Patient length of stay (LoS) varies enormously between and within different hospital specialties. For example, Paediatric care length of stay is frequently biased towards shorter LoS, but occasionally a child might stay a very long time, which can cause a disproportionate 'blocking' effect. LoS for Geriatric care, however, can be expected to show very different characteristics from that of Paediatric care. Here LoS is much larger and the blocking effect can be extreme when elderly patients stay for months rather than days and become so called 'bed-blockers'.

- *Limited resources*
  - Self-explanatory – hospitals must treat increasing number of patients through diminishing bed numbers. There is a need to efficiently and effectively plan and manage all hospital resources with particular emphasis on inpatient beds, operating theatres, hospital workforce, and expensive critical care resources.

Developed models should be able to aid with both the planning and management of resources. Appropriate detailed models that can evaluate a variety of scenarios could be powerful tools for good planning and management decisions.

- *Planning tools*
  - Capacity planning in hospitals is largely a strategic decision. For example the total number of beds in a new hospital and the number of beds in various specialities are very major concerns; here the planning horizon could be about 10 years.
  - Planning tools should allow hospital staff to examine in detail the likely capacities required over time, for example the annual hospital business planning cycle each financial year. They should enable the user to identify the likely consequences of changes in numbers, and distribution across the hospital, of beds, theatres and workforce.

- Planning for services across a region e.g. the number and location of outpatient clinics to serve the population's needs; the number and distribution of critical care beds in a health authority.

- *Management tools*
  - Management of available capacities could be from day to day or over longer periods such as winter months and summer months. An example would be a planned transfer of surgical beds to elderly medical patients in winter.
  - Management tools should allow the user to examine resources in detail over smaller time intervals in order to maximise their utilisation and provide a more efficient use of healthcare resources. For example, the consequences of changing daily arrival patterns for elective patients and re-scheduling of nurses.

All of these features point towards a need for sophisticated hospital capacity models. The literature review, as presented in the previous two sections of this chapter, has profiled the vast amount of literature on this topic, which dates back many years and has seemingly involved the spectrum of OR techniques. Given the wealth of work that has already been done in this area, it is both surprising and disappointing that it has not found greater application. The review has unfortunately highlighted many concerns regarding the adopted methodologies and stated assumptions of various proposed models that essentially make them redundant in a real-life setting. Some of the common themes of concern are given below.

- *Over-simplistic* – many models have failed to recognise the complexity of the hospital dynamics, with assumptions proving far too restrictive. This raises concerns over their applicability and the reliability of their conclusions. Examples include: hospital wards accommodating all arrivals and so the concept of a refusal does not exist; arrival rates taken from the same negative-exponential distribution throughout time when in reality demand is known to be highly time-dependent in many specialities (for example, a large surgical emergency demand on Monday mornings; no planned activity over the weekend). With today's improved computing power, there is no longer a need to over simplify processes. The necessary complexity can be readily captured in the development of practical operational models.

- *Deterministic* – a huge number of papers on the topic fail to acknowledge and incorporate the stochastic nature of healthcare. Such models use average conditions, for example average bed occupancy, average arrivals, and average LoS. It is widely known that variability is high within a healthcare environment (for example, LoS for different patient types have very different distribution profiles). Such models can expect to under-estimate true resource needs (Shahani 1981, Shahani *et al.*, 1994 and Davies 1985).

- *Lacking flexibility* – nearly all models have been designed for specific case studies, for example a specifically named hospital ward, an individual operating theatre, or for a particular hospital workforce. These models lack flexibility and transferability of conclusions from one healthcare setting to another. This is particularly unhelpful to the healthcare profession since great effort is seemingly required to model individual elements within and between different hospitals. In practice, high-level hospital processes are generic processes and widely transferable from one hospital to another. There is a need to develop flexible models that allow for the 'fine-tuning' of parameters to reflect local conditions but which can be used by a variety of hospitals without the need for completely overhauling the model.

- *Lacking granularity* – many models suffer from an isolated approach to modelling; namely that the models are built and experimented with for specific studies and conclusions drawn at a localised level. This lack of granularity (being unable to model at various interacting levels of the hospital system) has a great potential to give misleading and unrealistic conclusions. For example, modelling the use of an operating theatre is the subject of many published papers. An operating theatre however is an intricate and delicately balanced part of the hospital system as a whole - it is not necessarily possible to increase throughput in the theatre unless there is an inpatient bed for the patient in the first instance. Many models fail to recognise this obvious interdependency.

- *Lacking user-friendliness* – practical models should be sufficiently accessible for end-users (e.g. healthcare managers) to understand the workings of the model and to use the model in an experimental way themselves. Many developed applications clearly involve the 'black-box' misgivings and are too complex and/or too difficult to use.

3.5 Chapter Summary

The provision of hospital resources is a matter of considerable public and political concern and has been the subject of widespread debate. For several years hospital managers have been under considerable pressure to reduce hospital capacities and increase patient throughput in the name of operational efficiency. There is considerable scope for Operational Research methods to be widely used for this purpose.

The literature study has revealed a vast array of papers on this topic. The dynamics governing a hospital, and the flow of patients through it, means that the necessary models should reflect the complexity, uncertainty, variability and limited resources. There is a great need for necessarily detailed and realistic operational tools to aid with the planning and management of hospital resources. Many of the proposed models have failed in a number of ways to meet this need.

A common current practice is to plan and manage hospital capacities through a simple deterministic spreadsheet approach using average values only. It may be shown mathematically that this deterministic approach can lead to significant bias and inaccurate results, typically under-estimating the true resource needs (Appendix B). It is evident that the necessary models will be stochastic in nature, flexible, integrated, versatile and easy to use by various hospital managers to examine in detail different local and global concerns. Of the reviewed methods, it would seem that Simulation would best handle the key issues and needs that have been identified, and solve the necessary complex models (Appendix C).

The participating NHS Trusts, Royal Berkshire and Battle Hospitals, and Portsmouth Hospitals (see section 1.4), requested the development of appropriately detailed models to evaluate a variety of scenarios for good planning and management decisions. Royal Berkshire and Battle Hospitals Trust wanted to model and evaluate detailed options within the re-engineered hospital designs. Apart from this 'one off' application, the Trust was also looking to develop a forecasting and planning tool, for beds, theatres and workforce, which could be used in long term planning and during the annual business planning cycle. Portsmouth Hospitals NHS Trust required similar models for use in its submission of the PFI outline business case.

Given the needs of the Trusts, the complexity, uncertainty, variability and limited resources (as discussed earlier in this chapter), and in addition the requirement of a planning *and* management tool, there is a need for sufficiently detailed and flexible models for managing capacities. The following chapter proposes a generic framework that specifically addresses these needs.

Chapter 4 – Requirements, Methodology and Generic Framework

4.1 Chapter Introduction

Previous chapters have highlighted the need for necessarily detailed, stochastic, flexible and user-friendly operational models to aid with both the planning and management of hospital resources. This chapter outlines the adopted general methodology with reference to a list of user requirements from the participating hospital NHS Trusts. A generic framework for modelling hospital resources is proposed which forms a central aspect of the developed simulation models discussed in subsequent chapters.

4.2 Forming a Structured Approach

The literature review in the previous chapter has illustrated the immense diversity of possible methodological approaches for the planning and management of healthcare resources. It has highlighted a need for necessarily detailed and flexible models that can be easily and quickly fine-tuned to reflect local conditions by hospital managers themselves. It is therefore important to delineate a coherent research and development approach consistent with the stated objectives of the thesis and the needs of the participating hospitals.

To meet this need, an evolutionary development methodology was adopted. This approach requires a constant dialogue with the end-users (hospital consultants and managers). Models may then be created and enhanced alongside the potential users themselves and this clearly forms a more structured approach than developing models in isolation and then attempting to fit real-life processes and needs into the assumed framework.

Figure 4.1 illustrates the evolutionary approach adopted during the time spent with The Royal Berkshire and Battle Hospitals NHS Trust. For a detailed description of evolutionary system development see Jenkins (1985).

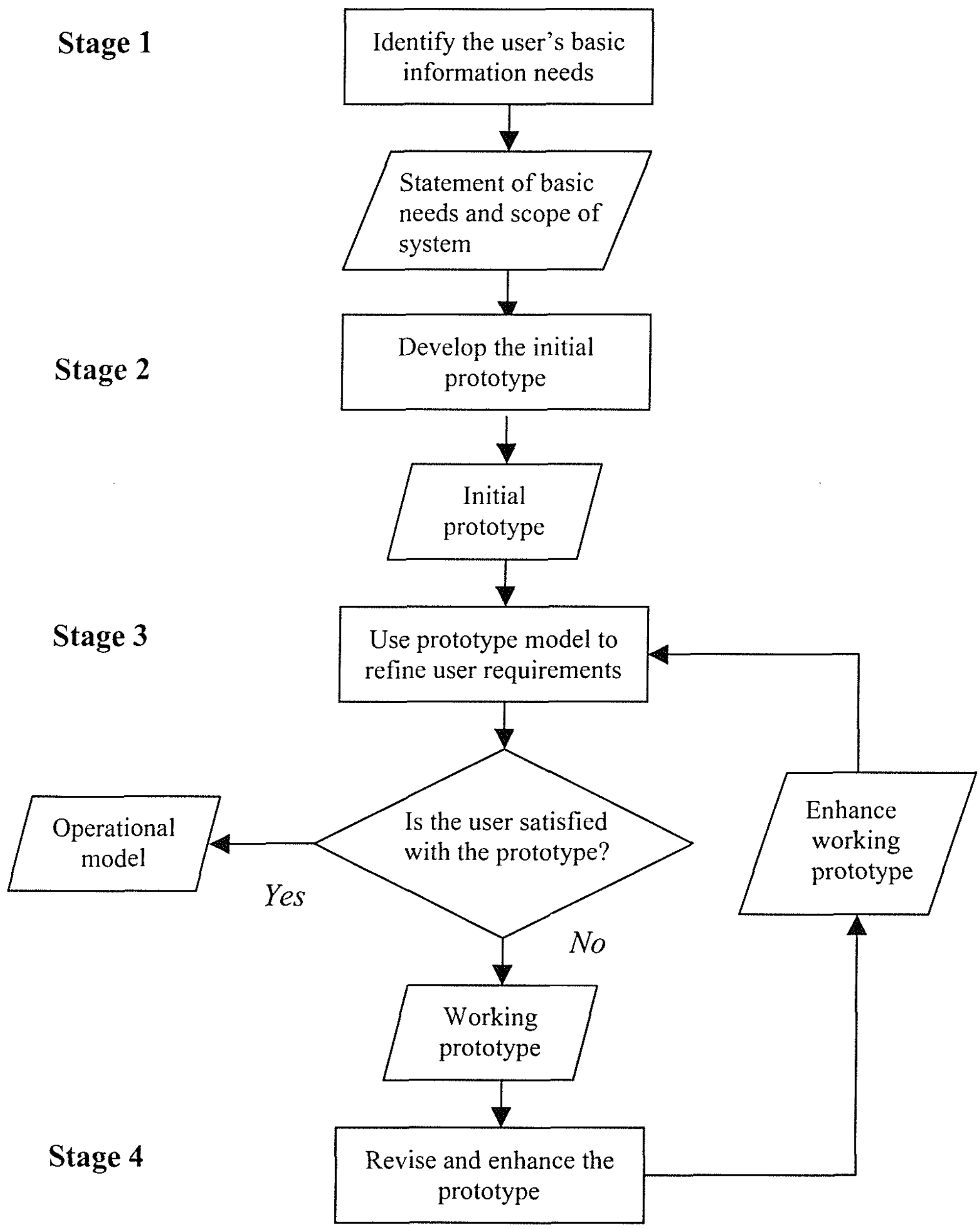


Figure 4.1: Evolutionary model development

The iterative process identified attempts to avoid ultimate failure of the application by involving end-users at all design stages. This is particularly important in a healthcare environment where the operational researcher must gain the trust of hospital staff and be able to demonstrate the benefits of the new technology to replace or enhance existing methods. The evolutionary approach highlights two main components of the research, which are both described in detail in the subsequent sections of this chapter.

- Identifying user requirements to inform the model development (Stage 1).
- A generic framework in which the models are built (Stages 2 to 4).

A considerable amount of time was spent working on-site at the hospitals and liaising with specialty managers, consultants and other hospital staff including bed-managers and nurses. The majority of time was spent sitting on various working groups and within process re-design teams, although contact was sometimes informal and unstructured. These meetings provided a rich source of insight into management processes at different levels and consequently added to the understanding of hospital needs and to designing a practically useful, but mathematically correct, planning and management tool. A range of structured methods were used to elicit feedback when necessary. Such techniques included questionnaires, structured interviews and more soft-system OR methodologies for brainstorming and cognitive mapping activities. A summary of knowledge elicitation is given below:

- *Formal interview* (specialty managers, consultants, nurses and other "system-owners") – structured examination of key aspects of care system often using questionnaires, flow diagrams and demonstration of prototype models.

- *Informal discussions* (all applicable staff) – communication about particular operations of the healthcare system.

- *Observation* – observing processes, for example accompanying the bed-managers on their morning tour of the hospital wards.

- *Examination of data sets, literature and records* – survey of collected and reported data relating to specific healthcare services.

4.3 Defining User Requirement Profiles

4.3.1 General requirements

Over a period of time the following set of criteria addressing user requirements was identified. These user requirements form the basis for the design rationale that enabled the development and enhancement of prototype models. The reader will note that many of the criteria have previously been highlighted as those qualities lacking from previous approaches, as detailed in the literature study in the previous chapter. This is no coincidence and further stresses the need for an improved methodological approach to the planning and management of healthcare resources.

- **Flexibility and versatility** – ability to model a number of different scenarios within hospitals at various levels (from ward level to hospital as a whole) with minimal effort. The model must allow the fine-tuning of parameters to reflect local conditions but without the need for many different models (a flexible model with modules as opposed to a suite of different models).

- **Ease of use** – sufficiently accessible for end-users to understand and readily use the model to examine various *what-if* scenarios.

- **Integration and granularity** – a hospital wide approach to modelling of resources. These must be the ability to model various inter-dependent hospital resources such as beds, theatres and nursing needs within one granular model at different levels as necessary. There is a need to model hospital bed requirements alone or in combination with theatres and/or hospital workforce planning.

- **Validity** – confidence can only be gained once the model has been validated against past data and experiences, and scenarios make clinical and managerial sense.

4.3.2 Hospital resource requirements

After informal and formal discussions with various hospital decision-makers, such as specialty clinical managers, and the responses to a questionnaire (see Appendix E), the following set of complicating factors were evolved. Necessarily detailed hospital capacity models should account for such features and be able to provide appropriate information for scenario modelling and decision-making.

- Emergency patients arrive at random and must be admitted with minimum delay and with priority over elective patients.
- Patient demand may depend on time of the day, day of the week and month of the year. Thus bed requirements should be considered over time and not assumed to be static.
- Patient LoS is highly variable, incorporating short LoS patients and those who stay a very long time.
- Different patient types have different LoS distributions.
- Patients cannot be admitted into a bed-pool if no beds are currently available. However, bed managers will decide whether another suitable bed is free elsewhere in the hospital. In effect, each patient has a *priority list* of suitable hospital bed-pools that is used before either deferring (elective) or transferring (emergency) patients.
- Different specialities have different rules for elective deferral times, ranging from re-admission the next day to some weeks in the future. Likewise, for some patients, only a pre-defined number of deferrals may be permitted before they are given emergency status.
- There are different theatre sessions for different specialties within the hospital. Furthermore, the number and duration of sessions may change by day of the week.

- Some of the patients admitted to inpatient beds require an operation and queue for theatre (*procedure patients*) and some patients stay in the bed but do not require theatre (*non-procedure patients*).

- Patient operation time is highly variable. Different patient types have different operation time distributions.

- There are known hospital rules concerning the amount of time that a theatre session may over-run. These rules have a direct bearing on whether to admit a patient to theatre or whether to cancel the operation until the next available session.

- The hospital day is divided into three shifts – early, late and night. Patients often require different amounts of care over time. In turn this equates to the requirement of different grades of nurse by shift and by day for each hospital bed-pool or ward.

- Nurse-dependency varies between different patient groups within the hospital. Nursing requirements are usually directly linked to the type of operation or clinical diagnosis.

Furthermore, managers provided a provisional list of possible model applications. The following set of criteria specifies the "*what-if*" scenarios to be evaluated. A valid operational model should allow decision-makers to evaluate (quantify) the consequences of various planning and management decisions.

- Examine what the likely bed requirements for speciality bed-pools will be over time and their relationship with other hospital speciality bed-pools.

- Analyse what effects changing various admission rules may have on the efficiency of the bed-pool.

- Understand the consequences of various daily elective (planned) patient admission schedules, in the light of daily bed provisions and a stochastic emergency demand.

- Evaluate various planning and management options, such as the effects of temporary transfer of beds from a surgical ward to a medical ward, creating new clinical groupings, and combining existing bed-pools. Important considerations in this evaluation include refusals (transferred or deferred patients), bed-pool occupancy levels and the interaction between various care-units.

- Appraise different operating theatre scheduling options. For example, whether to operate on a first come first served basis (FCFS), longest operations first (LF) or shortest operations first (SF). Furthermore to evaluate the impact of separate dedicated day-case theatres or whether to incorporate day-cases into existing elective theatre sessions.

- Calculate likely nurse needs over time for each specialty bed-pool or hospital ward. This should be by grade of nurse, by month, day of week and shift each day. This must account for patient case-mix and resulting patient nursing dependencies over time.

The desired user requirements and features of the hospital system point towards a need for a sophisticated hospital resource capacity-planning tool. This should meet the general criteria as specified by hospital decision-makers and potential end-users, as discussed in the previous section.

4.3.3 Critical care requirements

Although a critical care unit (CCU) forms a part of the overall hospital system, to an extent this unit exhibits distinctive planning and management challenges that are rarely seen elsewhere in the hospital, thus the need for separately defined user requirements. The extreme costs of critical care (section 3.3.4), the relatively few beds available and the critical medical condition of the patients admitted intensify the planning and management issues. There is currently a great need to better plan and manage critical care beds at both a local and regional level (Department of Health, 2000). During lengthy discussions with participating CCU personnel, the following set of characteristics concerning the flow of patients through a CCU were evolved:

- Some CCUs comprise of an intensive care unit (ICU) only whereas others have both ICU and high dependency units (HDU). Currently in the UK the average number of ICU beds is 6 but the range is at least 2 to 22. One third of Trusts did not have high dependency beds in 1999 (Department of Health, 2000).

- In those hospitals where both intensive and high dependency beds are available, some have physically separate units and others have a combined unit. The Department of Health critical care review recommended that that existing division into high dependency and intensive care based on beds be replaced by a classification that focuses on the level of care that individual patients need, regardless of location.

- For separate units, intensive care patients may only be admitted to ICU and not to HDU. High dependency patients may be admitted to ICU if no bed is available in HDU.

- For a combined CCU, admission is limited by both numbers of beds and numbers of available nurses. Typically one nurse can care for either one ICU patient or two HDU patients.

- Some units have *holding beds* where emergency patients can wait for a bed to become available. Typically patients can only remain on this bed for a maximum of 24 hours.

- Patients may be discharged early from a bed if certain criteria are satisfied. These usually consist of the patient having stayed a minimum time on the unit, they are on their last day of stay and they will survive. ICU patients may be discharged early to HDU although some may return direct to ward care. HDU patients are early discharged to ward although some patients may deteriorate and require intensive care.

- If no bed is available for admission, and having already checked for possible early discharges and the arriving patient having already stayed the maximum allowed time on any available holding bed, emergency patients will then be transferred to another CCU.

- Elective patients are deferred for a certain time period, if, upon arrival at the CCU, the number of free beds has fallen below a minimum level. This level accounts for the provision of *emergency-only beds*.

- Elective patients can only be deferred for a maximum number of times after which a elective patient is deemed to have emergency patient status. Emergency status elective patients and 'true emergency' patients are always admitted if there is a free bed.

- Emergency demand may depend on day of the week and month of the year. Thus bed requirements should be considered over time and not assumed to be static.

- Patient LoS is highly variable. Typically CCU LoS is biased towards shorter LoS although occasionally some patients may stay a very long time.

- HDU case-mix and LoS is very different in nature to ICU. HDU usually accommodates more elective, less severe patients whereas the majority of ICU admissions are emergencies.

- Different patient types have different LoS distributions.

- Some regional CCUs provide specialist beds, such as those for patients with major head injuries. Other CCUs within the same region will send these patients to the specialist unit.

- Within a region of co-operating CCUs, there often exists a priority matrix governing the movement of transferred patients from one unit to another available unit.

There was an evident desire amongst the participants to permit any developed models to examine the consequences of the following options on unit occupancy, transfer and deferral rates:

- Changes to the number of ICU and/or HDU beds.

- Changes to the number of available nurses in a combined CCU.

- Changes to the number of holding beds for ICU, HDU or a combined CCU.
- Changes to the number of emergency-only beds.
- Changes to the rules governing early discharge.
- Changes to the distribution of beds within a region of co-operating CCUs.
- Changes to the sharing of capacities amongst CCUs, such as emergency-only beds in some, but not all, units.

The complex characteristics of a CCU and region of CCUs, coupled with desired user requirements, indicates a need for a sophisticated CCU capacity planning and management tool. There is a likely need for a model for a single CCU and a further model for co-operating CCUs within a region. Developed models should meet the general criteria as specified by critical care service decision-makers and potential end-users.

4.3.4 Proposed model utilisation

The participating hospital Trusts requested models to evaluate detailed options within the new hospital re-engineering and re-design processes. It is anticipated that there will be a model for hospital inpatient resources (beds, theatres and workforce) and models for critical care services. It was envisaged that all models could be readily used for ad hoc studies within this framework. However the Trusts were also looking to develop a forecasting and planning tool that could be used in long term planning and during the annual business planning cycle. The multi-dimensional nature of the model utilisation requires the development of flexible models that can be fine-tuned to reflect local conditions. In turn this leads to a necessary generic framework in which various project-based studies may be realised. Such a generic framework is discussed in the following section.

4.4 Developing a Generic Framework

The defined user requirements coupled with the increased knowledge of the hospital processes and perceived model utilisation, led towards the evolution of a generic framework. Detailed flexible stochastic models may be developed within this framework. The generic framework is shown in Figure 4.2.

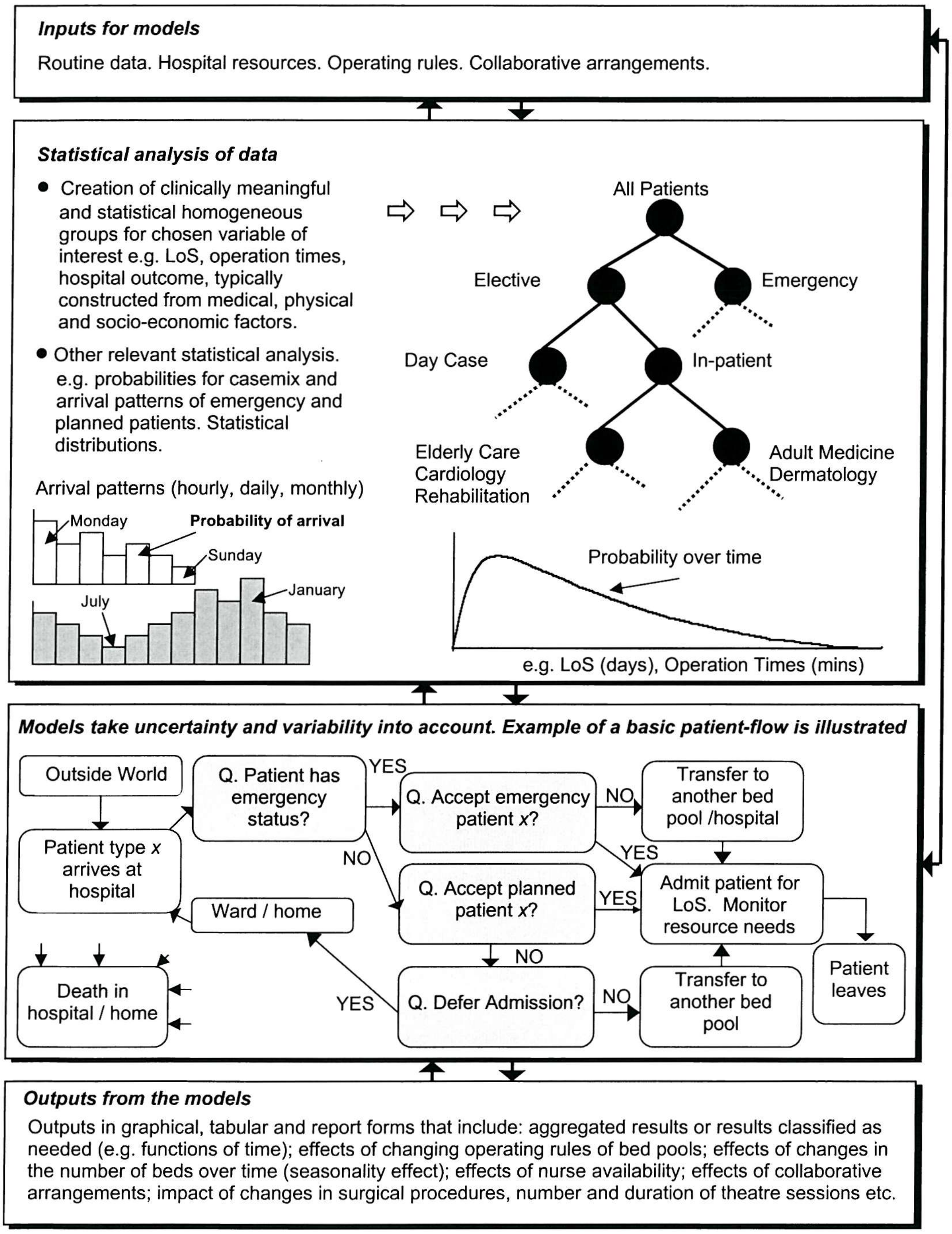


Figure 4.2: Developed generic framework for modelling healthcare resources

Further detail on each stage within this generic framework is provided below:

- ***Model inputs*** – routinely collected data, such as hospital admission and discharge dates, time of arrival, LoS, emergency or elective status, operation time etc. Expert data may be used if raw data is not available (see section 6.4). Hospital rules governing use of resources are defined.

- ***Statistical analysis of data*** – this stage plays a critical role in the success of developed models. The automated rapid classification of patient groups forms a key differentiator between this approach and other attempts to produce practical capacity planning and management tools. A specially designed statistical analysis program, called *Apollo*, has been developed to enable the creation of statistically and clinically meaningful patient groups and to obtain information about particular flows over time. Apollo can link with most databases that are used in hospitals and extract the necessary data for the statistical analysis. Some of the helpful features of Apollo are:

    - Rapid classification of patients using a binary splitting method similar to Classification and Regression Tree (CART) analysis.
    - Analysis of patient arrival times for the detection of patterns in the arrivals over time for any desired group of patients.
    - Fitting appropriate distributions. For bed capacities, distributions for LoS are important. For theatre capacities, distributions for operation times are necessary.
    - Linking the statistical analysis to the simulation model.

Apollo has been designed to link to the hospital simulation model, so that patient groups can be created and the key information on arrival profiles and LoS fed directly into the model. This is a major help to end-users in creating a clinically and statistically correct model. Fitted distributions avoid the need for average values (section 3.5.1). Patient arrival profiles acknowledge demand as a function of time. See chapter five for a description of Apollo and patient classification techniques.

- ***Simulation model*** –developed simulation models within this framework take individual patients through time as they pass through the chosen healthcare system (Figure 4.3). This could be inpatient beds and operating theatres (for hospital capacity model) or complex care-pathways in a critical care unit. Once again this approach differentiates models described in this thesis from other more simplistic models of aggregated patient flows. Patients staying within the healthcare system take their attributes (arrival profiles, LoS, operation times, survival rates etc.) from those described by Apollo during the previous stage. Within this framework, any proposed system can be modelled. Such models capture uncertainty, complexity and variability (section 3.4). Successful application requires a detailed understanding of patient-flows, as typically captured in a flow diagram. This is achieved through the evolutionary development process adopted within this work that involved end-user involvement during all model development stages (section 4.2).

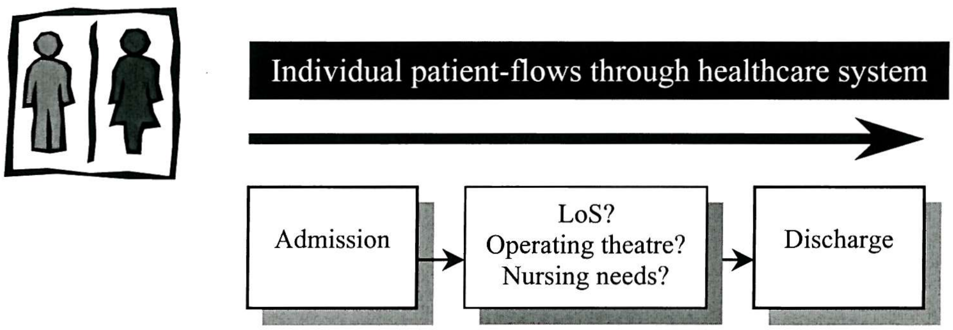


Figure 4.3: Individual patient-flows

- ***Model outputs*** – The simulation model should be designed to meet user requirements (section 4.3) so that model outputs provide the necessary information for end-users. Outputs may be in graphical, tabular and report forms that include aggregated results or results classified as needed (e.g. functions of time).

The framework was used to design and build a hospital capacity simulation model (Chapters 6 and 7) and critical care models (Chapter 8).

4.5 Chapter Summary

It is important to delineate a coherent research and development approach consistent with the stated objectives of the thesis and the needs of the participating hospitals. To meet this need, an evolutionary development methodology was adopted with involvement of hospital staff from the outset of the research. A range of structured methods were employed to obtain information and feedback whilst visiting a participating NHS Trust. Lists of complicating factors governing the flow of patients through hospitals were identified. Necessarily detailed hospital capacity models need to account for these features and be able to provide appropriate information for scenario modelling and decision-making. Provisional lists of model applications aided model enhancement alongside the potential end-users.

A generic framework has been evolved in the light of perceived user-needs and real-life hospital processes. Developed models for hospital resources and critical care units should be designed within this framework. A specially designed statistical analysis program has been developed to enable the creation of statistically and clinically meaningful patient groups and to obtain information about particular flows over time. This automated rapid classification of patient groups forms a key differentiator between this approach and other attempts to produce practical capacity planning and management tools. Developed simulation models within this framework take individual patients through time as they pass through the chosen healthcare system. These models take uncertainty, variability and complexity into account properly.

The generic framework has been shown to be applicable to many different NHS Trusts. Originally it was evolved at The Royal Berkshire and Battle Hospitals NHS Trust but was then adopted by Portsmouth Hospitals NHS Trust who recognised that the structure applied to them. Other Trusts have since approached the author expressing an interest in the framework and developed models.

Chapter 5 – Patient Classification Techniques

5.1 Chapter Introduction

The evolved generic framework (Chapter 4) incorporates the need for sophisticated patient classification techniques to be adopted. Necessary patient groupings may then be fed into developed simulation models and individual patients from each group passed through the particular healthcare system of concern. In order to capture the uncertainty and variability amongst the patient population, a number of classification techniques are considered and evaluated for their relative performances and practical usefulness. A statistical package incorporating a symbolic tree-based algorithm, CART, has been developed for use within the modelling process.

5.2 The General Classification Problem

As a direct consequence of individuality, patients typically differ in a number of medical, physical and socio-economic characteristics, for example by age, severity of illness, complications and speed of recovery. Groups of patients with healthcare needs, whether they represent those who suffer from a particular disease or those who are rushed into hospital as emergencies, are usually considered as groups of patients with similar needs. In fact these groups are typically heterogeneous and require more detailed modelling for classification. Resulting healthcare needs and corresponding resources vary from patient to patient within each group. The subsequent uncertainty and variability in the healthcare system can place great stress on the efficient and effective planning and management of resources.

From both a clinical and operational perspective, it is desirable to be able to divide this heterogeneous group into smaller homogeneous (in terms of some measure) sub-groups. Homogeneity brings the benefits of increased certainty in individual patient needs and resource utilisation. For example, given an individual patient we can classify them into a patient sub-group in which we know, from past experience and data, that their LoS in hospital is likely to be within a certain range of time with a given confidence. The LoS for this patient group will typically substantially differ from the predicted LoS of other groups. The purpose of classification in this example would be to produce tight LoS bands with high confidence. Thus with the added knowledge and confidence of how long individual patients are likely to stay in the hospital system, the potential for improved efficiency and effectiveness in hospital planning and management is vast.

An important criterion for a good classification procedure is that it not only produces accurate classifiers (within the limits of the data) but that it also provides insight and understanding into the predictive structure of the data (Breiman *et al.*, 1984). For example, finding which socio-economic and medical characteristics contribute to the risk of a particular disease not only provides valuable assistance in classifying individuals into risk groups with some certainty, but more generally has advanced the knowledge and understanding of the disease.

There are two elements to a general classification problem. Measurements are made on some case or object (for example a patient in a hospital setting with measurements including age, sex, clinical diagnosis, LoS, outcome etc.) and based on these measurements a prediction is made as to which class a case is in. The prediction is made following a pre-defined classification rule.

In mathematical terms we define X to be the measurement space containing $\underline{x} = (x_1, x_2, \ldots x_m)$, the measurement vector, where each x_i is a measurement taken on a case. The method should, given any $\underline{x}$ in X, have a classification rule to assign one of the classes $(1,2,3,\ldots J)$ to $\underline{x}$, where J is the number of classes. The classifiers are based on past experience using a combination of expert knowledge and past data with their relevant outcomes. For example, the classifiers to be defined could come from a hospital database combined with the expert knowledge of the consultants, specialty managers and other medical staff.

Each measured variable is *continuous*, *nominal* or *ordinal* in nature. A variable is continuous if the measured value is a real number (e.g. LoS, age, height). A variable is nominal if it is a finite categorical set with no natural ordering (e.g. sex, hospital ward, clinical diagnosis). A variable is ordinal if it is a finite categorical set with a natural order.

5.3 Comparison of Classification Algorithms

There exist many different classification algorithms, but their relative merits and practical usefulness for healthcare problems in particular remain unclear. Thus a need arises to evaluate their relative performances. Intrasubject comparisons have been considered in the past, for example, within statistics (Remme *et al.*, 1980), within symbolic learning (Clark and Boswell, 1991) and within neural networks (Xu *et al.*, 1991).

Other authors, for example King *et al.* (1995), have compared different algorithms for non-healthcare datasets, but little or no research has been conducted on the relative merits of various techniques for healthcare problems and in particular consideration of the practical usefulness for (use by and interpretation of) medical personnel.

5.3.1 Algorithms and datasets

Four techniques have been considered, with the intention of representing the spectrum of classical statistical and more recent advances in computer-based approaches:

- Discriminant Analysis
- Regression Models (Multiple and Logistic)
- Tree-based Algorithms (CART)
- Artificial Neural Networks

In considering suitable datasets, the primary criterion was that chosen datasets were of real-world interest to the medical profession. In order to evaluate how the different classification approaches perform on different types of data, a number of datasets were chosen with the intention of representing those with different sizes (number of records), number of variables (fields), the level of variance or deviance (an indicator of how "messy" the data is) and the ratio of continuous to categorical variables in the dataset.

The four selected datasets are described in Table 5.1. A summary of the classification studies is shown in Table 5.2.

Table 5.1: Dataset descriptions

Dataset	Description
A	An intensive care dataset containing routinely collected ICU data within the UK. Information on a number of socio-economic and medical variables. Primary interest in predicting LoS and outcome.
B	Routinely collected data from a hospital patient management system for predicting LoS on the ward.
C	A comprehensive maternity dataset collected as part of a commissioned study on predicting complications at birth. Contains a number of socio-economic and medical variables.
D	A large diabetes dataset, containing information on various diabetic complications collected for over 30 years from a leading unit in the UK.

Table 5.2: Summary of classification studies

Study	Dataset	Dependent Variable		Number of Records	Variance/ Deviance	Number of variables
		Description	Nature			
1	A	LoS in ICU (days)	Cts	582	11.8 (mean 2.4)	7 (2 Cts, 5 Cat)
2	A	Outcome (death or survival)	Cat	582	0.13 (13%)	7 (2 Cts, 5 Cat)
3	B	LoS in hospital (days)	Cts	17,974	56.6 (mean 4.3)	5 (2 Cts, 3 Cat)
4	C	Chance of complicated delivery	Cat	2,402	0.24 (24%)	16 (8 Cts, 8 Cat)
5	D	Predicting onset of Retinopathy	Cat	4,056	0.33 (33%)	14 (12 Cts, 2 Cat)

(Cts = continuous variable; Cat = categorical variable)

5.3.2 Evaluation criteria

The algorithms are evaluated using four criteria. Two are objectively measurable: the accuracy and the computing time taken to produce results. There are also two subjective criteria: the comprehensibility of the results and the ease of use of the algorithm to relatively naive medical users.

Accuracy

There is no generally accepted measure or agreement on the appropriate loss function or accuracy of a classification tool. Therefore to compare the relative performances of each of the four techniques, accuracy has been measured as the percentage of cases (patients) that the algorithm classifies correctly with a categorical dependent variable or the correlation coefficient (r) between the observed and predicted responses with a continuous dependent variable.

Run-Time Speed

Training and testing times were calculated. Training time is the time taken to learn plus the time taken to classify the training data. Test time is the time it takes to classify the test data. All algorithms were performed on the same PC with a Pentium III processor of speed 600MHz with 128 Mb RAM.

Comprehensibility and Ease of Use

This is a measure of the extent to which the algorithm produces comprehensible results that are easy to interpret and understand, particularly by medical staff. This was measured by consulting various medical personnel within the participating NHS Trusts. They were also asked to estimate the ease of use of each technique. This is based on the amount of time required to understand the algorithm and prepare the data, the amount of tuning necessary and the time required to produce correct results.

The subsequent four sections of this chapter describe in turn each of the four chosen classification algorithms. This is followed by a presentation of how each performed on the dataset together with a critical comparison. The majority of time was spent researching and developing CART and Neural Network tools, thus the emphasis on the description of these two techniques as opposed to the more classical, and generally wider understood, discriminant analysis and regression analysis approaches.

5.4 Discriminant Analysis

(Canonical) Discriminant Analysis (DA) is a technique using least squares methods to separate data into two or more groups. Data points are characterized by several variables; the optimal *discriminant function* is assumed to be a linear function of the variables and is determined by maximizing the between group sum of squares for fixed within group sum of squares. This technique is well described (for example in Manly, 1998) and a detailed discussion will not be given here.

DA is primarily used to classify cases into the values of a categorical dependent, usually a dichotomy. If it is effective for a set of data, the classification table of correct and incorrect estimates will yield a high percentage correct. In general there are several purposes for DA:

- To investigate differences between groups.
- To determine the most parsimonious way to distinguish between groups.
- To discard variables which are little related to group distinctions.
- To classify cases into groups.
- To test theory by observing whether cases are classified as predicted.

Discriminant analysis shares all the usual assumptions of correlation, requiring linear and homoscedastic relationships and untruncated interval or near interval data. Like multiple regression, it also assumes proper model specification (inclusion of all important independents and exclusion of extraneous variables). DA also assumes the dependent variable is a true dichotomy since data which are forced into dichotomous coding are truncated, attenuating correlation.

DA is an earlier alternative to logistic regression (see 5.5.3), which is now frequently used in place of DA as it usually involves fewer violations of assumptions, is robust, and has coefficients which many find easier to interpret.

5.5 Regression Algorithms

Regression analysis is concerned with investigating the relationship between several variables in the presence of random error. In particular we build a model in which one of the variables (the *dependent variable*) is expressed as a linear combination of the remaining variables (which are referred to as the *independent* or *explanatory variables*).

The method of least squares is used to estimate the parameters of the model from a given data set. This process of estimation is often referred to as *fitting* the model. The following description is intended to provide an introduction to the topic. The reader is referred to Draper and Smith (1966) for a comprehensive examination of the subject.

5.5.1 General model with additive error

We seek a model which will enable us to represent the relationship between the dependent variable, y, and the set of independent variables $x_1, x_2 \ldots, x_k$. In general this relationship will not be completely deterministic and the relationship will contain a random experimental error term, ε. In order to investigate the relationship a value of y, y_j, is determined when $x_1, x_2, \ldots, x_k$ take the values $x_{1j}, x_{2j} \ldots, x_{kj}$ respectively. This is repeated for $j = 1, \ldots, n$. The general model used to represent the relationship is of the form:

$$y_j = \Phi(x_{1j}, x_{2j}, \ldots, x_{kj}) + \varepsilon_j \qquad j = 1, \ldots, n$$

where Φ denotes some function of the x's and contains unknown parameters which have to be estimated.

Attention is usually limited to the general linear model in which Φ is a linear function of the unknown parameters, such that:

$$y_j = \alpha + \beta_1 x_{1j} + \beta_2 x_{2j} + \ldots. + \beta_k x_{kj} + \varepsilon_j \qquad j = 1, \ldots, n$$

5.5.2 General model with multiplicative error

In some situations it is more appropriate for the model to include a multiplicative error term rather than the additive term used above. The general model then becomes:

$$y_j = \Phi(x_{1j}, x_{2j}, \ldots, x_{kj})\delta_j$$

where δ_j denotes the random experimental error term.

This model is often appropriate for data for which the relationship between the variables is clearly non-linear. In this case a suitable logarithmic transformation can sometimes be applied to achieve an approximately linear relationship with an additive error term.

5.5.3 Logistic regression

Binomial (or binary) logistic regression is a form of regression that is used when the dependent is a dichotomy and the independents are continuous variables, categorical variables, or both. *Multinomial logistic regression* exists to handle the case of dependents with more classes. Logistic regression applies maximum likelihood estimation after transforming the dependent into a logit variable (the natural log of the odds of the dependent occurring or not). In this way, logistic regression estimates the probability of a certain event occurring.

Logistic regression has many analogies to ordinary least squares (OLS) regression: the standardized logit coefficients correspond to beta weights and a pseudo R^2 statistic is available to summarize the strength of the relationship. Unlike OLS regression, however, logistic regression does not assume linearity of relationship between the independent variables and the dependent, does not require normally distributed variables, does not assume homoscedasticity and in general has less stringent requirements. The success of the logistic regression can be assessed by looking at the classification table, showing correct and incorrect classifications of the dichotomous, ordinal, or polytomous dependent. Goodness-of-fit tests are available as indicators of success as is the Wald statistic and other tests of the model's significance.

5.6 Classification and Regression Trees (CART)

5.6.1 The foundations of CART

Classification and Regression Trees (CART) is a classification method that has been successfully used in many healthcare applications. Example applications include creating case-mix groups (Smith *et al.*, 1992), minimum data requirements (Hornberger *et al.*, 1995), cancer survival groups (Garbe *et al.*, 1995) and Intensive Care (Ridley *et al.*, 1998).

Breimen *et al* (1984), the founders of the technique, use the following University of California study as a means of introducing the reader to the technique. The study uses measurements recorded when a heart attack patient is admitted to hospital and attempts to classify the patients into low-risk and high-risk groups. Nineteen variables are recorded in the first twenty-four hours, including age, blood pressure and seventeen other ordinal and binary variables summarising the patient's condition. The CART method produces a tree that, by answering a series of yes/no questions, can be used to classify the patient. The authors found that it was possible to identify a high risk group of those patients not surviving more than 30 days based on minimum systolic blood pressure, age and whether sinus tachycardia was present. Figure 5.1 shows the tree produced in the study. In the tree, the letter F indicates low risk and G for high risk.

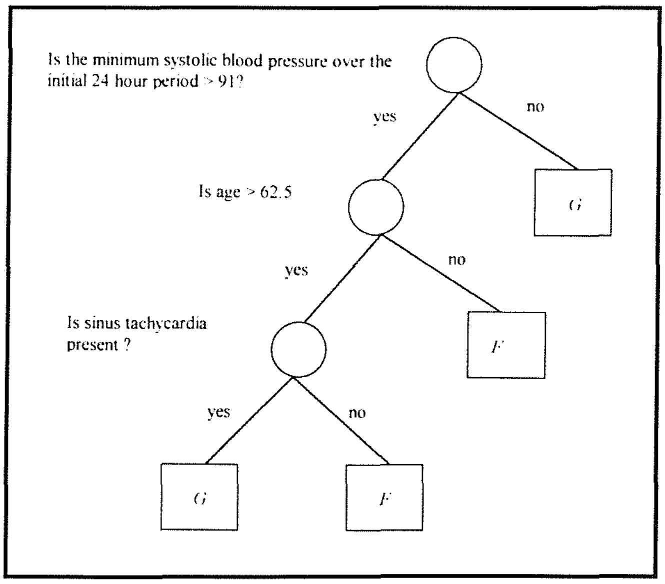


Figure 5.1: CART analysis of cardiac patients (from Breimen *et al.*, 1984)

5.6.2 The method

The first step in producing a tree is to decide which variable is to be predicted (LoS, operation times, survival rates etc.). If the variable is ordinal then the variance is used to measure the *purity* in the group, if the variable is categorical then deviance is used. An algorithm is used to split the original dataset into sub-populations of increasing purity (decreasing variance or deviance). At each junction of the tree is a *node*. A *terminal node* is a node at the end of the branch of the tree. At each node in the tree the algorithm searches through each of the independent variables in turn. For each variable it finds the best binary split that produces a node with the smallest variance or deviance. Then it selects the variable that has produced the best binary split (best of the best). The *parent node* will thus be split on this variable with the split as defined. This may have the effect of leaving one of the *child nodes* with a higher variance/deviance than the parent node. The algorithm however continues to branch from each of these child nodes until defined stopping rules have been fulfilled.

An issue in CART analysis is when to stop the partitioning, i.e. when do we say that the variance has not significantly reduced? It would be possible to create a tree where each terminal node has zero variance by having just one case in each node. However this would be statistically irrelevant and practically useless. It is necessary therefore to introduce *stopping rules* so that terminal nodes have sufficient size to yield reasonable and statistically robust results. Stopping rules include:

- Stop when nodes contain a certain number of cases.
- Stop when reduction of variance is below a certain threshold.
- Stop when a maximum number of terminal nodes (or layers) have been produced.

Care must be exercised when defining stopping rules and should account for the number of cases in the dataset. A terminal node with less than 30 cases, for example, can be expected to yield little predictive power and lack statistical robustness. Standard bounds are no less than 50 cases per node, a significance level of 1% on the reduction of variance in order to split a node, and a maximum of around 10 terminal nodes. Once a tree is constructed statistical summaries can be produced at the terminal nodes which can be used to form the classes.

5.6.3 CART components

There are 4 components required to construct a regression tree:

1. A set of questions of the form: Does x_i belong to the set A. The answer to such questions induces a split of the predictor space, cases associated with A and those with the complement of A. The sub-samples form the nodes.
2. A goodness of split criterion $\Phi(s, t)$ that can be evaluated at any split s at any node t.
3. A means of determining the appropriate size of the tree.
4. Statistical summaries at terminal nodes of the tree, for example, node averages and frequency distributions.

5.6.4 Validating the trees

Once a tree has been produced it should be validated to give an estimate of the accuracy of its classifications. The same data that is used to construct the tree cannot be used to test the classifications, as the estimate will be over-optimistic. This is overcome by splitting the data into two sets, A and B. The cases in A must be independent and identically distributed to the cases in B. This has the drawback that it reduces the sample size used in the construction of the tree. Set A can be used to train the data and build the tree (training set) and set B to test the robustness and validity by forcing the data through the tree (testing set).

For smaller samples there is a technique called *V-fold cross validation*. This involves three stages:

1. Split data into v sub-sets
2. Classify on A-A_v for each v
3. Cross validate over all samples, combine miss-classification rates to measure accuracy

Standard statistical methods can be used to compare the values of the training set and test set at any node. The significance test to compare data sets can be described as the following:

Let n_1 = number in group at node from training set
n_2 = number in group at node from test set

(A) Categorical dependent variable

Let r_1 = number responding 'yes' in group at node from training set
r_2 = number responding 'yes' in group at node from test set

Let π_1 and π_2 be the true proportions responding 'yes' in the training and testing populations respectively.

For a large sample (>20) r_1 is approximately normal distributed with mean $n_1\pi_1$ and variance $n_1\pi_1(1-\pi_1)$, and r_2 with mean $n_2\pi_2$ and variance $n_2\pi_2(1-\pi_2)$.

Under the null hypothesis of $\pi_1 = \pi_2 = \pi$ we obtain:

$$Variance\left(\frac{r_1}{n_1} - \frac{r_2}{n_2}\right) = \frac{\pi(1-\pi)}{n_1} + \frac{\pi(1-\pi)}{n_2}$$

and

$$Expectation\left(\frac{r_1}{n_1} - \frac{r_2}{n_2}\right) = 0$$

where π is estimated by $p = \dfrac{r_1 + r_2}{n_1 + n_2}$

A 95% confidence interval for $\pi_1 - \pi_2$ is given by:

$$\frac{r_1}{n_1} - \frac{r_2}{n_2} \pm 1.96\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$$

(B) Ordinal (continuous) dependent variable

Let $\bar{x}_1$ = the mean value within the group at node from training set

$\bar{x}_2$ = the mean value within the group at node from test set

s_1^2 = the estimated group variance at node from training set

s_2^2 = the estimated group variance at node from test set

Let μ_1 and μ_2 be the true means in the training and testing populations respectively. Let σ_1^2 and σ_2^2 be the true variances in the training and testing populations respectively estimated by s_1^2 and s_2^2 for large n_1 and n_2.

Under the null hypothesis of $\mu_1 = \mu_2$ we derive:

$$Variance(\bar{x}_1 - \bar{x}_2) = \left(\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}\right)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)$$

and

$$Expectation(\bar{x}_1 - \bar{x}_2) = 0$$

Hence a 95% confidence interval for $\mu_1 - \mu_2$ is calculated as:

$$\bar{x}_1 - \bar{x}_2 \pm t_{\alpha/2,(n_1+n_2-2)} \sqrt{\left(\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}\right)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

Confidence intervals can be used to judge whether or not the two data sets are statistically the same. If the interval produced at any given node contains zero then it suggests the test data supports the training data; if zero is not contained in this interval then the node should be treated with caution.

5.6.5 The CART algorithm details

Classification

1. For the current group calculate the total variance within the group.

2. For each value of independent variable, calculate the total value of the dependent variable in that group ($\sum x$), the total value squared ($\sum x^2$), and the number of items of data in that group (*N*).

3. Sort the values of the independent variable into increasing order of the mean value of the dependent variable.

4. For each independent variable calculate the best point at which to split the sorted mean values to produce the minimum variance.

5. Split the data based on the best independent variable in order to reduce the total variance calculated as:

$$\frac{\sum_{\text{all groups}}\left[\sum x^2 - \frac{\left(\sum x\right)^2}{N}\right]}{\text{Total observations}}$$

6. Choose a suitable sub-group of the data as the current group, and repeat the above steps until either the data is split into groups of size less than a minimum number, or the reduction in variance obtained by a split of the data is below a minimum value.

Regression

1. For each pair of adjoining subgroups, calculate the change in variance resulting from the amalgamation of these two groups.

2. Combine the sub-groups whose combination will result in the least gain of variance.

3. Repeat the above steps until the desired numbers of sub-groups are obtained.

5.6.6 Computational time issues

With uncensored data numerated covariates, for example patient age, these are ordered such that the analysis is carried out on ($< x$, $\geq x$) and is performed n times, where n is the number of individual ages. We have O(n) computations to make. Enumerated covariates, A, B, C and D with X = {A, B, C, D}, are ordered such that E(A) ≤ E(B) ≤ E(C) ≤ E(D). Therefore it is logical to look at the groupings {A}, {B, C, D}; {A, B}, {C, D}; {A, B, C}, {D} etc., with order O(n-1)

With censored or categorical data numerated covariates, analysis is carried out with order O(n). However if we are analysing an enumerated split, for example hospital specialties A, B, C and D with X={A, B, C, D}, we have to investigate all possible groupings: {A}, {B,C,D}; {A,B}, {C,D}; {A,B,C},{D}; {B},{A,C,D}; {B,C}, {A,D}; {C}, {A,B,D}; {A,D}, {B,C} with order O(2^{n-1} - 1) for nominal and O(n^2 - 1) for ordinal variables. The computation time for the construction of the tree sequence for censored data is of obvious concern, although a possible solution for large number of enumerated types would be the use of factorial designed experiments.

5.7 Artificial Neural Networks

Artificial neural networks are parallel computing devices consisting of many interconnected simple processors. In essence they are attempting to mimic the behaviour of the most powerful asset known to man, the human brain. Although each processor is quite simplistic, a collection of these units (a *network*) gives rise to a powerful computational tool. Each processor in the network is only aware of signals it periodically receives and the signal it periodically sends to other processors, and yet such simple local processors are capable of performing complex tasks when placed together in a large network of orchestrated cooperation.

Artificial neural networks have their roots in work performed in the early part of the twentieth century, but only during the 1990s, after the breaking of some theoretical barriers and the advances in computing power, have these networks been widely accepted as useful tools. A plethora of books and papers have been published on artificial neural networks. This section aims to provide the reader with a general overview of the topic. For a more comprehensive and exhaustive description, the following references are useful starting points: Anderson (1995), Haykin (1999), Callan (1999) and Kay and Titterington (1999).

5.7.1 The basic components

The neural network is the collection of units that are connected in some pattern to allow communication between the units. These units, also referred to as *neurons* or *nodes*, are simple processors whose computing ability is restricted to a rule for combining input signals and an activation rule that takes the combined input to calculate an output signal. Output signals may be sent to other units along connections known as *weights*. The weights usually excite or inhibit the signal that is being communicated. The net input of weighted signals received by a unit j is given by

$$net_j = w_0 + \sum_{i=1}^{n} w_i x_i$$

where w_0 is the biasing signal, w_i the weight on input connection ij, x_i the magnitude of signal on input connection ij and n is the number of input connections to unit j. An illustrative schematic of a single network unit with three incoming signals is shown in Figure 5.2.

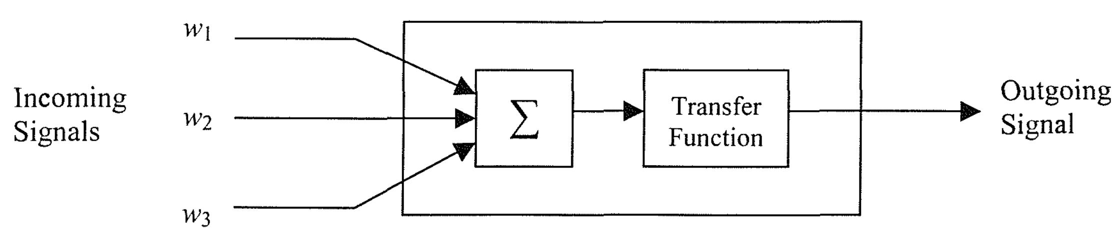


Figure 5.2: A single network unit

The summed value *net* is passed on to a second processor within the unit, the *transfer function*, which computes the output value of unit *j* determined by:

$$o_j = f(net_j)$$

where o_j is the unit output and $f(\ )$ is the output transfer function. Common transfer functions include the step, sigmoid and hyperbolic tangent sigmoid functions.

One of the intriguing aspects of neural networks is that although they have units with limited computing capability, when many of these units are connected together, the complete network is capable of performing a complicated task. Figure 5.3 illustrates an example of a simple two-layered feedforward network (the input layer is not normally regarded as a layer as its purpose is simply to enter the input values).

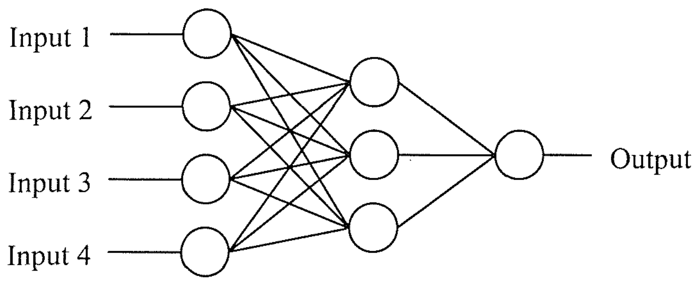


Figure 5.3: A two-layered feedforward network

5.7.2 The backpropagation network

As opposed to a traditional and rigorous programming approach, where the developer blueprints every command to be executed, a neural network is left to itself to learn the underlying theories of the problem and the procedures required for solving it. This process of learning is known as network training. The *knowledge* obtained by the network is stored in its weights and biasing values.

A popular training algorithm for feedforward networks is the backpropagation algorithm. Feedforward networks that are trained using this approach are commonly known as backpropagation networks (BPN). The general concept of this technique is to constantly perform corrections to the individual weights and bias values, with the objective of reducing the output error, corresponding to the magnitude it has contributed to the output error. In other words, the algorithm redistributes the "blame" to the individual weights and biases according to their contributions to the overall error.

The backpropagation algorithm is a supervised learning method in which the developer first selects a set of training data from historical records. The training data would consist of samples of inputs, I, together with a corresponding set of targeted output(s), t. For each data point in the training set, the algorithm sweeps through the network twice. The forward sweep first propagates the input vectors through the network to compute the output values at the output layer (with the weights and biasing values of the network initialised to small arbitrary values). The unit's total output error, e, obtained by summing the individual differences between the network's output and the targeted outputs (t-o), are then in turn propagated backwards through the network to determine how the weights are to be changed during training. Common error functions used include the sum of squared error (SSE) and the mean squared error (MSE).

An *epoch* is a complete cycle through each data point in the training set. Network training terminates when the sum of all errors is below a predefined target error.

The main parameter of the backpropagation algorithm is the networks *learning rate*, r. In order to make the effect of weight changes smoothly, the transfer function of the units must be smooth. The learning rate should be kept low to allow a smooth and steady descent on the error surface. A commonly adopted learning rate is the logistic sigmoid function:

$$o_j = \frac{1}{1 + e^{(-net_j)}}$$

The amount of correction that is to be made on a particular weight connected to the output layer is computed using the following equation:

$$\partial w_j = r e_j o_j (1 - o_j) I_j$$

where ∂w is the weight change, r the network learning rate and e_j the SSE of unit j.

Thus the new value of the weight is found by:

$$w_{new} = w_{old} + \partial w$$

Further discussions on the backpropagation algorithm may be found in Mehrotha *et al.* (1997).

5.7.3 Architecture selection

An important consideration affecting the likely success or failure of a neural network is the choice of the number of units in the input, hidden and output layers. With prior knowledge of the dataset, respectable initial values may be chosen. For most problems however, the initial values are unclear and often further complicated by the *curse of dimensionality* which states that the number of data points required for training increases non-linearly as each input variable is added.

The choice of the number of units to be used in the hidden layers also strongly affects the generalisation capability of the network. Generalisation in the neural network context refers to the ability of a network to arrive at a configuration that is able to correctly process input data that has never been presented to it before. As a general rule of thumb (Callan, 1999), start with a two-layered network comprising of 30% to 50% of the total number of units in both the input and output layers in the hidden layer. While the lack of units in the hidden layers would obviously cause the network to be insufficiently powerful to model the problem at hand, the presence of too many units may cause the network to *overfit*. Under such circumstances, the network has begun to memorise the training set and will not allow for flexibility in processing new input sets. Overfitting is observed when a higher test error is observed when compared to the final training error.

Another issue is that of *overtraining*. This occurs when the training data is presented to the network on too many occasions, which will result in the network memorising the training set. A practical solution is to periodically test the network with a separate data set known as the validation set.

5.7.4 Dataset selection

The choice in the dataset used for training, validation and testing is an important factor for the success in network training and generalisation. The dataset must firstly be representative of the problem. Specifically, the data must include all previously observed eventualities. A guideline by Baum and Haussler (1989) in determining the training set size is found by calculating:

$$N > \frac{W}{\varepsilon}$$

where N is the number of training data points required, W is the total weight and biasing values in the network and ε is the proportion of allowed errors in testing.

The numeric values of training data may also need to be pre-processed prior to application. For example, the data will usually need to be scaled so as to prevent domination by variables that are computationally larger in magnitude. A crude method of scaling would be to divide the observations of a particular variable set by the largest observation. This then limits the largest value within the set to unity.

5.7.5 Avoidance of local minima

The backpropogation training algorithm is a gradient descent method. If the output errors were plotted against the possible range of weights and biasing values, a two-dimensional view of that graph would be similar to that shown in Figure 5.4. This graph is known as the *network error surface.*

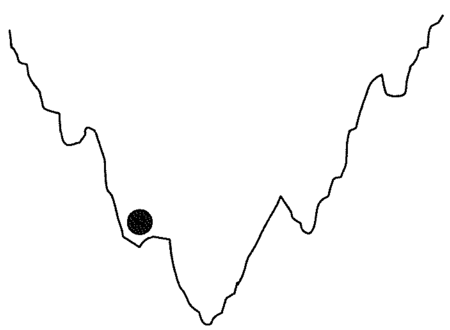


Figure 5.4: An example of network's error surface

The algorithm aims to minimise the error by taking the steepest descent on the error surface. However, as the above figure shows, the error surface will consist of a combination of uphill and downhill slopes and valleys. It is therefore possible for the algorithm to be trapped in one of the many local minima.

The inclusion of a momentum constant can help overcome this event whilst in the process of searching for a global minimum. This constant provides additional magnitude in the error descent by simply adding a fraction of the previous weight change to the current computed change. Thus, if the previous weight change resulted in a large downhill shift in error and the network is now in a local minima, the previous weight change can help the network to overcome the small valley(s).

The amount of correction that is to be made on a particular weight connected to the output layer is thus modified to include a momentum constant α.

$$\partial w_j(n+1) = re_j o_j(1-o_j)I_j + \alpha\partial w_j(n)$$

5.8 Results of the Comparison Study

5.8.1 Summary of results

Initial time was spent randomly splitting each of the four datasets into training and testing datasets (with approximately half the total number of observations in each). The same training and testing datasets were used for each of the classification techniques in each study. The discriminant and regression models were built using SPSS (version 10). The CART trees were constructed using a developed package *Apollo* (section 5.9). NeuroSolutions (version 3) was used to build, train and test the neural networks. For each study, the models were trained and then tested. Their accuracy and run-time were recorded. Further details may be found in Appendix D.

Table 5.3 summarises how each of the classification techniques performed in each of the five studies. Recall that with a categorical dependent variable, the percentage of correctly classified cases is presented (rescaled as a number between 0 and 1). With a continuous dependent variable, the value of the correlation coefficient (r) is given. The time taken to test the models is shown in seconds. Appendix D contains more detailed information with illustrative results for each of the four techniques.

Table 5.3: Performance of the different classification techniques

Study	Discriminant		Regression		CART		Neural Nets	
	Accuracy	Time	Accuracy	Time	Accuracy	Time	Accuracy	Time
1	0.23	12	0.29	1	0.32	1	0.33	39
2	0.67	6	0.81	1	0.87	2	0.87	39
3	0.45	165	0.62	5	0.60	20	0.34	620
4	0.54	7	0.68	2	0.78	6	0.79	165
5	0.73	3	0.75	2	0.79	12	0.74	129

5.8.2 Discussion of research findings

Discriminant Analysis (DA)

DA is used to classify cases into the values of a categorical dependent. A large amount of set-up time was therefore required to split the continuous dependent variables into necessary groups. For this purpose, groups were derived using percentile splits with a total of ten groups constructed.

DA consistently gave the lowest accuracy and took a longer time to run than both the Regression and CART approaches, although it was faster than the Neural Network tool. DA's poor performance may be explained by its linear structure that seems unable to tune itself to the structure of the datasets. In contrast, for example, regression models may be tuned to the dataset through the use of interaction terms. In DA the assumptions of a linear structure appear too restrictive. DA performed particularly poorly for predicting LoS (studies 1 and 3). The LoS data has high skew and kurtosis which appears to severely disrupt the performance of the discriminants.

The resulting discriminant functions and their coefficients were not always easier to interpret and often shed little light on the structure of the data. A survey of healthcare professionals revealed that they often struggled to make clinical sense of the discriminant functions. This is a further downfall of the technique.

Regression Models

Models with both additive and multiplicative error terms were examined and the best accuracy recorded. When the number of variables is large, and thus running a full-factorial model is unfeasible, regression models ideally require that the user has at minimum a general understanding of the variables and may therefore evaluate suitable interaction terms. As a consequence, a large amount of time should be afforded to the consideration of appropriate model terms. Various models may then be built and evaluated.

The accuracy was typically comparable with that of CART and Neural Network approaches. It tended to do better when the number of records (data points) was large and actually produced the best correlation coefficient r for predicting hospital LoS (0.62). For studies 1 and 3 (continuous dependent variables) the best results were obtained after a logarithmic transformation was applied to the dependent, although a number of other transformations were also evaluated.

Careful selection of interaction terms were considered and this process was greatly aided by the results from CART. Tree-based structures graphically illustrate the variables that cause the tree to split and branch, with nested variables (*parent* and resulting *child* branches further down the tree) providing a good initial indication of potential interaction terms to include in the regression model. Further detail of the usefulness of CART in the appreciation of possible interaction terms may be found in Appendix D. Once the desired model had been chosen, the actual run-time was the quickest of all of the techniques. Even with 17,974 records (study 3) it only took 5 seconds to run.

Logistic regression performs particularly poorly compared to CART and Neural Networks when there are a high number of categorical independent variables (study 3). This is consistent with what is known about the properties of the discriminant and logistic regression algorithms; the properties of high skew (>1) and kurtosis (>7) along with the presence of binary/categorical variables disrupts the performance of these algorithms (King *et al.*, 1995). Such conditions are well suited to symbolic learning algorithms, as observed by the performances of CART and Neural Networks (study 3).

When considering ease-of-use and interpretability, regression models are fairly straightforward to interpret and understand although there is a large set-up time and a need to run a number of different models accounting for interaction terms and transformations to the dependent variable as necessary.

CART

The performance of CART was consistently as good as, or better than, Regression and Neural Networks models and always more accurate than Discriminant Analysis. It appeared to perform well over all datasets, indicating that the number of records or number of variables does not affect its relative accuracy compared to the other techniques. It performed particularly well on datasets with high skew and kurtosis, indicating that they are furthest from the (multivariate) normal. Symbolic learning tools, like CART, are generally non-parametric. That is, they do not make any assumptions about the underlying distributions. This is why we observe a consistently good performance relative to the other approaches. Although run time is a potential concern for tree-based methods, the observed times were low even for larger datasets and those with a high proportion of categorical explanatory variables.

The CART output is simple and straightforward to interpret. Surveyed healthcare professionals particularly favoured the clear pictorial way in which the tree was constructed and captured on screen. It was found that this technique was the easiest to clinically interpret, allowing the user to discover which variables were of importance and furthermore their relative importance and quantifying the point at which to split on each variable. Such a method can challenge perceived beliefs or reinforce existing clinical judgements. A great practical advantage of this tool is the ability to combine expert clinical/managerial knowledge with the power of statistical analysis. From any node in the tree, it is possible for the user to split the node on a user-selected explanatory variable. The CART algorithm may then be continued from the created child nodes. The combination of CART and local expert clinical knowledge proved to be a powerful tool during discussions with, and use of the tool by, the surveyed medical staff.

Neural Networks

The performance of the Neural Net varied throughout the study depending on the features of the dataset, although overall it performed well. A large amount of initial effort was required to train and validate the models. Run time was by far the slowest of the studied classification tools, because of the large number of parameters, with the neural net taking over 10 minutes to train the largest dataset (study 3). Improvement in the speed of neural networks is a large research area (for example see Fahlman, 1999).

A major drawback of neural networks is that they are difficult to set up to produce good results. For example, to run backpropogation properly, a number of parameters needed to be adjusted. For example, an important decision concerns the number of layers and the initial step-size and momentum rates of the backpropogation network. Any small changes in these parameters can decrease the performance substantially. After running a number of networks and gaining an appreciation of the tool, an insight was made on choosing suitable starting conditions. These appeared to indicate an initial step-size of 0.7, a momentum rate of 0.5 and a network with two-hidden layers although there is no guarantee that the results are the best that could be achieved. An automatic method of parameter selection for backpropogation is another important current research topic.

Accuracy was restricted (especially when comparing the results with those obtained from CART and Regression models) when the network was handling datasets with high variance or deviance (studies 3 and 5). The tool was equally or more accurate than all of the other tools in studies 1 and 2 with the smallest dataset (582 records).

As with Regression and DA, Neural Networks are not always easier to interpret. Coupled with the large set-up time, long run-times and multitude of possible network configurations, this tool is perhaps best suited to an experienced user. Extreme care should be exercised if intending to give the tool to a healthcare professional with limited statistical and Neural Network knowledge.

5.8.3 General conclusions

In order to capture the uncertainty and variability amongst the patient population, a number of classification techniques have been considered and evaluated for their relative performances and practical usefulness in predicting various healthcare indicators using a number of real-life datasets. This research has indicated that in practice there is no single *best* classification tool but instead the best technique will depend on the features of the dataset to be analysed and any preferences of end-users. The research has made a start in investigating what these features are with particular emphasis on healthcare data. A summary of the main findings are as follows:

- Overall the results were very promising with each tool making a statistically significant contribution in each study (values of r and the percentage correctly classified were all significant at the 99% level).

- Regression models consistently had the fastest run-times, although the difference in times compared to CART and DA is likely to be insignificant in practice. Neural Networks require significantly more time to train and validate models.

- In general CART, Regression and Neural Network classification approaches gave similar accuracies, although CART was the only tool to give consistently good results. DA performed poorly throughout the study.

- CART was well suited to datasets with large skew (>1) and kurtosis (>7) and where there were a large proportion of categorical independent variables. CART makes no assumption about the underlying distribution, hence why CART performed consistently well. In contrast, these conditions limit the performance of discriminant and regression models, where the data is furthest from the (multivariate) normal.

- Neural Networks produced the best accuracy when dealing with smaller datasets but performed slightly disappointingly when handling dependent variables with high levels of variability or deviance.

- If ease of use and human understanding are high priority, symbolic algorithms such as CART should be chosen.

- A number of healthcare professionals were surveyed for ease of use and interpretability of the four techniques. The main concern focussed on the form of the DA discriminant function and the associated coefficients and weights of the Regression and Neural Network models respectively. Typically these were of seen to be difficult to interpret and often shed little light on the structure of the data.

A survey of hospital staff from the participating NHS Trusts has revealed that tree-based tools, such as CART, do have a greater practical appeal than that of the other tested techniques. This is a measure of the extent to which the CART algorithm produces comprehensible results that are generally easier to interpret by medical staff than the results of other algorithms, and on the time it took for hospital staff to understand the technique, prepare the data and actually perform the analysis to produce correct and meaningful results.

In practice clearly a balance must be made between the accuracy and interpretability of a proposed technique. Accuracy is undoubtedly important, especially when considering a number of healthcare variables such as predicting death or survival. We might however wish to avoid a situation in which we are obtaining accurate predictions but where the form of the classifier is complex and little confidence and knowledge is gained on the data structure. Such a *black box* approach is limited in producing interpretable classification rules both for understanding the prognostic structure and for the planning and management of healthcare in general.

The evolved generic framework incorporates the need for a patient classification technique to be adopted. Derived patient groupings may then be fed into developed simulation models and individual patients from each group passed through the particular healthcare system of concern. In order to capture the uncertainty and variability amongst the patient population, it was decided that a tree-based classification tool be utilised. CART performed well in the research and was particularly well received by the participating NHS Trusts. A statistical package incorporating a CART tree-based algorithm has been developed for use within the modelling process.

Integration of the mathematical modelling work and statistical analysis is demonstrated with the hospital resources and critical care modelling work (Chapters 6, 7 and 8). The successful integration has been enabled by the software development constructed in object-orientated code within the Windows environment. The developed classification package, *Apollo*, is presented in the next section of this chapter.

5.9 The Apollo Statistical Package

A statistical package, Apollo, has been developed as part of the evolved generic framework for modelling healthcare resources. Apollo incorporates a tree-based algorithm, similar to CART, that assists in the production of clinically and statistically meaningful healthcare groupings. For example, these could be patient groupings based on LoS, operation times or survival rates. Derived groupings may be automatically saved and fed in to developed simulation models within the genetic framework.

Apollo has been designed to enable healthcare personnel to create appropriate groupings of patients and carry out the necessary statistical analysis. At the highest level of functionality, Apollo may be used as a:

- Data exploratory tool, allowing the user to explore and understand in greater detail their data. For example, Apollo permits manual splitting of the data into desired groupings and the rapid extraction of a number of key statistics, time-dependent profiles and continuous distribution fitting.

- Tree-based algorithm tool for classification and prediction, allowing the user to derive statistically meaningful, easy to interpret homogeneous groupings. This aids understanding of the structure of the data, enabling the user to define interpretable classification rules as necessary.

In the context of the planning and management of hospital resources, likely patient groups would be based on LoS as the dependent variable. Independent variables could include any routinely collected, or other, data such as patient age, sex, status, specialty and clinical diagnosis (HRG). Apollo has been specifically designed to enable users to create groups of patients, capture the necessary statistical information for each group, namely demand profiles and LoS fitted distributions, and then automatically fed these into a hospital capacity simulation model. Apollo has been designed for ease of use with screens designed in a Windows environment using Delphi software, and the ability to save patient groupings for loading into a developed simulation model (Chapter 6).

A high-level appreciation of the functionality of Apollo is given in Figure 5.5.

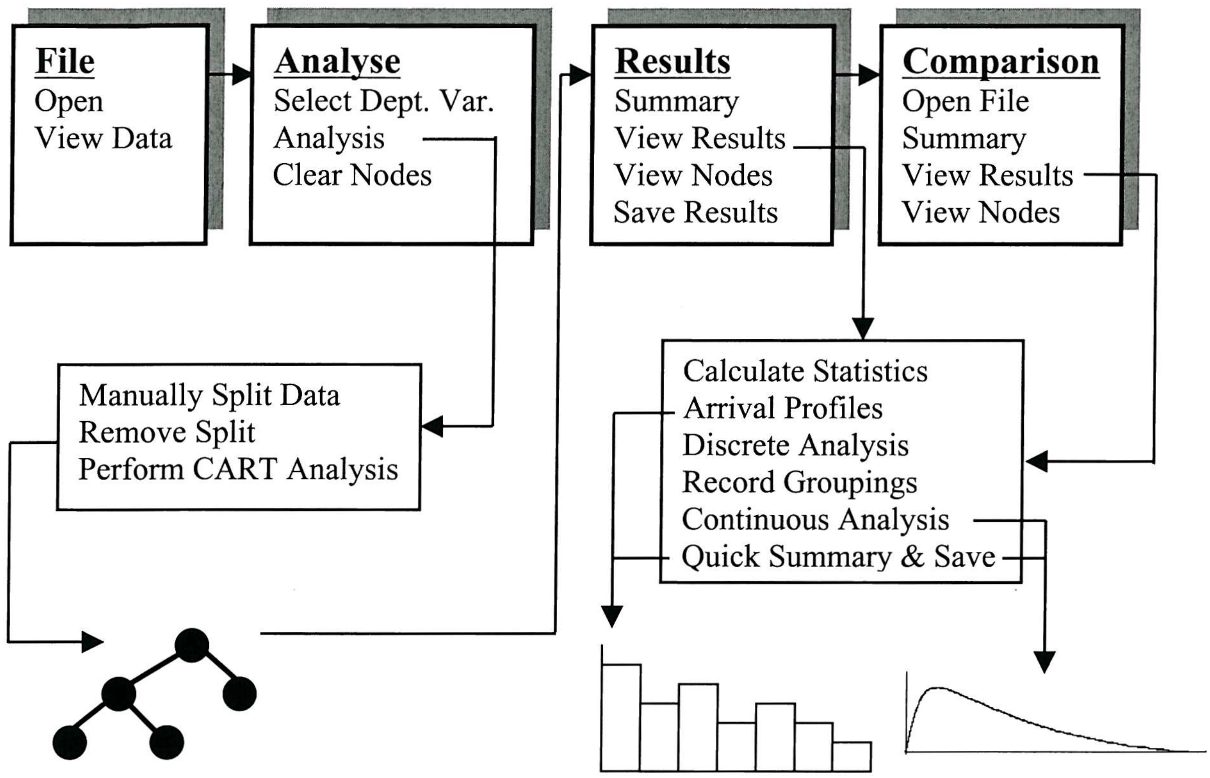


Figure 5.5: High-level Apollo functionality diagram

The subsequent sections of this chapter illustrate different aspects of this functionality through various screen-shots and discussions.

5.9.1 Loading, viewing and constructing a tree

Data files may be loaded in to Apollo from a current choice of .dbf or Excel formats. Figure 5.6 shows the main screen which forms the central control of the program. Through this menu the user can load and save data, view the data table and resulting tree and perform the necessary splits to construct a tree.

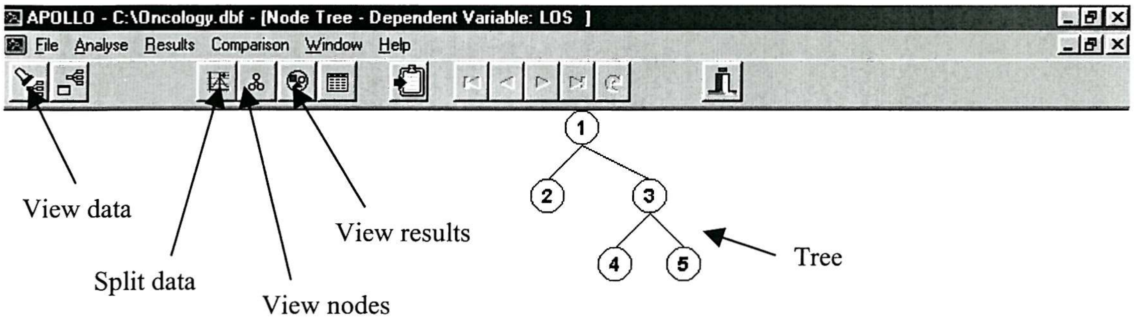


Figure 5.6: Apollo main screen

Having chosen the dependent variable from those available in the data table, the user can commence the splitting of the data (binary splits) adopting either a manual or CART approach, or a combination of both from the list of independent variables (Figure 5.7). For example, splitting the whole dataset (node 1) into two subgroups (nodes 2 and 3) representing those less than 50 years of age and those aged 50 or over.

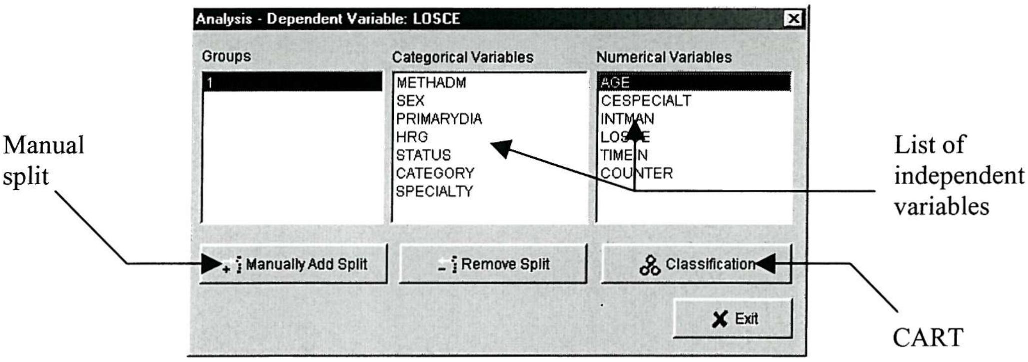


Figure 5.7: Splitting the data to construct a tree

5.9.2 CART analysis

If the tree-based algorithm is selected, the user will need to provide the necessary information in order to construct a tree. The dependent variable has already been selected. Independent variables to use in the classification algorithm must be chosen from the list of those available (e.g. sex, age, specialty). These may be nominal (categorical) or scale (continuous) variables. Additional information required includes the desired number of final groups and the minimum number of patients in each group.

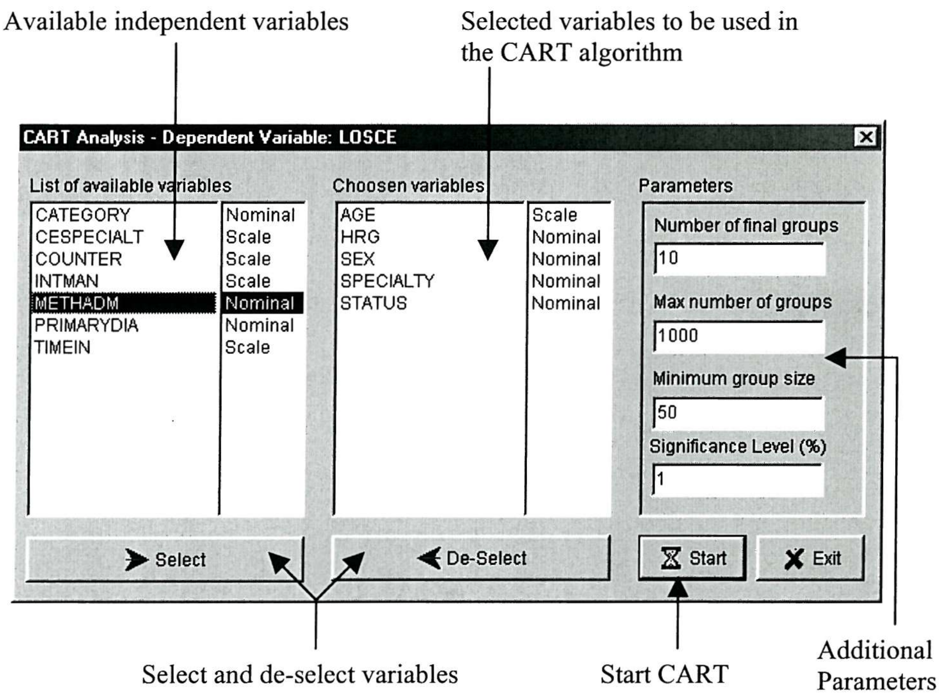


Figure 5.8: CART parameters

5.9.3 Viewing the results

A number of statistical indicators are available for each node within the tree. These include group mean, variance and inter-quartile range (for a continuous dependent variable) or percentage split and deviance (for a categorical dependent variable), together with rapid access to arrival profiles (for month, day and hour) and distribution fitting of continuous variables. A summary of nodes form provides easy access to node results (Figure 5.9).

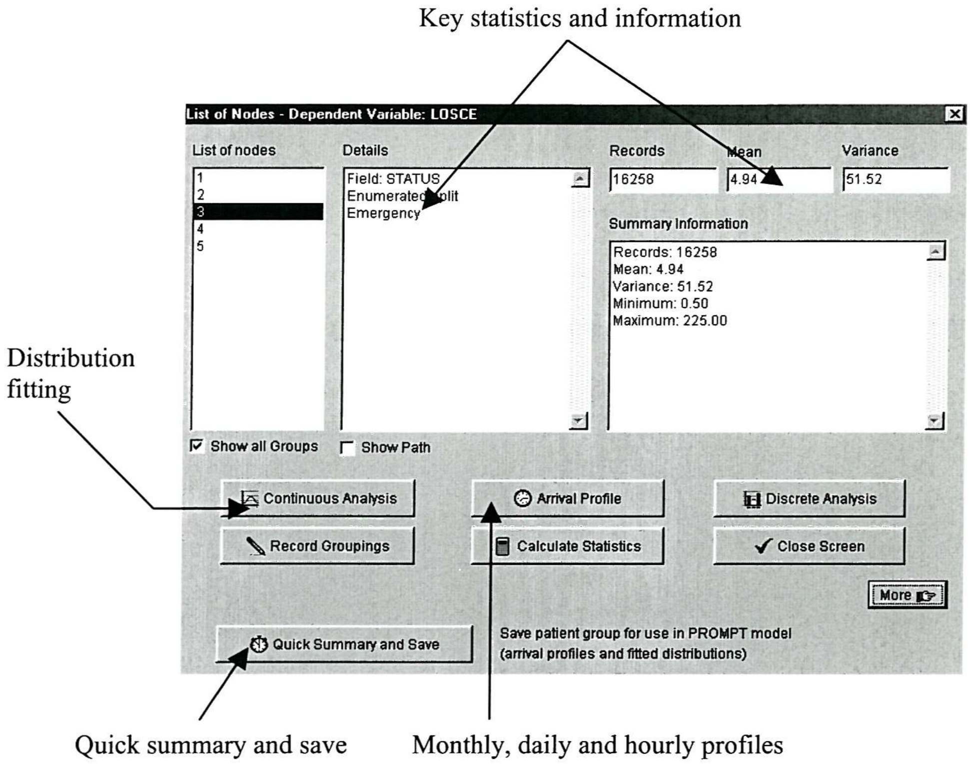


Figure 5.9: Summary of nodes form

The quick summary and save button allows the user to rapidly view and save the key results for use within the developed hospital capacity model (Chapter 6). For the planning and management of hospital beds, LoS distributions and patient arrival profiles are of particular help (Figure 5.10).

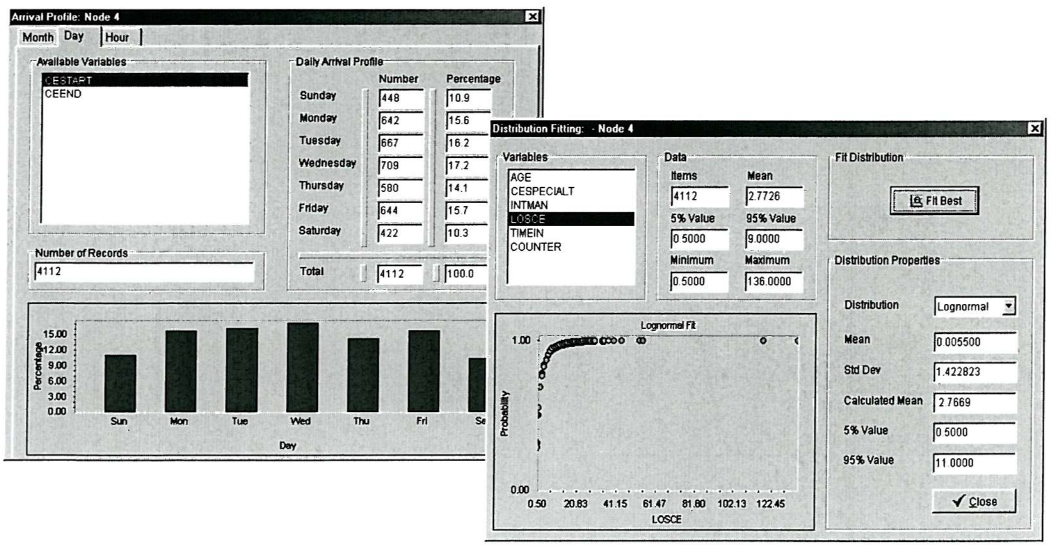


Figure 5.10: Patient group arrival profile and fitted LoS distribution

5.10 Chapter Summary

The research has explored the use of classification techniques for the creation of patient groupings. Necessary patient groupings may then be fed into developed simulation models and individual patients from each group passed through the particular healthcare system of concern. In order to capture the uncertainty and variability amongst the patient population, a number of classification techniques have been considered and evaluated for their relative performances and practical usefulness. Research has shown that there is not necessarily a single *best* classification tool but instead the best technique will depend on the features of the dataset to be analysed. The research has made a start in investigating what these features are with particular emphasis on healthcare data.

A survey of healthcare staff has however revealed that symbolic, tree-based tools, such as CART, do have greater practical appeal than that of the other tested techniques. This is a measure of the extent to which the CART algorithm produces comprehensible results that are generally easier to interpret by medical staff than the results of other algorithms and on the time it took for hospital staff to understand the technique, prepare the data and actually perform the analysis to produce correct and meaningful results.

A statistical package, Apollo, has been developed as part of the evolved generic framework for modelling healthcare resources. Apollo incorporates the CART tree-based algorithm that assists in the production of clinically and statistically meaningful healthcare groupings. For example, these could be patient groupings based on LoS, operation times or survival rates. Derived groupings may be automatically saved and fed in to developed simulation models within the framework.

Chapter 6 – A Simulation Model for Hospital Resources

6.1 Chapter Introduction

This chapter outlines the development, structure and validation of a simulation model for hospital resources, PROMPT, that has been developed in conjunction with the Royal Berkshire and Battle Hospitals NHS Trust. The model has been designed within the evolved generic framework for modelling of healthcare resources (Chapter 4).

6.2 PROMPT

6.2.1 Developing PROMPT

During the initial phase of work with the Royal Berkshire and Battle Hospitals NHS Trust, it was decided that a working title of the model would be desirable. For this purpose the acronym ***PROMPT*** *– Patient and Resource Operational Management Planning Tool* was chosen. This helped give the work an identity and facilitated easy reference to the model. The operational modelling approach captured in PROMPT can help to evaluate the implications of various options for patient care. In particular, the model allows *what if..?* scenarios to be examined for:

- Hospital beds (Core model)
- Operating theatres (Add-on module)
- Use of human resources, such as nurses, doctors and anaesthetists (Add-on Module)

PROMPT was developed within the evolved generic framework as detailed in section 4.4. A high-level summary on constructing a PROMPT model is shown in Figure 6.1.

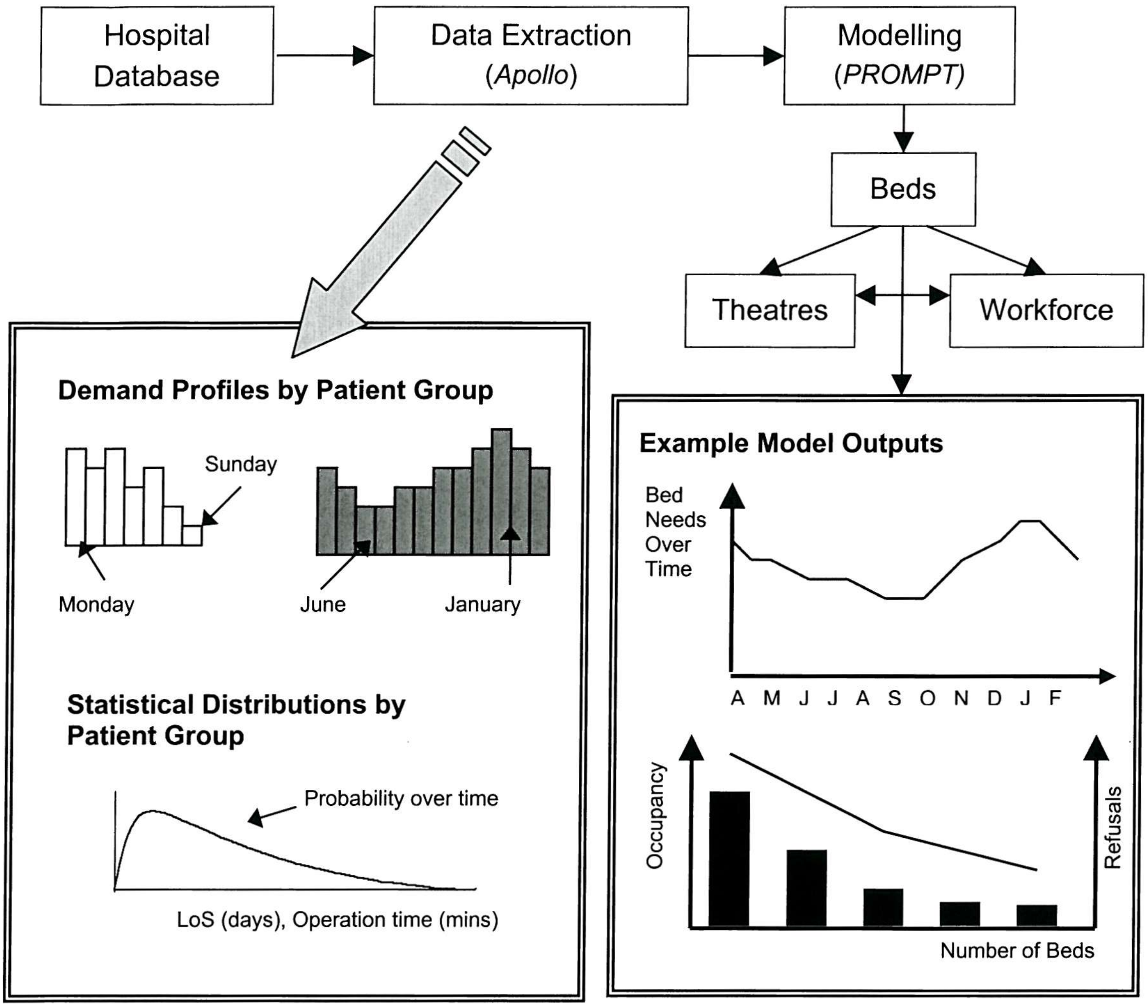


Figure 6.1: Constructing a PROMPT model

The statistical module, Apollo, links directly to the hospital database. The necessary statistical analysis is conducted within Apollo, including the creation of patient groups with demand profiles and distributions derived for each group. These groups are then directly fed into the PROMPT simulation model, and individual patients pass through the hospital system through time. Patient demand and LoS/operation time distributions are sampled from the appropriate patient group statistics. Bed needs together with optional workforce and theatre needs are assessed during each run of the simulation. Example outputs include daily numbers of beds in use, number of refusals over time, workforce rosters and theatre session utilisation.

6.2.2 PROMPT functionality

Over-time, and with an increasing knowledge of the hospital processes and perceived model utilisation, a schematic diagram of patient-flows though a hospital system was developed. This is suitably generic, allowing for the model to be readily used by other hospitals. Figure 6.2 identifies the structure of the developed simulation model. This diagram was evolved together with hospital staff, and so reflects their requirements and re-enforces their knowledge of the model structure. It draws together the various elements of the PROMPT model.

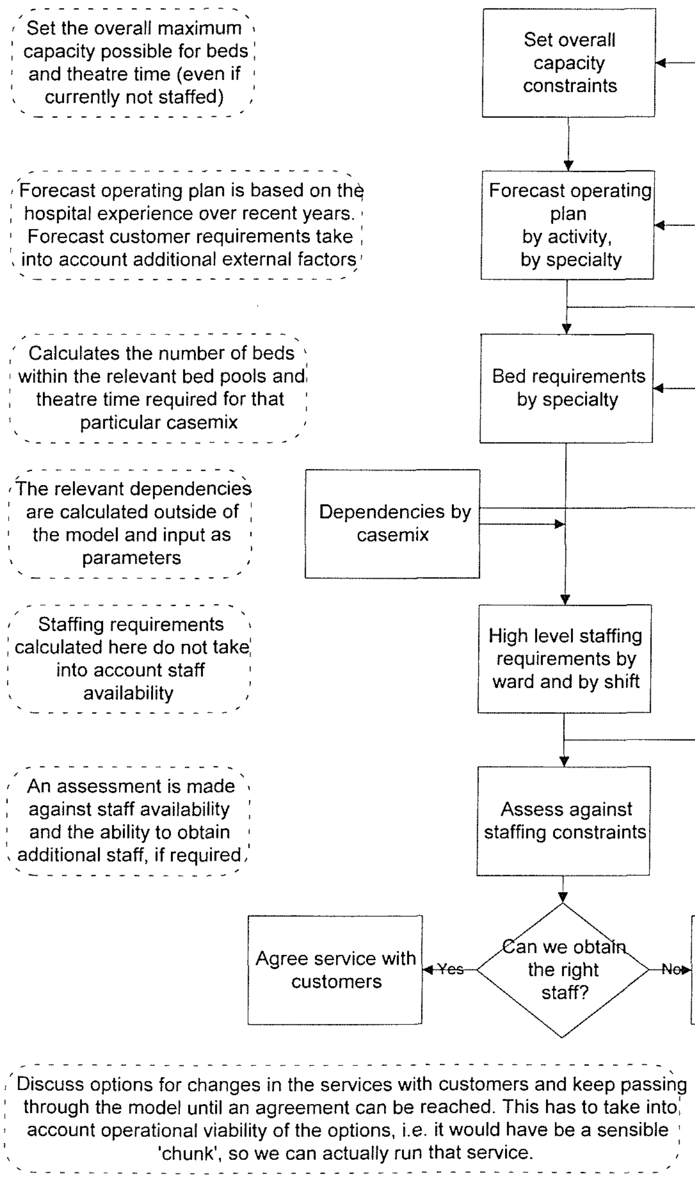


Figure 6.2: Schematic workings of the capacity model

PROMPT was developed in a Delphi environment using a three-phase simulation shell TOCHSIM (Hawkins *et al.*, 1992), developed at the University of Southampton (Appendix G). The use of Delphi enables the model to be easily tailor-made for the hospital, whilst the resulting familiarity of a Windows environment aids ease of use. The TOCHSIM simulation shell can reflect the experiences of the hospital very rapidly. For example, one year of hospital time corresponds to a few seconds of TOCHSIM time.

The model simulates the flow of patients through user-defined *care-units*, which may represent a ward, a specialty bed-pool or the hospital as a whole. Bed numbers for each care-unit are entered into the model. The bed numbers can be changed monthly and daily to represent the step-up and step-down of seasonal bed requirements and those wards which have some beds in use on a part-time basis. User-defined patient groups capture the hourly, daily and monthly arrival rates and reflect the observed (or other) demand patterns. For each hospital specialty, patient categorisation techniques were employed to capture the necessary variability in LoS and estimate the appropriate statistical distributions. A particularly useful classification method is the Classification and Regression Tree analysis (CART – see Chapter 5). Each specialty manager supplied the necessary information on admission rules and deferral times, whilst the bed-managers constructed patient priority listings (Appendix E).

The developed model takes individual patients through time as they arrive and pass through the hospital. For an arrival at the hospital, the model will attempt to acquire a suitable bed. Each patient has a defined *priority list* of suitable hospital bed-pools that they may stay in. Acquiring a bed is achieved by assessing the current bed situation for each of the defined care-units in turn on the priority-list and admitting, if possible, the patient into an free bed. Patients are classified as *outliers* if they stay in beds from a care-unit other than the first appropriate choice (for example, a general medicine patient stays in a general surgical bed). If no bed can be found, then a patient is allowed to wait for a user-defined time for a bed to become available. After this time, if still unsuccessful, an emergency patient will be transferred (out of the hospital) whilst an elective patient will be deferred and told to come back after a user-defined number of days (or a random time up to a maximum time limit). The user may also specify that a given number of beds be only accessible to emergency patients.

Throughout the admission process, emergency patients are always given priority over elective patients.

Once a patient has acquired a bed, the LoS is sampled from an appropriate statistical distribution (current choice is from a Weibull, lognormal, normal, gamma or negative-exponential). The patient will be discharged on completion of their LoS and the bed will become available.

6.2.3 Evaluating bed capacity options

Hospital managers requested a number of key statistical indicators, tables and graphs. The results collected are broadly at two levels; those at the care-unit level, such beds in use over time, transfers and deferrals, and those at the patient level, such as LoS and patient waiting times.

Bed utilisation is one of the most important measures of each specialty's workload. Bed occupancy, which is used in the model, calculates the proportion of time that a bed is occupied.

$$\text{Bed Occupancy} = \frac{\text{number of bed days used}}{\text{number of bed days available}} * 100\%$$

An important concept about bed utilisation that is often sadly forgotten or misunderstood is that of its relationship with the refusal rate. A refused admission occurs when no bed is available for an arriving patient.

$$\text{Refusal Rate} = \frac{\text{number of refused admissions}}{\text{total number of referrals (or admissions + refusals)}} * 100\%$$

As patient demand increases, both bed occupancy and refused admissions increase. It is important to understand how bed allocation and bed capacity planning effect both bed occupancy and refusals. Refused emergency admissions will typically be found a bed in another hospital specialty and will become an outlier. Outliers put stress on other specialties, as they are in effect an unexpected emergency demand. Refused elective admissions result in patient deferrals and distress to the patient as well as having a consequence on elective waiting lists. Case studies later in this chapter will demonstrate the complex relationship between occupancy and refusals, and the implications on hospital resources.

6.2.4 Operating theatre module

Building on the foundation of the bed capacity model, as described above, an add-on operating theatre module was developed. This module helps the user to evaluate operating theatre capacities. It is necessary to refer back to the provision of beds when considering the planning and management of theatres, since without a bed in the first place, an in-patient cannot be sent to theatre. Two classes of patients now exist: *procedure patients* (who require an operation) and *non-procedure patients* (who do not require an operation). Non-procedure patients stay in the bed and are discharged on completion of their LoS. The flow of procedure patients is now considered in more detail: patients acquire a bed, queue for theatre, have an operation, return to the hospital bed and stay for a LoS before discharge. As with LoS, the patient operation time will vary between patients, and thus needs to be captured using a statistical distribution rather than an average time in the PROMPT model. Various theatre session and patient scheduling options may be examined.

Numbers and durations of theatre sessions are defined in the model. These are likely to vary according to the day of the week. A number of theatre scheduling rules have been incorporated to meet expressed user requirements. These are first come first served (FCFS), longest operation times first (LTF), shortest operation times first (STF) and longest time first followed by shortest first after a user-defined cut-off time (LTSC, also known as top-down bottom-up scheduling). Furthermore, the user can examine the effects of scheduling day-case patients first in any session before elective in-patient operations. Each speciality manager supplied the necessary information on numbers and duration of theatre sessions together with current scheduling rules (Appendix E).

6.2.5 Evaluating theatre capacity options

Hospital managers requested a number of key statistical indicators, tables and graphs from the theatre module. Examples include graphs of number of patients in theatre, operation time distributions and patient waiting times for theatre.

Session utilisation is a key measure of the theatre's workload. Session utilisation, which is used in the model, calculates the proportion of time that the theatre is busy.

$$\text{Session Utilisation} = \frac{\text{total time theatre in use during session}}{\text{total time theatre available during session}} * 100\%$$

Other key statistical indicators include the under and over-run time distributions. An over-run occurs when a patient's operation is not completed until after the official theatre session closing time. This will depend on theatre admission rules governing the latest time an operation may commence in light of the expected operation time and session end time. The model reflects these rules by incorporating a user-defined maximum over-run time permitted (as a percentage of the total session time) and by considering the patient's expected operation time as sampled from the appropriate statistical distribution. Under-runs (when a theatre closes early) can be avoided by careful scheduling of operations. Various theatre session and patient scheduling options may be considered and their impact on theatre utilisation, under-run and over-run times examined.

6.2.6 Workforce planning module

Quantifying hospital workforce needs over time, such as the required number of nurses by grade, is a non-trivial exercise. Workforce needs are influenced by many different factors, such as patient case-mix and patient demand. Using the bed model as a basis for simulating patients passing through the hospital (arrival, LoS, discharge), an add-on workforce module for PROMPT enables the user to further evaluate workforce needs over time (down to monthly, daily and shift needs). During their LoS, patients might pass through different stages of care-needs. Initially, for example, a patient may be classed "high dependency" and remain in this state for a day. Then they might move into an "Intermediate dependency" stage for the next 48 hours, before moving onto "low dependency" for the remainder of their stay. Care-needs will obviously differ between each dependency stage, as one would expect intense care for high dependency patients, with the level of care gradually becoming less intense during their stay (Figure 6.3).

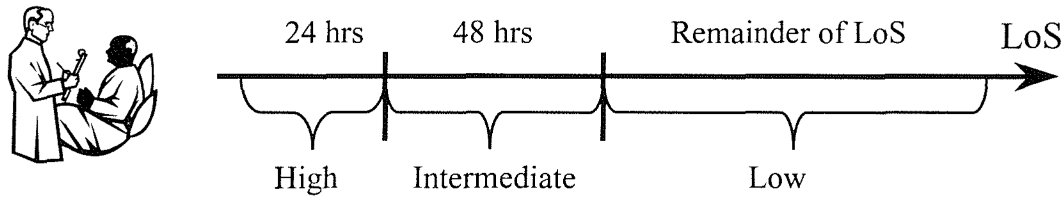


Figure 6.3: Illustrative patient nursing dependency states

Workforce requirements will depend on the following variables:

- Number of dependency states a patient experiences.
- How long a patient stays within each of these states.
- Patient-to-Resource dependencies (or ratios) for each state (see below).

The model allows any workforce resources to be defined (for example, Grade A nurse, Grade B nurse, Skilled Nurse, Untrained Nurse, Porter etc.). For each patient group in the model, necessary workforce needs are structured for the patient during their stay in hospital. The participating Trusts agreed that these needs should be provided in the form of patient-to-resource dependencies (or ratios). For example, a ratio of 2 for a grade A nurse indicates that a grade A nurse can care for 2 patients from this group who stay in hospital - or equivalently, one patient requires 0.5 grade A nurses. These dependencies are provided for each workforce resource.

The user may create as many dependency states as necessary, together with the expected percentage of time that the patient will stay in that state (as a percentage of their expected total LoS). For each state, dependency ratios may then be entered against each resource. Furthermore, the ratios should be defined for each hospital shift - early, late and night. This reflects the varying care-needs across the day. During the night, for example, whilst a patient sleeps, the necessary level of care may be less than when the patient is awake during the daytime.

The workforce module calculates likely workforce needs (rosters) over time for each specialty bed-pool or hospital ward. The resulting needs are displayed by resource and further broken down by month, day of week and shift each day. Because the module is linked to the bed capacity model, and operating theatre module if so desired, the consequences of changes to other hospital capacities, such as the number of beds and duration of theatre sessions, will be reflected in the workforce rosters. This integrated and granular approach is essential to reflect real-life complex patient flows through a hospital and avoids many of the failings of existing approaches as discussed in the literature review (section 3.4) and furthermore meets the expressed needs of a sufficiently detailed and flexible hospital capacity model (section 4.3.1).

6.2.7 Data requirements

Access was given to the hospital's patient management system. This large database contained information on individual patients, including episode admission date, episode discharge date, LoS, operation time, specialty, admission method (emergency or elective) and management intent (day-case or inpatient). Apollo is used to create statistically and clinically meaningful patient groups and to obtain information about particular flows over time. Apollo can link with most databases that are used in hospitals and extract the necessary data for the statistical analysis.

6.2.8 Novel features

The PROMPT model has been designed with the needs of the hospital and review of existing literature in mind. The review (Chapter 3) highlighted the need for a sophisticated dynamic hospital resource model and PROMPT has been evolved to avoid a number of issues that have made previous models essentially redundant in a real-life setting. Some of the novel features of this work, which help provide its place in literature include:

- The stochastic nature recognises and incorporates the complexity of hospital dynamics. The model uses statistical distributions to help capture the large amount of variability, for example in length of stays and operation times. Arrival profiles help to mimic complex monthly, daily and hourly arrival patterns.

- The flexible structure permits the model to be fine tuned to reflect local conditions, which can be used by a variety of hospitals. This avoids the development of hospital or ward specific models. For example, the concept of a *Care Unit* was derived alongside hospital managers to represent any configuration of hospital care, such as a specialty bed-pool, day-case unit or ward. Many different patient groups may be defined, enabling a variety of different hospital case-mixes to be captured. Virtually all hospital parameters may be changed to reflect current conditions and for ease of use in scenario modelling by healthcare managers.

- An integrated approach helps to capture the wider, more global picture of healthcare provision. For example, PROMPT models the use of beds, theatres and human resources and is unique in examining all three major hospital resource components in a single simulation tool. Many other models suffer from an isolated approach in that they ignore the complex integrated system and instead draw conclusions at a localised level. Clearly this has the potential to provide misleading and unrealistic conclusions.

- PROMPT has been designed for ease of use in mind. Practical models should be sufficiently accessible for end-users who are not experts in simulation, such as healthcare managers and clinicians. Ease of use has been achieved through development in a Windows environment (Delphi) and through the evolutionary development methodology (section 4.2), which created prototype models alongside end-users.

- The use of Apollo for creating patient groups and automatically feeding them into the simulation model is novel. This aims to provide the user with an easy to use interface and to help in the creation of statistically and clinically meaningful groups.

- Detailed hospital resource models should aid managers with both planning and management issues. PROMPT has been developed with these needs in mind. For example, it may be used to examine the consequences of different daily theatre schedules (management decision) or the total bed needs for the forthcoming financial year (planning issue).

Together, the above features help to demonstrate that Apollo and PROMPT have a unique place in healthcare modelling.

6.3 Simulation Structure

A patient-flow through the hospital is assumed to consist of a finite number of states. For example, the LoS in the hospital bed or the operation time in the theatre. Patients have to queue before making the transition between states. For example, queueing for a bed to become free or queueing for a theatre session. The simulation model follows individuals as they progress through the hospital system, from the first time they arrive in an attempt to obtain a bed until they leave the hospital having completed their stay.

In order to simulate the passage through the hospital, information on dwelling times in the various states and transition probabilities are required. In other words a semi-Markov structure has been chosen to model patient-flows (Appendix B). A number of distribution functions are available for use in the program to sample transition times: Lognormal, Weibull, Normal, Exponential and Gamma. The Weibull distribution is frequently used when data availability is limited (see section 6.4).

PROMPT was developed in a Delphi environment using a three-phase simulation shell TOCHSIM (Appendix G). The three-phase approach consists of two different events: *Bound* (B) events and *Conditional* (C) events. The necessary events for the PROMPT structure are:

- ***Bound events***
  - Generate initial arrival times and cause next arrivals within each patient group.
  - Check, and if necessary adjust, bed numbers to reflect step-up and step-down of beds over time (daily and monthly event).
  - Open theatre sessions as necessary (daily event).
  - Cause post-operation LoS to commence when patient arrives back from theatre to ward.
  - Cause patient to leave hospital on completion of their LoS or cause re-admission arrival time for deferred elective patients.

- ***Conditional events***
  - Start patient stay (having already found an available and suitable hospital bed).
  - Start operation in theatre (having queued and been admitted in to a suitably available and open theatre session).

Appendix G provides further information on the TOCHSIM shell with pseudo-code for the PROMPT model.

An object-orientated programming (OOP) approach was adopted within the three-phase structure. OOP is a natural evolution from earlier innovations to programming language design (for example see Hirata and Paul, 1996). It is more structured and more modular than previous approaches. Three main properties characterise an OOP language:

- *Encapsulation* – combining a record with the procedures and functions that manipulate it to form a new data type, called an *object*.
- *Inheritance* – defining an object and then using it to build a hierarchy of descendant objects, with each descendant inheriting access to all its ancestors' code and data.
- *Polymorphism* – giving an action one name that is shared up and down an object hierarchy, with each object in the hierarchy implementing the action in a way appropriate to itself.

TOCHSIM with Objects has benefited from the language extensions to give the full power of OOP: more structure and modularity, more abstraction and reusability built into the simulation shell.

An object type is a structure consisting of a fixed number of components. Each component is either a *field*, which contains data or a particular type, or a *method*, which performs an operation on the object. Similar to a variable declaration, the declaration of a field specifies the field's data type and an identifier that names the field. Similar to a procedure or function declaration, the declaration of a method specifies a procedure, function, constructor, or destructor heading.

An object type can *inherit* components from another object type. If T2 inherits from T1, then T2 is a descendant of T1, and T1 is an *ancestor* of T2.

Inheritance is transitive in nature. If T3 inherits from T2, and T2 inherits from T1, then T3 also inherits from T1. The domain of an object type consists of itself and all its descendants.

The inherited object is the generic simulation object used by the following objects in TOCHSIM: queues, resources, all statistic objects, the simulation timer, and the simulation model.

6.4 Probability Distributions for LoS and Operating Times

The necessary distributions for use within PROMPT, namely LoS and operation times, can be represented by a number of statistical distributions. Within the Apollo statistical add-on package, distributions can be fitted and parameters passed across for use in the simulation. Individual patient LoS and operation times may then be sampled from the appropriate distributions within the model. This automated process is particularly useful for end-users. An illustrative Apollo patient group fitted LoS distribution is shown in Figure 6.4.

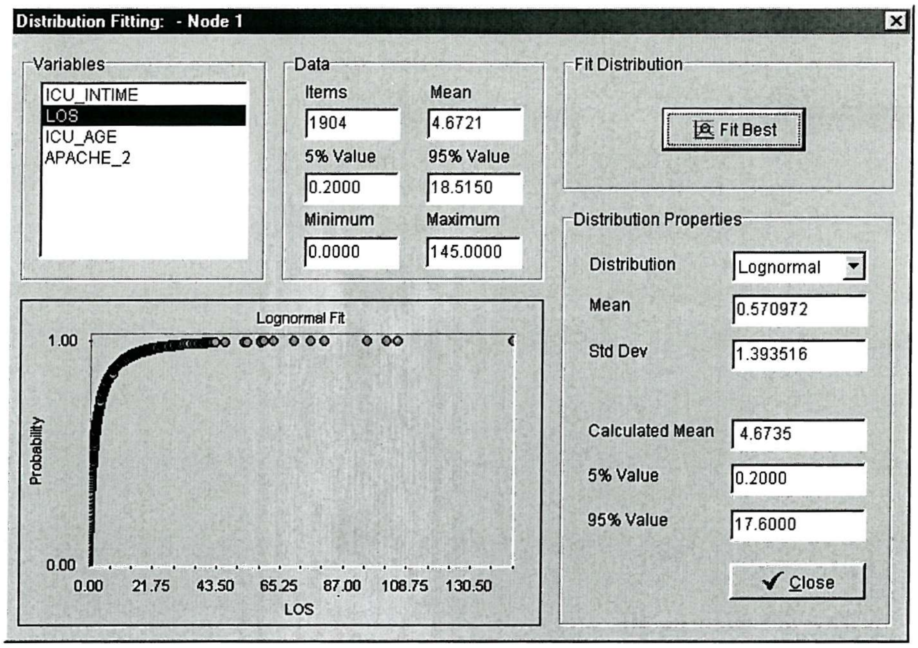


Figure 6.4: Illustrative fitted LoS distribution (Apollo screen-shot)

Complex hospital models must deal with circumstances when LoS and/or theatre data is absent or limited. When detailed data is not available, information through aggregated data or through expert opinion, such as a likely mean value and/or percentage points, may be the only source of input for the model. Rapid choice of distributions and the use of different distributions in a sensitivity analysis are needed in practical simulation modelling work.

Typically a choice is made from a number of standard statistical deviates, such as Normal, Exponential, Weibull, Gamma or Lognormal. In the absence of data, Weibull variates are particularly useful.

We discuss a method for estimating the parameters from the Weibull distribution function. This method has been used within PROMPT and allows the user to rapidly fit LoS and operation time distributions. Currently Normal, Lognormal, Weibull, Negative Exponential and Gamma have been implemented (Appendix F), although other variates, including discrete distribution, can be easily included if necessary.

6.4.1 Point estimation for the Weibull distribution

Several methods of point estimation for Weibull parameters are given by Orman (1995). For a density $f(x)$ the maximum likelihood function of a sample n $(x_1, \ldots, x_n)$ is given by:

$$L(x_1,\ldots,x_n)=\prod_{i=1}^{n} f(x_i)$$

The maximum likelihood estimates of the parameters α, β can be obtained by solving:

$$\tilde{\alpha}^{\tilde{\beta}} = \frac{1}{n}\sum_{i=1}^{n} x_i^{\tilde{\beta}}$$

and

$$\frac{1}{\tilde{\beta}} = \frac{\sum_{i=1}^{n} x_i^{\tilde{\beta}} \ln x_i}{\sum_{i=1}^{n} x_i^{\tilde{\beta}}} - \frac{1}{m}\sum_{i=1}^{n} \ln x_i$$

The Newton-Raphson method can be used to estimate the parameters. The maximum likelihood method is the preferred approach when data is available. In the absence of data, Weibull variates are particularly useful and point estimation is possible with a mean value and/or percentile points as described below.

6.4.2 Point estimation for the Weibull distribution given mean and one percentage point

With ease of use in mind, Orman (1995) gives an algorithm that calculates an estimate of the parameters of the Weibull distribution from a defined mean value and one percentage point.

The user defines a mean value m and a percentage value for which $F(x) = p$. The user provides the value of p.

With $q = 1 - p$, the starting values of the two parameters are

$$\alpha_0 = m \qquad \text{and} \qquad \beta_0 = \frac{\ln(-\ln q)}{\ln(x/\alpha_0)}$$

At the next iteration,

$$\alpha_1 = \frac{m}{\Gamma(1+1/\beta_0)} \qquad \text{and} \qquad \beta_1 = \frac{\ln(-\ln q)}{\ln(x/\alpha_1)}$$

The ith iteration, $i \geq 0$, gives

$$\alpha_i = \frac{m}{\Gamma(1+1/\beta_{i-1})} \qquad \text{and} \qquad \beta_i = \frac{\ln(-\ln q)}{\ln(x/\alpha_{i-1})}$$

Care must be exercised for percentiles in the neighbourhood of 63% where the algorithm does not always converge. This feature is the result of the mathematical nature of the probability density function of the Weibull variate (Shahani *et al.*, 1994) since $p \cong 0.63 \cong 1 - 1/e$ and leads to the condition:

$$\beta \ln\left(\frac{t}{\alpha}\right) = 0$$

which shows that $\alpha = t$, $\forall\ \beta$. Thus this value of p will not discriminate between different values of β and so the range around this value should be avoided when trying to solve for β.

6.4.3 Point estimation for the Weibull distribution given two percentage points

Dubey (1967) describes a method for estimating the parameters α and β of the Weibull distribution using two percentages from a sample set of data points. The information given corresponds to the equations:

$$F(x_1) = p_1, \qquad F(x_2) = p_2,$$

The parameters can be obtained by solving these two equations to get:

$$\beta = \frac{\ln[-\ln(1-p_1)] - \ln[-\ln(1-p_2)]}{\ln(x_1) - \ln(x_2)}$$

and

$$\alpha = \exp[|W\ln(x_1) + (1-W)\ln(x_2)|]$$

where

$$W = 1 - \frac{\ln(-\ln(1-p_1)}{\ln[-\ln(1-p_1)] - \ln[-\ln(1-p_2)]}$$

6.4.4 Algorithm AS47

A simplex optimising algorithm (Nelder, 1965 and O'Neil, 1971) has been selected for estimating the parameters of a chosen distribution. The parameters are estimated by minimising a helpful second degree function, the Chi-square value. The simplex optimising algorithm was chosen as it is computationally easy to implement and use of the Chi-squared value is asymptotically equivalent to the maximum likelihood function (Jones, 1997).

The user may define a mean and/or one or more percentile values for Weibull, Gamma, Lognormal, Normal and Exponential. Other distributions may be added if necessary. The AS47 algorithm for solving the two dimensional problem in which a function $f(x)$ is minimised with respect to two parameter values is described below. The simplex routine minimises the Chi-square value and converges to provide a solution.

Algorithm AS47 is built into both the Apollo and PROMPT models to enable the user to fit the necessary statistical distributions. This may be an automated process within Apollo so that the chosen distribution and parameters are fed into PROMPT, or within the PROMPT model itself to fit a distribution with a mean and/or percentile values only. Either way, it permits the simulation model to reflect the inherent variability amongst the population of patients who stay at the hospital and thus avoids the dangers of using deterministic average values only.

Algorithm AS47

STEP 1. Start with three initial points x_1, x_2, x_3 where $x_i = (P_1(i), P_2(i))$ and P_i represents the parameter i.

STEP 2. Calculate the function value at these points $y_i = f(x_i)$.

STEP 3. Find the reflection, x_R, through the centroid x_C, of the point with the maximum function value, say x_1.

STEP 4. If the reflection value, y_R, is minimum then expand to x_E. The minimum of y_R and y_E determines which point replaces x_{max}.

STEP 5. If the reflection value, y_R, is maximum then contract to x_A or x_B, depending on whether y_2 or y_3 is lower, and replace x_R.

STEP 6. The iteration, steps 3 to 5, continues with the simplex distorting in shape according to the slope it encounters, until the variance of the values y_i are less than a determined level.

6.5 Capturing Patient Demand Profiles

Alongside the necessary statistical distributions, additional information is required to allow PROMPT to reflect observed (or other) patient demand patterns. An automated process within the Apollo statistical module captures the hourly, daily and monthly arrival rates which are fed directly into the PROMPT simulation model. An illustrative Apollo patient group monthly demand profile is captured in Figure 6.5.

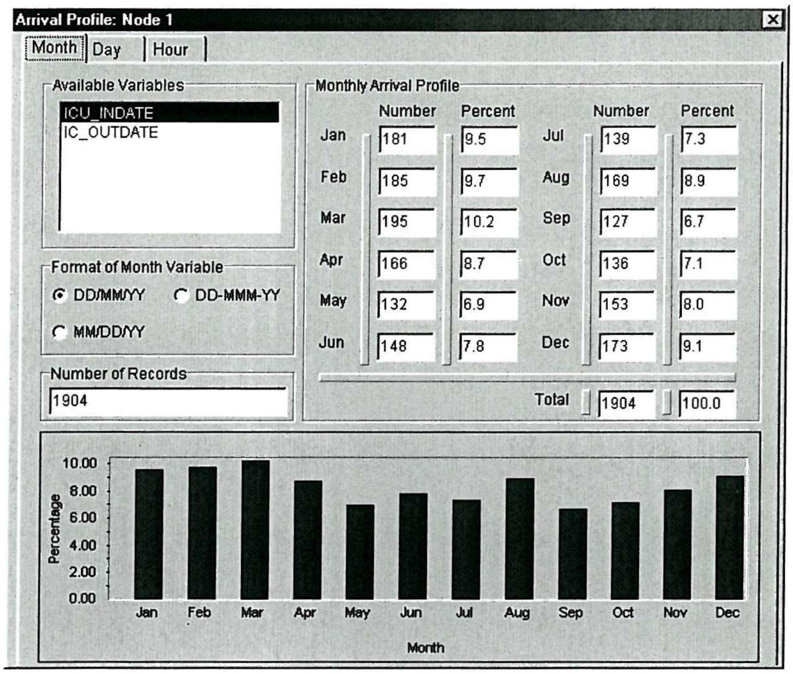


Figure 6.5: Illustrative monthly patient demand (Apollo screen-shot)

In the model observed (or other) arrival patterns are entered for each month, day and four-hour time block within each day. This is necessary to reflect in detail fluctuating demand over time. Expected annual total numbers of patients are provided for each group. Individual patient group arrival profile numbers need not sum to the corresponding annual total but instead represent a profile of relative demands. For example, if the total number of annual patients was six hundred and every month had a monthly profile value of one, then we would expect the same demand of fifty patients per month across the year. If January had a value of two and every other month a value of one, then we would now expect double the demand in January relative to the other months (92 patients in January and 46 every other month). This methodology has been adopted since it is likely that hospitals will use a number of previous years demand profiles but wish to change the total number of predicted patients for scenario modelling whilst preserving the observed demand structure.

In general let m_i be the demand profile entered for month i ($i = 1,\ldots,12$), d_j for day j ($j = 1,..,7$) and h_k for hour-block k ($k = 1,..,6$). Furthermore let n be the annual total number of patients for a particular patient group.

Inter-arrival times for arrivals from within each patient group are sampled from the Negative Exponential distribution. The expected number of arrivals, $\lambda_{i,j,k}$ for any given month i, day j and four hour-block k is given by:

$$\lambda_{i,j,k} = \left(\frac{nm_i \big/ \sum_{i=1}^{12} m_i}{y_i} \right) \left(\frac{7d_j}{\sum_{j=1}^{7} d_j} \right) \left(\frac{h_k}{\sum_{k=1}^{6} h_k} \right)$$

where y_i indicates the number of days in month i (28, 29, 30 or 31).

The probability of r arrivals within a given four-hour block may be calculated by using the Poisson distribution, such that:

$$P(R = r) = \frac{\lambda_{i,j,k}^{r} e^{-\lambda_{i,j,k}}}{r!} \qquad r = 0, 1, 2, \ldots$$

6.6 Model Validation and Verification

Throughout the research work constant validation and verification was conducted in order to increase confidence in the model operations. Essentially verification ensures that you are solving the problem correctly and validation ensures that you are solving the correct problem. Validation was largely achieved through the use of the generic framework and the adopted evolutionary model development. This ensured that end-users contributed to the model development at all stages and ensured that the simulation model reflected the original conceptual schema.

For model verification, a range of techniques was used including statistical comparisons of simulated output against real data. The following summary lists the validation and verification techniques that have been adopted. These techniques are described in Sargent (1991).

***Animation*:** The model's operational behaviour is displayed graphically as the model moves through time. An optional run-time graphical display shows the current simulation time, numbers of patients stayed, numbers of refusals and occupancy rate for each care-unit within the hospital. This is particularly useful to verify that the simulation behaviour (e.g. numbers of arrivals and beds in use each day) matches real-life behaviour. Graphical displays, such as the one developed in PROMPT, are becoming increasingly popular amongst simulation end-users. Graphics provide more confidence as the user can see the processes on screen which prevents a "black-box" approach. Graphics however can considerably slow down run-time and so the user may turn off the graphical display at any time during the simulation run. Other graphics available on completion of the run include plots of patient arrivals over time. Such graphs also permit model verification. Figure 6.6 shows the graphical display during a simulation run of the PROMPT model.

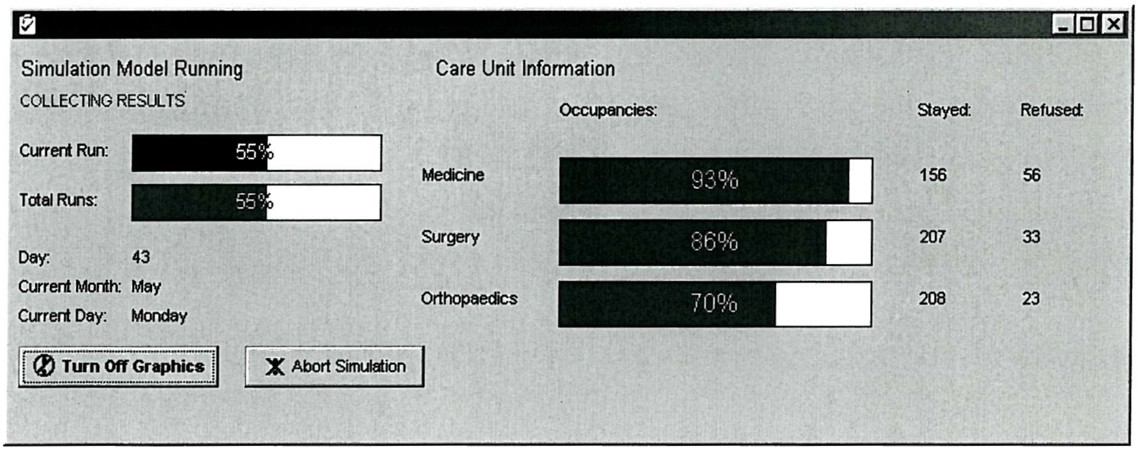


Figure 6.6: Animation within PROMPT

***Comparison to other models*:** Various simulation results may be validated by comparison to other models. As an example simpler cases of the simulation model may be compared to known results of an analytical model. Validation of PROMPT has included the comparison of the developed simulation model against some queueing models. Further details may be found in section 6.7.

***Degenerate tests*:** The degeneracy of the model's behaviour is tested by an appropriate selection of values for the input and internal parameters. For example, running the simulation with arrivals only on a Monday but an operating theatre session only on Thursday. Suitable checks are made to ensure that all patients are waiting until Thursday for their operation and then returning to ward for re-commencing their LoS.

***Extreme-condition tests*:** The model structure should be plausible for any extreme and unlikely combination of levels of factors in the system. As examples: providing no beds in the hospitals for a 100% refusal rate; no theatre sessions so that operations are not permitted and patients can wait all year in bed; no arrivals and so 0% care-unit occupancy.

***Historical data validation*:** In the presence of historical data, part of this is used to build the model and the remaining data used to determine if the model behaves as the system does.

***Internal validity*:** Several replications of a stochastic model are made to determine the amount of variability in the model. Inconsistency in the results should cause concern over the model's validity. Graphical plots are used in PROMPT to display standard deviation from the multiple simulation runs that can be used to gauge the run-to-run simulation variability.

***Traces*:** The behaviour of different types of specific entities in the model are traced through the model to determine if the model's logic is correct. Examples include tracing a patient through PROMPT to ensure that they follow the correct care-pathway and receive the correct resources (theatre and nursing) as necessary.

***Parameter variability – Sensitivity analysis*:** This validation technique consists of changing the values of the input and internal parameters of a model to determine the effect upon the model behaviour and output. For example: shortening LoS should reduce occupancy and refusals; examining the relationship between numbers of beds, occupancy and refusal rates; sensitivity of predicted demand on Trust's performance. Sensitivity analysis, and in particular the design of experiments, is a major topic of research. Currently there is a great need for more research in this area (Cheng and Holland, 1997, Cheng and Kleijnen, 1999 and Cheng and Lamb, 2000).

6.7 Comparison of Queueing Models and Developed Simulation

As a necessary part of the model validation process, the developed simulation has been compared to analytical queueing models. As the complexity of the healthcare system increases, correspondingly analytical models become harder to formulate and solve. Thus only two queueing models have been built for illustrative validation purposes. A simulation model enables the programmer to incorporate far more complexity, variability and uncertainty within the stochastic framework. Analytical models fail to capture many of the necessary processes, but nevertheless permit validation for a number of simplistic scenarios, as detailed below.

6.7.1 Queueing models

A number of queueing models, adapted from work by Cohen (1956), Saaty (1961) and Erlang (1917), were considered. A description of each model and the relevant steady-state solutions are provided. The corresponding simulation runs are shown alongside for validation. Example data values (numbers of patients and LoS) have been taken from the Royal Berkshire and Battle NHS Trust's database.

Key to queueing models

s : Number of beds (fixed)
λ_i : Arrival rate of patient type i
μ_i : Service rate (1/LoS) of patient type i
ρ : Traffic intensity (λ/μ)
LQ_i : Average length of the i'th patient queue (number of patients)
WQ_i : Average waiting time on the i'th patient queue (days)
W_i : Average total time spent in hospital by i'th patient type (days)
π_j : The steady-state, or equilibrium probability, of state j

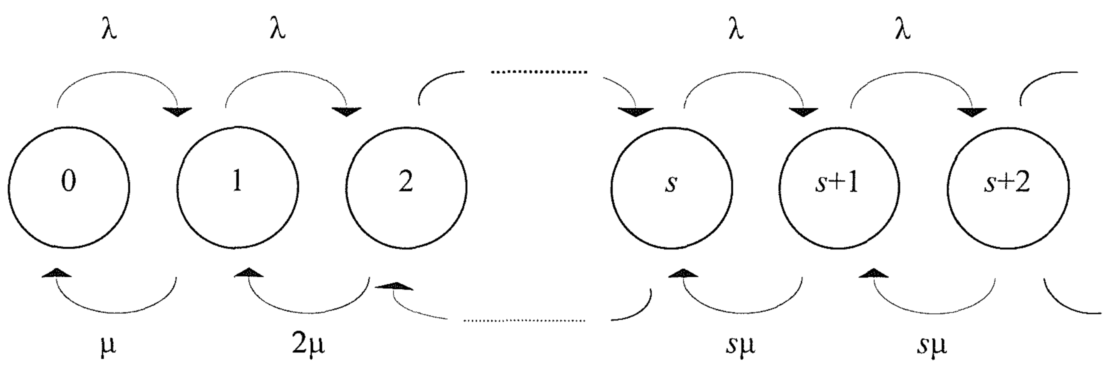
6.7.2 $M / M / s / GD / \infty / \infty$ queueing model

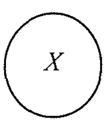
Model description

- All patient admissions to a single care-unit taken from a single queue, Q_a.
- All patients have a mean LoS of 6.2 days.
- Patients are referred at a rate of 1,100 per year.
- LoS and inter-arrival times are assumed to have Negative-Exponential distributions.
- There is an infinite population of patients.
- There is no limit on the length of the arrival queue.
- Arrivals are taken from the queue on a first come first served basis (FCFS).

Analytical solution

The model state transitions are shown in the transition diagram of Figure 6.7.





represents the state in which there are X patients in the system (care-unit)

Figure 6.7: Transition diagram for $M / M / s / GD / \infty / \infty$ queueing model

If j servers (beds) are occupied, service completions occur at a rate $\underbrace{\mu + \mu + \mu + \ldots}_{j} = j\mu$

Thus:

$$\lambda_j = \lambda \qquad (j = 0,\ldots,\infty)$$

$$\mu_j = j\mu \qquad (j = 0,\ldots,s)$$

$$\mu_j = s\mu \qquad (j = s+1,\ldots,\infty)$$

Substituting these expressions into the following basic queueing theory formulae,

$$\pi_j = \pi_0 c_j \qquad \text{where} \quad c_j = \frac{\lambda_0 \lambda_1 \lambda_2 \ldots \lambda_{j-1}}{\mu_1 \mu_2 \mu_3 \ldots \mu_j}$$

gives:

$$\rho = \frac{\lambda}{s\mu}, \qquad \pi_0 = \frac{1}{\left(\sum_{i=0}^{s-1} \frac{(s\rho)^i}{i!}\right) + \frac{(s\rho)^s}{s!(1-\rho)}}$$

$$\pi_j = \frac{(s\rho)^j \pi_0}{j!} \quad (j = 1,\ldots,s), \qquad \pi_j = \frac{(s\rho)^j \pi_0}{s! s^{j-1}} \quad (j = s+1,\ldots,\infty)$$

$$P(X \geq s) = \frac{(s\rho)^s \pi_0}{s!(1-\rho)}, \qquad LQ_a = \frac{P(X \geq s)\rho}{1-\rho},$$

$$WQ_a = \frac{LQ_a}{\lambda} = \frac{P(X \geq s)}{s\mu - \lambda}, \qquad W_a = WQ_a + \frac{1}{\mu} = \frac{P(X \geq s)}{s\mu - \lambda} + \frac{1}{\mu}$$

Illustrative queueing model results

Table 6.1: Illustrative results from a $M / M / s / GD / \infty / \infty$ queueing model

s	ρ	$P(X \geq s)$	LQ_a	WQ_a	W_a
20	0.94	0.69	10.09	3.35	9.56
21	0.89	0.51	4.17	1.38	7.59
22	0.85	0.37	2.09	0.69	6.90
23	0.81	0.26	1.13	0.37	6.58
24	0.78	0.18	0.63	0.21	6.42
25	0.75	0.12	0.36	0.12	6.37

Figure 6.8 shows the complex relationship between beds and a number of hospital system indicators.

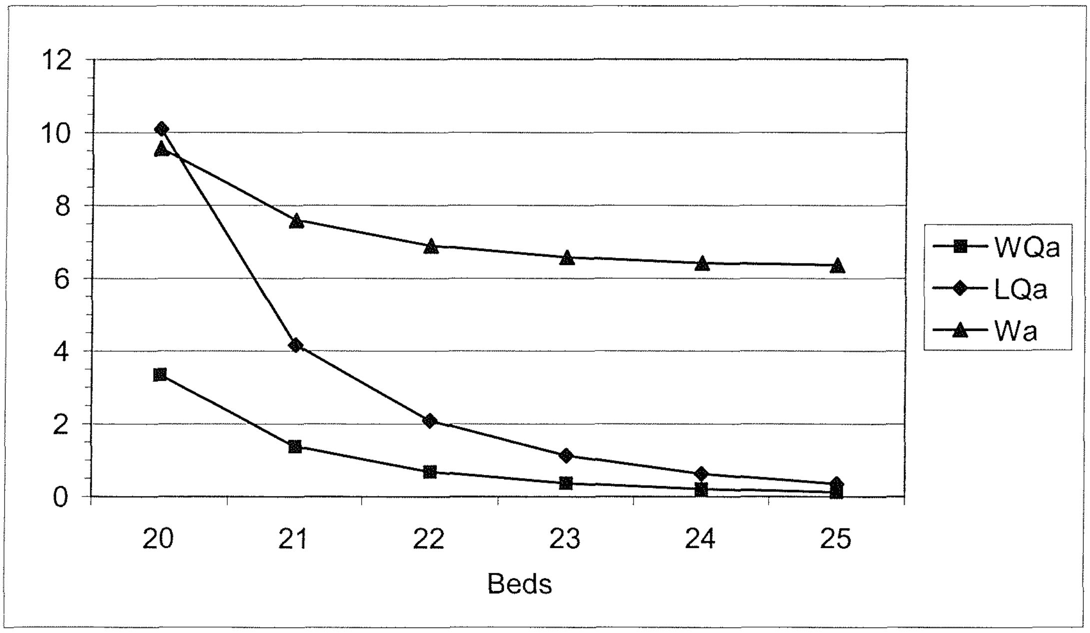


Figure 6.8: Relationship between number of beds, queue waiting time, queue length, and time in hospital using $M / M / s / GD / \infty / \infty$ model

Simulation model

The developed simulation model was run under the same assumptions as for the queueing model. Figure 6.9 shows the corresponding simulation activity flow diagram.

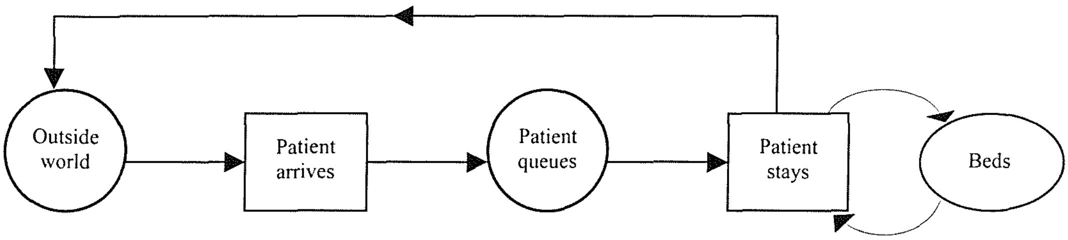


Figure 6.9: Activity flow diagram for $M / M / s / GD / \infty / \infty$ simulation model

Model comparisons

The simulation was run for one-year with a five-year warm-up and for a total of fifty runs.

Table 6.2: Comparison of models for a $M / M / s / GD / \infty / \infty$ queue

Model	s	LQ_a	WQ_a	W_a
Queueing	25	0.35	0.12	6.37
Simulation	25	0.36	0.11	6.36

Although the assumptions are very limiting, this model permits a first approximation to real-life and is useful for validating the simulation. It becomes possible, and with increased confidence, to progress on to more complex and realistic queueing models for validation purposes.

6.7.3 $M_i / M / s / NPRP / \infty / \infty$ queueing model

Model description

- All patient admissions to a single care-unit taken from two queues, Q_e for emergency patients and Q_p for planned (elective) patients.
- All patients have a mean LoS of 6.2 days.
- Emergency patients are referred at a rate of 850 per year.
- Planned patients are referred at a rate of 250 per year.
- LoS and inter-arrival times are assumed to have Negative-Exponential distributions.
- There is an infinite population of patients.
- There is no limit on the length of the arrival queue.
- Emergency patients always have priority over planned patients when both queues are not empty; arrivals are then taken from the queue with priority on a FCFS basis.

Analytical solution

Let (x) be the state in which x beds $(0 \le x \le s)$ are occupied, $f(x)$ be the probability of state (x) and h be the average LoS. Cohen (1956) can be adapted for this model to produce the following helpful functions:

$$f(x) = \left(\frac{s - \lambda_e - \lambda_p}{s - \lambda_e - \lambda_p + (\lambda_e + \lambda_p)E(s, \lambda_e + \lambda_p)} \right) \left(\frac{(\lambda_e + \lambda_p)^x}{x!N(s, \lambda_e + \lambda_p)} \right)$$

where

$$N(a,b) = \sum_{i=0}^{a} b^i / i! \qquad \text{and} \qquad E(a,b) = \frac{b^a}{a!N(a,b)}$$

$$P(X \ge s) = \frac{sE(s, \lambda_e + \lambda_p)}{s - \lambda_e - \lambda_p + (\lambda_e + \lambda_p)E(s, \lambda_e + \lambda_p)}$$

$$WQ_e = \frac{hP(X \geq s)}{s - \lambda_e}, \quad WQ_p = \frac{shP(X \geq s)}{(s - \lambda_e)(s - \lambda_e - \lambda_p)}$$

$$LQ_e = \frac{\lambda_e P(X \geq s)}{s - \lambda_e}, \quad LQ_p = \frac{s\lambda_p P(X \geq s)}{(s - \lambda_e)(s - \lambda_e - \lambda_p)}$$

Model results

Table 6.3: Illustrative results from a $M_i / M / s / NPRP / \infty / \infty$ queueing model

s	ρ	$P(X \geq s)$	LQ_e	LQ_p	WQ_e	WQ_p	W_e	W_p
20	0.94	0.69	1.81	8.28	0.78	12.10	6.99	18.31
21	0.89	0.51	1.13	3.04	0.48	4.45	6.69	10.66
22	0.85	0.37	0.70	1.38	0.30	2.02	6.51	8.23
23	0.81	0.26	0.44	0.69	0.19	1.01	6.40	7.22
24	0.78	0.18	0.27	0.36	0.12	0.52	6.33	6.73
25	0.75	0.12	0.16	0.19	0.07	0.28	6.28	6.49

Simulation model

The developed simulation model was run under the same assumptions for the queueing model. Figure 6.10 shows the corresponding simulation activity flow diagram.

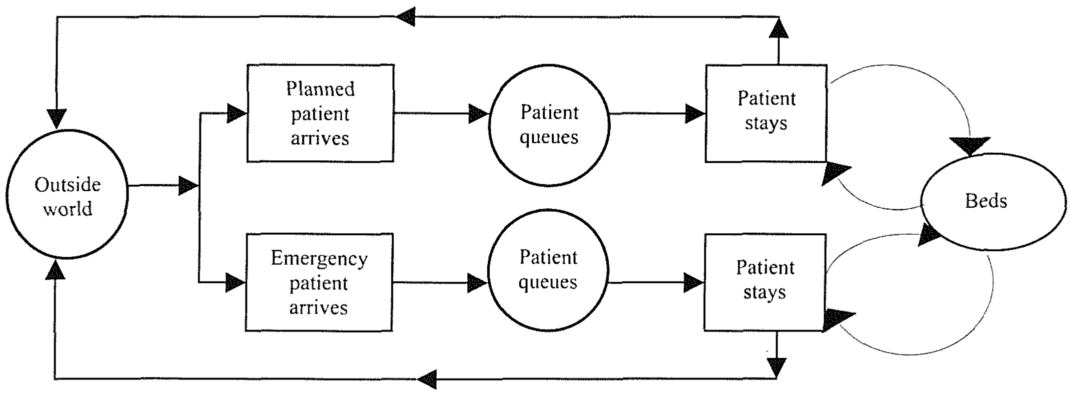


Figure 6.10: Activity flow diagram for $M_i / M / s / NPRP / \infty / \infty$ simulation model

Model comparisons

The simulation was run for one-year with a five-year warm-up and for fifty runs.

Table 6.4: Comparison of models for a M_i / M / s / $NPRP$ / ∞ / ∞ queue

Model	s	LQ_e	LQ_p	WQ_e	WQ_p	W_e	W_p
Queueing	25	0.16	0.19	0.07	0.28	6.28	6.49
Simulation	25	0.16	0.20	0.08	0.23	6.29	6.44

As part of the model validation process, the developed simulation has been compared to two analytical queueing models. Consecutive runs of the PROMPT model have demonstrated excellent reproductively of the analytical solutions and has imparted the necessary confidence for model use. Slight discrepancies, particularly with planned patient queues, arise because the PROMPT model permits an unconstrained length on this queue. Further validation against a number of more complex queueing models would be desirable. Analytical models however become harder to formulate and solve when moving beyond the M_i / M / s / NPRP / ∞ / ∞ queueing model presented. The validation process has additionally granted an insight into necessary warm-up times and numbers of simulation runs required for PROMPT to reach a steady state. To avoid a significant warm-up bias, it was observed that a warm-up of at least three years was necessary, although to be sure it is proposed that five years should be used.

6.8 Designing a User-friendly Simulation

The adopted evolutionary model development approach (section 4.2) involved potential end-users from the outset of model development. Many of the users are likely to have limited experience or practice in using computer simulation models. This adopted process however enabled hospital personnel to play a major role in the "look and feel" of the final model. Delphi software enabled the simulation to be readily tailor-made for the hospitals within a familiar Windows environment. PROMPT has also been designed to link to spreadsheets and other Windows programs so that model outputs can be copied into other software. This results in a system for solving a wide range of healthcare problems.

Some illustrative PROMPT screen-shots are shown within subsequent sections of this chapter. These are divided into distinct model elements: PROMPT menu, the main screen and file management; how to create care units, patient groups, human resources and operating theatres; simulation parameters and outputs.

6.8.1 PROMPT menu

The menu screen forms the central control of the program, through which various options may be selected. Through this menu (Figure 6.11), the user may access (by clicking on the icons) either the Apollo statistical package or various PROMPT modules, namely: beds, beds and human resources, beds and operating theatres, or beds, human resources and operating theatres. These options may be activated within the model itself at a later time.

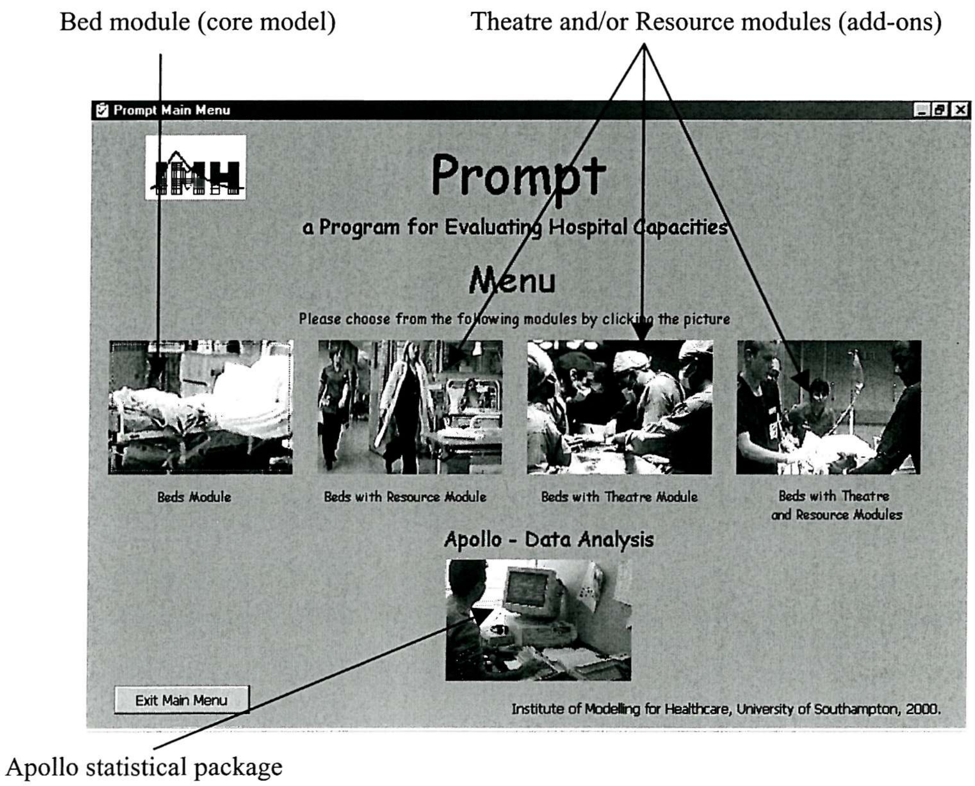


Figure 6.11: PROMPT menu screen

6.8.2 The main screen

Having selected the desired PROMPT module, the user will be presented with the main screen (Figure 6.12), which forms the central control in building a model of the healthcare system. The familiar Windows environment allows for easy selection of options and facilitates the construction of the necessary care-units and patient groups before running the simulation.

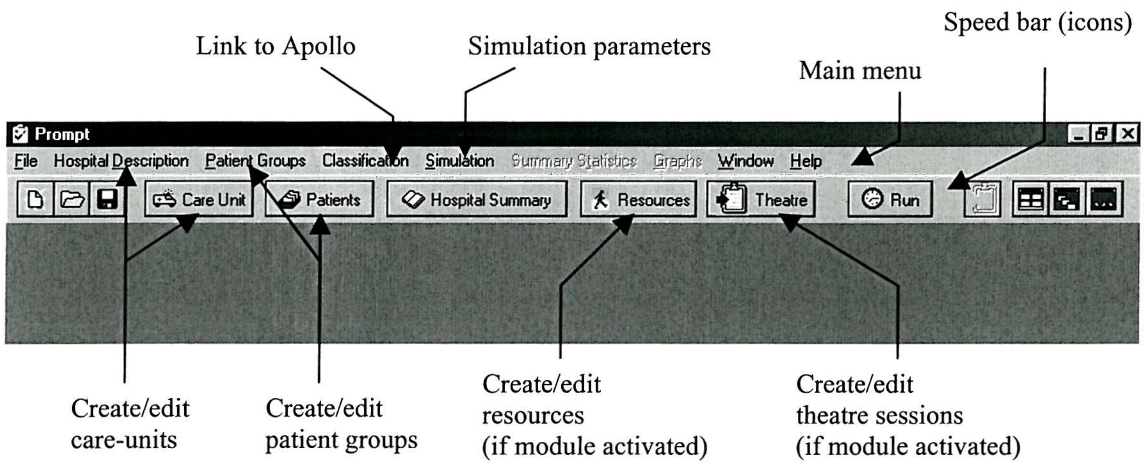


Figure 6.12: PROMPT main screen

6.8.3 File management

There are three main types of file that can be loaded and saved in PROMPT.

- ***Patient groups (*.pat)*** – Information on individual patient groups. The user can save a patient group in Apollo and directly load it into PROMPT. All patient group files have the extension .pat
- ***Care units (*.loc)*** – Information about each care unit (level of care) including bed numbers, resources and operating rules. All level of care files have the extension .loc
- ***Prompt files (*.prm)*** – Load and save all information (patient groups, care units, theatres and resources). This should be used once all parameters have been entered into the model. PROMPT files have the extension .prm

6.8.4 Care units

A care unit is defined to be any appropriate unit within the hospital where patients receive care. Typically this will be a speciality bed pool (e.g. General Medicine, ENT, Oncology) but may be an individual ward or the hospital as a whole (as one large care unit). The way in which a care unit is set-up by the user will depend on the nature of the study. For example, if we wish to plan bed numbers at the speciality level (e.g. 80 General Surgery beds, 50 Paediatric beds) then speciality bed-pool care-units should be defined.

One or more care units can be created depending on whether the user is interested in examining in detail one care-unit or the dynamics between care-units. This is achieved via the various patient groups in the model and their associated *care unit priority list.* Different patients in hospital require different care units. As an example, for a cardiac patient only a cardiac care unit bed may be appropriate, whereas for a general medicine patient, a general surgical bed may suffice on an occasion where no general medicine beds are available. Providing each care unit has been defined, any number of priority lists may be used. Use of beds, occupancy rates and outlier relationships can be studied (an outlier being a patient who couldn't be given a bed in their first choice care unit but is found a bed in another suitable care unit from their priority list).

The following variables are defined for each care unit:

- ***Bed numbers:*** user-defined bed numbers by month of year and day of week.
- ***Operating rules:***
  - *Permitted waiting times*: how long a patient may wait for until a bed becomes free.
  - *Operating rules for elective patients*: deferral time and number of deferrals allowed before patient is given emergency status
  - *Emergency-only beds*: the number of reserved beds for emergency patients only.

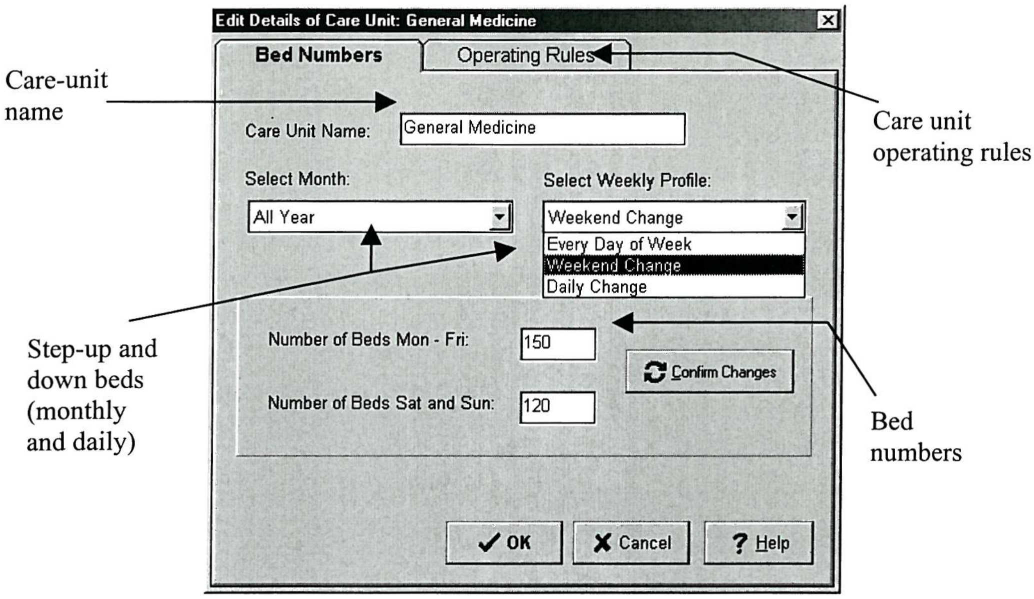


Figure 6.13: Care unit parameters

6.8.5 Patient groups

Fundamentally, there are two types of patient in the model:

- ***With procedure patients*** - those who require surgery.
- ***Without procedure patients*** - those who don't require surgery.

If the theatre module is turned off, then the user will only see details on without procedure patients. Furthermore, for each of these patient types, patients are either *emergency* or *elective*. An emergency patient always has priority in the model over elective patients (for example, if one bed is available and both an elective and emergency patient are waiting, then the bed would be given to the emergency patient).

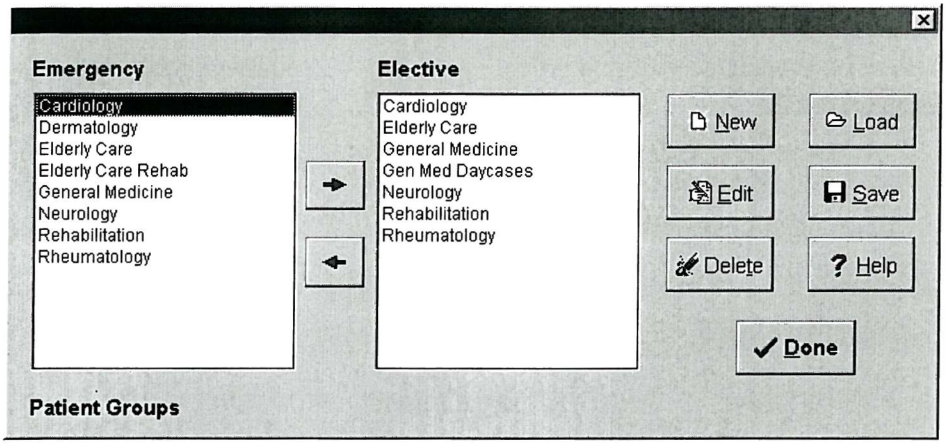


Figure 6.14: Illustrative list of patient groups

A patient group is described by:

- ***Patient group general:*** the group name, type and a text description.

- ***Length of stay (LoS):*** the LoS distribution governs the probability of the length of time a patient is likely to remain in the hospital bed.

- ***Operation time (with procedure patients only):*** the operation time distribution governs the probability of the time a patient is likely to spend in theatre.

- ***Arrival rate profiles:*** the arrival rate profiles describe how the patients arrive. The group's arrival pattern will depend on the month of the year, day of the week and potentially hour of the day. The yearly average arrival rate governs the number of patients that will be referred to the hospital for the time period being studied (typically one financial year).

- ***Deferral rules (elective patient types only):*** elective patients also have defined deferral rules (information used for when it becomes necessary to defer an elective patient), which may be at the patient level or care unit level. The rules incorporate *deferral time* (how long into the future before the patient should try again to obtain a bed) and *deferrals until emergency* (number of times a patient may be deferred before receiving emergency status with priority).

- ***Care details:*** each patient group is given a care unit priority list. When a patient is admitted, they will attempt to acquire a bed in the care unit(s) listed. One or more of the defined care units may be placed onto the priority list in the order of preference. In the case where no bed can be found from any of the listed care units, elective patients will be deferred and emergency patients transferred.

- ***Resource needs:*** if the resource module is turned on, the user will have access to this page. Further information may be found in section 6.8.6.

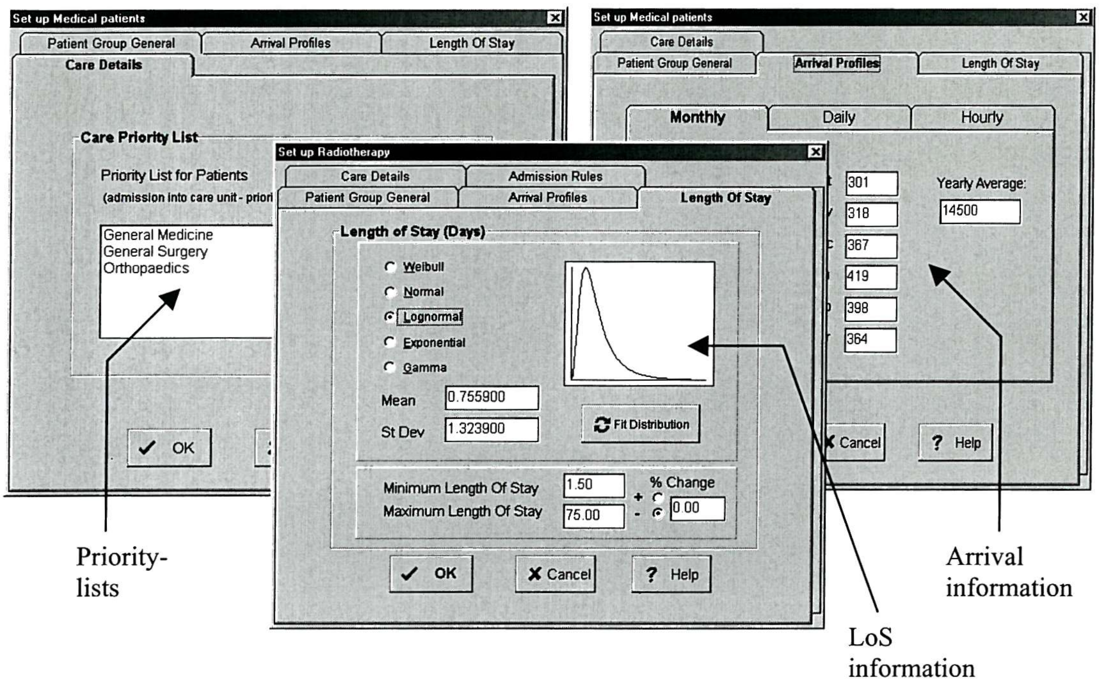


Figure 6.15: Patient group parameters

6.8.6 Human resources

For each patient group, the user may define the necessary resource needs for the patient during their stay in hospital. These needs are provided in the form of patient-to-resource ratios (or dependencies). These dependencies may be provided for each defined resource. Dependencies may also be defined for different stages during the patient's stay.

The user may create as many dependency states as necessary, together with the expected percentage of time that the patient will stay in that state (as a percentage of their expected total LoS). For each state, dependency ratios may then be entered against each resource. Furthermore, the ratios should be defined for each nurse shift shown - early, late and night. This reflects the varying care-needs across the day. During the night, for example, whilst a patient sleeps, the necessary level of care may be less than when the patient is awake during the daytime

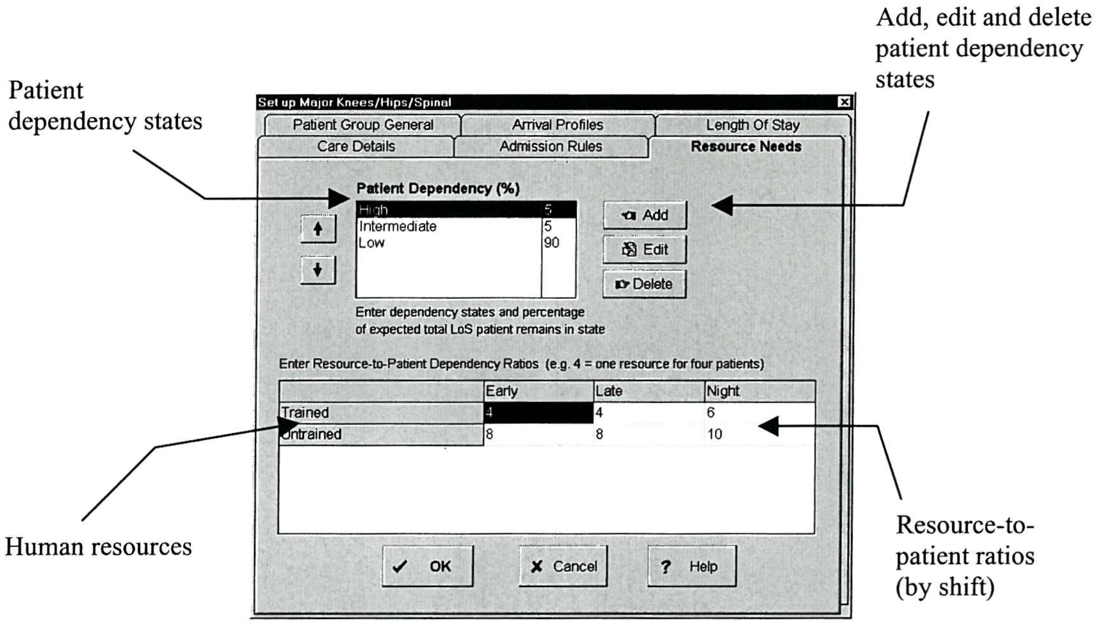


Figure 6.16: Workforce parameters

6.8.7 Operating theatres

When a *with procedure* patient arrives at the hospital and acquires a bed, the patient will then queue for theatre. When the theatre becomes free, the patient will be operated on. This process depends on a number of variables:

- ***Theatre times:*** the user must enter a *start time* and *duration* of each theatre session. More than one session may run in parallel. The number of theatre sessions is likely to depend on day of the week and possibly month of the year. As with defining bed numbers, the session times for the theatre are entered via month and weekly profile drop-down-boxes.

- ***Session operating rules:*** the user should define the maximum time that a session may overrun. This should be as a percentage of the total session time.
- ***Session scheduling rules:*** the user must chose from a number of session scheduling techniques.

The user may also define any number of appropriate *theatre resources*. For each resource, a resource description and quantity is defined. Each time a session opens these resources will be used. For example, an Anaesthetist with quantity one will result in one Anaesthetist being used for every session. Summary statistics of resource use over time and other key measures will be generated.

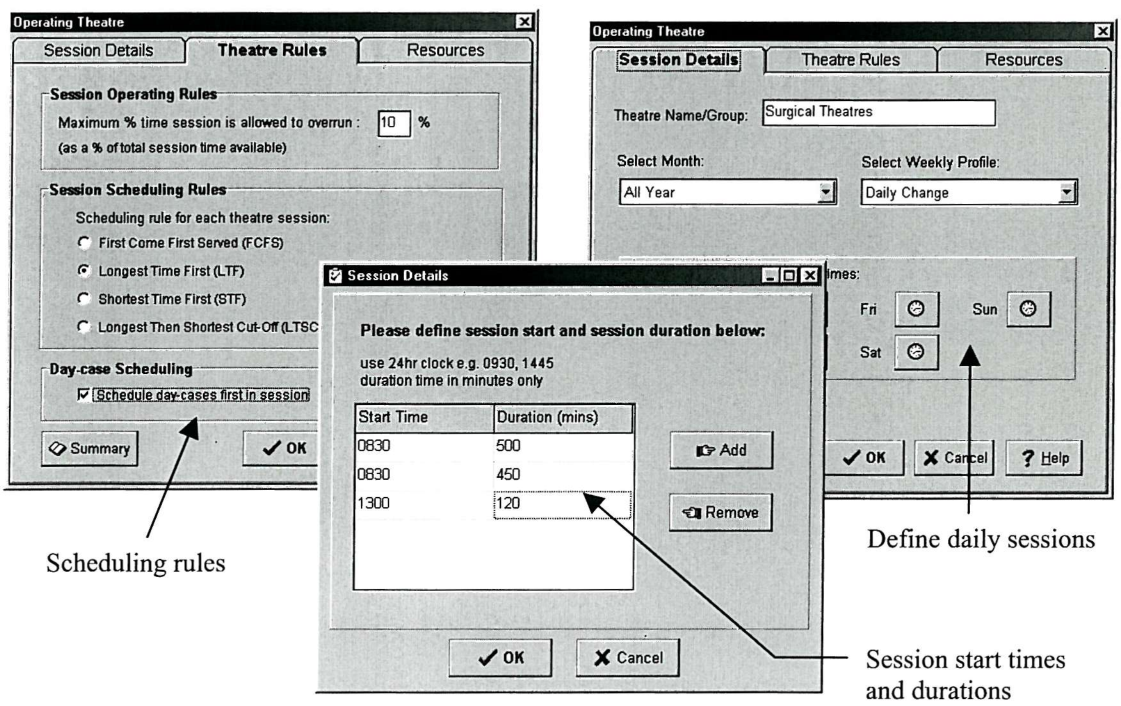


Figure 6.17: Operating theatre parameters

6.8.8 Simulation parameters

These screens allow the user to edit the simulation parameters and to run the model. The model has been designed to simulate one financial year (1st April to 31st March). The user has the possibility to change the following variables:

- ***Warm up time:*** the time (in years) that the model will run without collecting data.

- ***Number of runs:*** each run comprises of the warm up time plus one year where data is collected.

- ***First day of financial year:*** given an initial starting day (first day of financial year), the model calculates how many of each day there are in each month. For example, if April 1st is a Monday, then given April has 30 days, there will be 5 Mondays and Tuesdays in April and only 4 Wednesdays, Thursdays, Fridays, Saturdays and Sundays (the model will similarly calculate the number of each day for each month of the year).

- ***Cap patient LoS (check-box):*** for each defined patient group, there is a minimum and maximum length of stay (LoS). These values are used in the random sample of LoS for each patient from the given distribution. Thus the model will only sample LoS values that fall within this minimum and maximum interval. A further use of the model is to *cap* LoS. When we cap LoS, we effectively say that any LoS above the defined maximum will be set to the maximum LoS. This is useful, for example, in studies to evaluate the distribution of acute and post-acute bed needs whereby patients are transferred out of the ward after a certain cut-off LoS.

- ***Module options:*** there are two modules available - Theatres and Resources. The user may turn on/off these modules. The model will only give the user access to the relevant parts of the model depending on which options are currently activated.

- ***Graphics:*** the PROMPT model may be run with or without a graphical display. The graphics show the occupancy rates, and numbers of patients who stayed and who were refused in each care unit. This is particularly useful in validating the model, but is much slower than running the model with the graphics turned off (default). At any stage during the running of the model, with the graphics on, the user may turn off the graphical display. The user, through this option, can enable or disable the graphical display.

Once the required simulation parameters have been set, the simulation may be run.

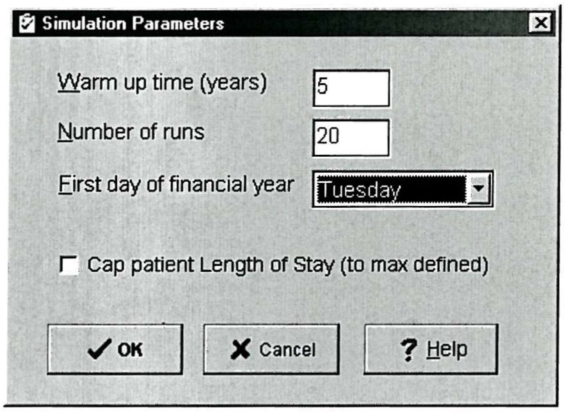


Figure 6.18: Simulation parameters

6.8.9 Simulation outputs

A number of statistics and graphs are available on completion of the simulation run(s). These are broadly at the patient level and care unit level, and include:

Graphs

- Beds in use over time.
- Patient frequency.
- Emergency transfers and elective deferrals over time.
- Deferral frequency.
- Number of patients moved over time (outliers).
- Patient LoS distributions.
- Number of operations over time.
- Patient waiting times for theatre.

Statistics

- Bed-days used.
- Occupancy rate, transfers, deferrals and outliers.
- Monthly occupancy and refusal rates.
- Workforce needs over time.
- Patient waiting times.
- Observed LoS.

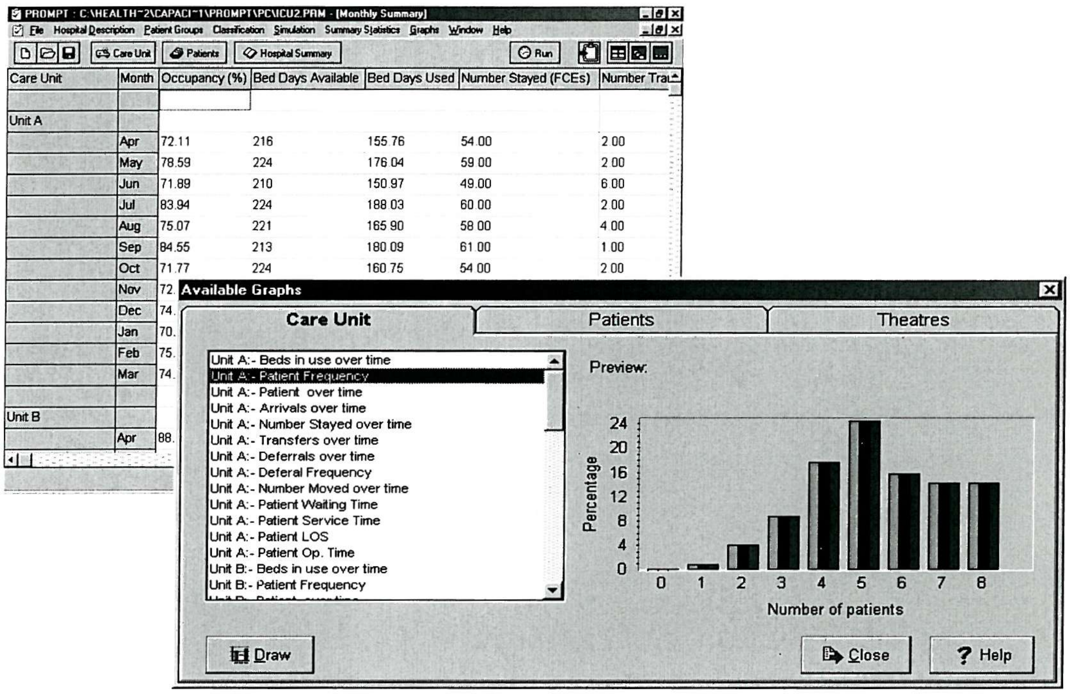


Figure 6.19: Illustrative model outputs

6.9 Chapter Summary

An operational model for hospital resources, PROMPT (*Patient and Resource Operational Management Planning Tool*), has been developed within the evolved generic framework for modelling of healthcare resources. PROMPT was developed in a Delphi environment using a three-phase simulation shell TOCHSIM. Over-time, and with an increasing knowledge of the hospital processes and perceived model utilisation, a schematic diagram of patient-flows though a hospital system was developed and incorporated into the model. This is suitably generic, allowing for the model to be readily used by other hospitals. This has been witnessed by the adoption of the framework by Portsmouth Hospitals NHS Trust who recognised that the structure applied to them. Other Trusts have since approached the author expressing an interest. The developed simulation takes individual patients through time as they arrive and pass through the hospital, whilst capturing and monitoring the necessary resource needs, such as operating theatres and nursing needs.

Throughout the research work constant validation and verification was conducted in order to increase confidence in the model operations. Verification was largely achieved through the use of the generic framework and the adopted evolutionary model development. This ensured that end-users contributed to the model development at all stages and ensured that the simulation model reflected the original conceptual schema. Validation techniques include the comparison of the simulation to analytical queueing models.

Many of the potential end-users are likely to have limited experience or practice in using computer simulation models. Delphi software enabled the simulation to be readily tailor-made for the hospitals within a familiar Windows environment for ease of use. The resulting practical simulation model incorporates the necessary complexity, uncertainty and variability, for solving a wide range of hospital planning and management issues. Chapter 7 will demonstrate PROMPT in use at the participating NHS Trusts.

Chapter 7 – Hospital Case Studies

7.1 Chapter Introduction

Previous chapters have discussed the development, structure and validation of an operational model, PROMPT, for the modelling of hospital resources. This chapter demonstrates model applications through case studies undertaken during the time spent with the participating NHS Trusts.

Expressed user needs and requirements helped to mould the model structure throughout its development. In particular, PROMPT has been designed to be used:

- as a tool for evaluating *ad-hoc* studies, such as the creation of a new care unit, hospital consolidation plans or changes to patient-flows through the hospital system.

- as a management and planning tool that can be used in long-term planning and during the annual business planning cycle.

During the time spent with both the Royal Berkshire and Battle Hospitals NHS Trust and Portsmouth Hospitals NHS Trust, the model played a fundamental role in a number of hospital studies to evaluate hospital process re-design options. It was also used in the annual business planning cycle with specialty managers to assess the likely scenarios for the forthcoming financial year, and as a tool for agreeing customer service levels with GP Fundholders, local Health Authorities, and in the case of Portsmouth, for the outline business case of the PFI project.

The primary purpose of the studies is to assess the general methodology in the context of the genuine concerns of healthcare managers. The experience of using the operational models and the evaluation of their effectiveness in a practical setting is therefore the primary objective in the research. Only a few case studies out of the many undertaken are presented in this chapter. It is hoped however that the selected studies will allow the reader to appreciate the spectrum of possible model uses.

7.2 Case Study One – Modelling the Provision of Adult Medicine Beds

One of the major specialties within the Royal Berkshire and Battles Hospitals NHS Trust, and indeed within any Trust, is Adult Medicine. This specialty has witnessed the largest growth in emergency demand over recent years and plays a pivotal role in the rest of the hospital system. Table 7.1 summarises the various patient groups who use the medical beds together with their status (emergency, elective, day-case or all in-patients), mean length of stay, length of stay inter-quartile range and observed demand (referrals in finished consultant episodes) for the 1996/97 financial year.

Table 7.1: Adult Medicine patient groups

Patient Group	Status	Mean LoS (days)	Inter-quartile Range (days)	Referrals
Cardiology	Emergency	3.2	1.0 - 4.0	1,603
	Elective	2.7	1.0 – 3.0	196
Dermatology	In-patients	15.7	5.8 – 19.8	27
Elderly Care	Emergency	10.9	4.0 – 13.0	2,062
	Elective	18.1	3.8 – 20.0	17
Elderly Rehab	In-patients	17.3	7.0 – 21.0	1,430
Gastroenterology	Emergency	1.5	1.0 – 1.8	500
	Elective	2.3	1.0 – 2.0	205
General Medicine	Emergency	5.3	1.0 – 6.0	6,890
	Elective	2.9	1.0 - 2.0	1,098
	Day-case	0.7	0.5 - 0.8	194
Neurology	Emergency	9.9	2.8 – 12.3	68
	Elective	4.7	2.0 - 5.0	33
Rehabilitation	Emergency	57.0	24.8 – 89.5	74
	Elective	55.0	14.8 – 69.3	35
Rheumatology	Emergency	9.6	1.0 – 13.3	248
	Elective	11.2	6.0 – 14.0	205
Thoracic Medicine	Emergency	5.0	1.0 - 7.0	154
	Elective	1.7	1.0 - 2.0	48
Totals	-	7.1	1.0 - 8.0	15,087

Table 7.1 illustrates the large variation in length of stay within and between patient groups. Traditional capacity planning methodology would have involved the overall mean LoS for emergencies and electives, equating to 7.5 and 4.3 days respectively. This fails to capture the inherent variability in LoS, and as this case study shall show, fails to estimate the true bed requirements. It is therefore necessary to obtain LoS distributions for each of the patient groups above. Furthermore, hourly, daily and monthly referral profiles are obtained for each patient grouping and used within the model, thus incorporating the necessary daily and seasonal variations. Using a constant daily arrival rate that is independent of time greatly increases the likelihood of producing misleading bed requirements.

The PROMPT model was run with observed referrals for one financial year for the following scenarios:

1. Using the developed model that captures demand over time and uses LoS represented by appropriate statistical distributions.
2. With demand over time but using only average LoS for each of the patient groups shown in Table 7.1.
3. With demand over time but using only an average LoS for emergency and elective patient groups (7.5 and 4.3 days respectively).
4. Assuming constant arrival pattern (no seasonal effect) and with average LoS for emergency and elective patient groups.

Clinical managers at the hospital typically employ method 4 to estimate bed requirements. Figure 7.1 shows how each of these bed-planning options compare against the observed admissions for 1996/97. Given that bed occupancy varies within each month, only the monthly mean occupied bed number is shown. Table 7.2 shows how each method compares with the prediction of overall bed occupancy and refused admission rate.

Figure 7.1 and Table 7.2 illustrate how the use of average LoS results in misleading forecasts of bed requirements. The developed simulation model performs well against the observed 1996/97 occupancy and refusal rates. Up to 5,600 bed days are underestimated using scenario 3. Occupancy and refusal rates for average LoS scenarios are consistently lower than those observed. For scenario 4, the summary statistics given in Table 7.2 do not capture the large mismatch between planned beds and bed needs.

Table 7.2: Comparison of bed planning performance measures

Method	Occupancy (%)	Bed-days used	Refusal Rate (%)
Observed	93.5	106,500	2.6
1	93.6	106,600	2.7
2	89.2	101,600	1.5
3	88.6	100,900	1.4
4	91.3	104,000	1.1

The seasonal bed requirements are readily appreciated from Figure 7.1. Standard capacity planning techniques, where monthly and daily variation are not incorporated, would produce deceptive results. This is seen in scenario 4 where demand is assumed to be constant over time. Such a prediction would greatly under and over estimate actual requirements during different periods of the year. Given the stochastic nature of the model, it is possible to determine confidence intervals around these predictions. To illustrate this, error bars (95%) are shown for model prediction (method 1).

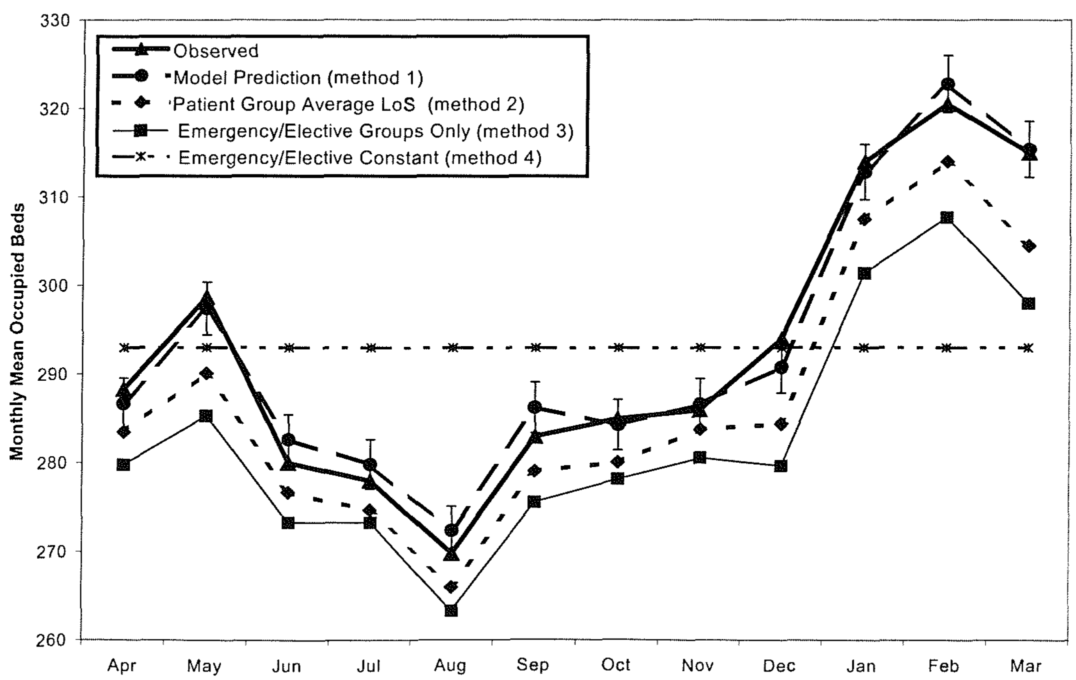


Figure 7.1: Observed and predicted Adult Medicine bed needs

Table 7.3: Comparison of different Adult Medicine bed configurations

Bed Configuration	Occupancy (%)	Refusal Rate (%)
Current (310 beds all year)	93.6	2.7
300 beds all year	94.9	4.7
320 beds all year	91.4	1.6
310 (current) with extra 20 beds Dec – Feb	93.0	1.9
320 beds Dec – May; 290 beds Jun – Nov	93.1	3.2
330 beds Jan – Feb; 290 Jun – Oct 310 Nov, Dec, Mar – May;	93.9	3.3

The developed bed-planning tool may also be used to examine the relationships between bed numbers, occupancy and refusal rates. Figure 7.2 and Table 7.3 illustrate how this relationship is complex and non-linear. Capacity planning options need to be explored in light of these sorts of calculations.

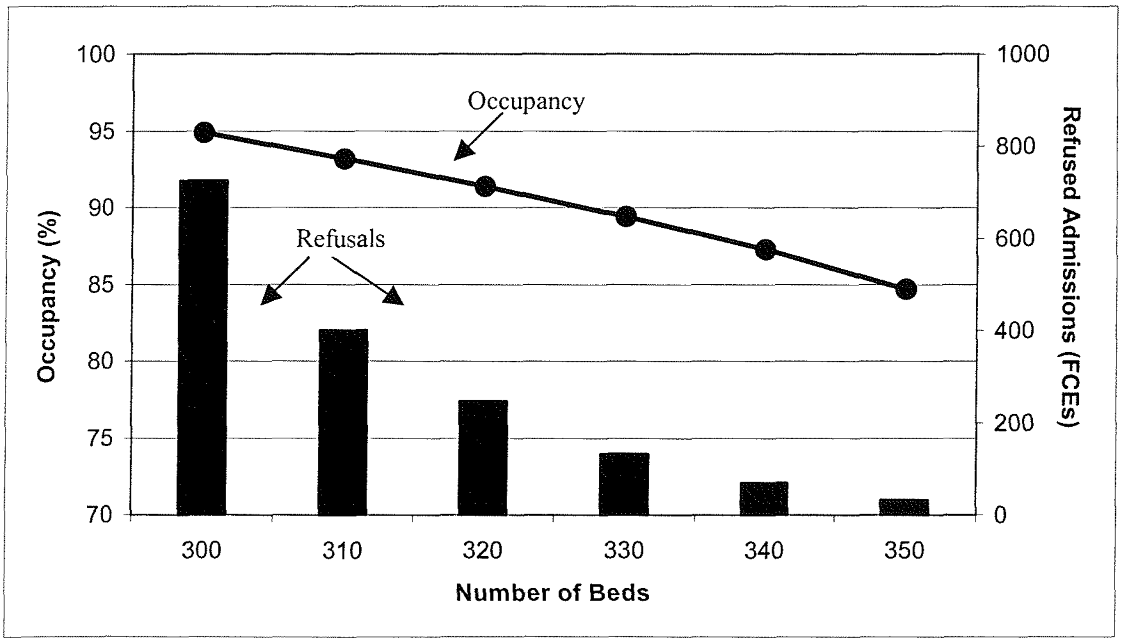


Figure 7.2: Beds, occupancy and refusals for Adult Medicine

7.3 Case Study Two – Creating a New Clinical Grouping

One great advantage of the flexible nature of the model is that the user is able to investigate the likely consequences of changes in capacities and hospital practices. This may be through changing current configuration of existing wards or through the creation of new clinical groupings or units. For this latter purpose, this case study outlines the pivotal role that the model played in helping managers at the Royal Berkshire hospital to investigate the care-needs of a proposed new clinical unit, respiratory medicine, and its likely effect on the remainder of the hospital system. Managers and clinicians needed to fully understand the consequences before implementation. They were particularly interested in the following:

- The bed needs of the new respiratory unit.
- Whether to *ring-fence* these beds or to share with other specialties, such as general medicine.
- The effect of removing respiratory patients from adult medicine and other hospital specialties and the corresponding potential bed-day savings, if any, from the creation of this unit.
- Bed-day savings and reduced bed needs as a result of likely reductions in LoS of respiratory patients.

Using the developed statistical module, it was possible to rapidly access and analyse information on respiratory patients, based on clinical coding, from the large hospital database. Respiratory patient groupings were created and demand profiles and LoS distributions were obtained. These patient groupings were then passed through a respiratory unit in the capacity model and the likely bed needs were evaluated (Figure 7.3). Furthermore, the adult medicine and respiratory bed-pools were considered together in the model in order to examine the consequences of removing the respiratory patients from the adult medicine bed-pool. The bed needs shown in Figure 7.3 show a marked seasonal demand. This graph indicates that beds could be stepped up and down over the year. It is noted with interest that before the model was used, managers had provisionally decided on a capacity of 25 beds kept constant across the year, based primarily on rough deterministic calculations. Their opinion was greatly changed after the modelling study.

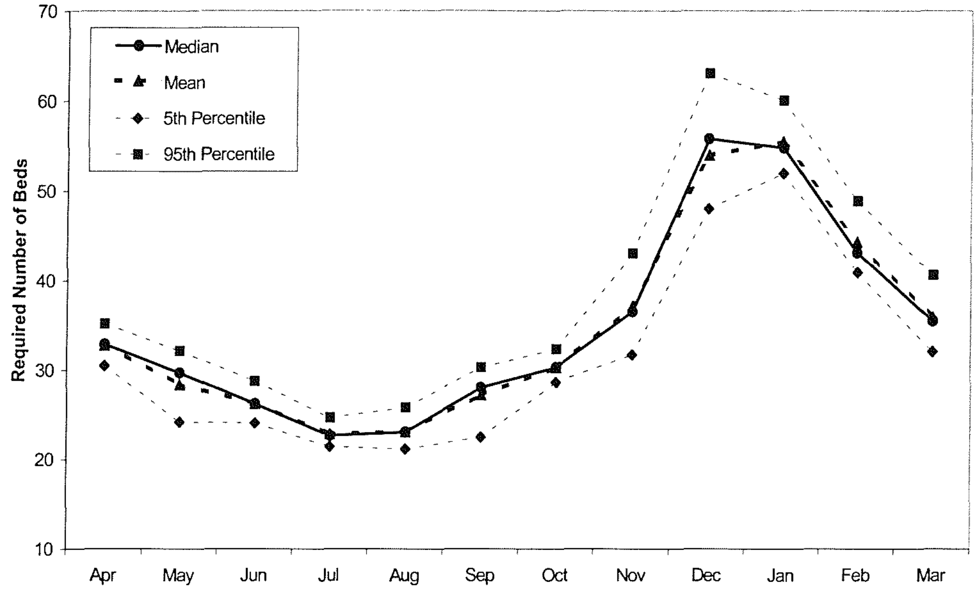


Figure 7.3: Respiratory bed-needs over time

The motivation behind the provision of a new clinical group is to provide focussed care to a particular class of patients. At the time of the study, respiratory patients could be found in up to 15 different wards across the hospital, many of which were outside of the general medicine specialty. It was hoped that the LoS for this group of patients would decrease, over time, within a dedicated unit. The flexible nature of the developed model enabled a number of different LoS reduction scenarios to be easily and rapidly explored.

It was decided that a realistic method was to attempt to achieve a proportional reduction in LoS because this proportional approach means that we have a greater chance of a larger LoS reduction in the upper range of LoS than for those who stay a short time in hospital. The scenarios of 10%, 20 and 30% reductions were chosen and the model was used to calculate the resulting bed needs (Table 7.4).

Table 7.4: Respiratory patient LoS reduction and resulting hospital bed-day savings

LoS Reduction	Adults		Elderly		Overall Mean LoS	Bed-day Savings
	Mean LoS	95% Value	Mean LoS	95% Value		
Current	5.7	16.0	11.8	33.0	8.7	-
10%	5.2	14.5	10.6	30.0	7.7	1,000
20%	4.8	13.0	9.5	26.5	6.8	2,500
30%	3.9	11.5	8.8	23.0	5.9	3,600

Table 7.4 shows the revised mean LoS, 95% point of this LoS distribution, and predicted hospital bed-day savings under the 10, 20 and 30% LoS reductions. This Table helps show that bed-day savings is not a linear function of LoS reduction given the stochastic nature of queues in the hospital system.

7.4 Case Study Three – Hospital Consolidation and Reconfiguration.

Portsmouth Hospitals NHS Trust is one of the largest in England, with an annual income of around £240 million (2001/02) and providing acute healthcare services for almost a million people. Internally the Trust launched a consolidation and reconfiguration programme of facilities. The Trust successfully applied for PFI status and was asked to submit a fully costed Outline Business Case (OBC), to appraise the options for delivering PFI benefits. The Trust needed to calculate the likely bed needs for the consolidated hospital and examine various reconfiguration plans, such as the effects of curtailing acute LoS and increasing the capacity of post-acute beds.

Working within the evolved modelling generic framework, the first phase included the necessary extraction and analysis of the large hospital database using Apollo. This database included five years worth of individual patient data, including LoS, episode start and end dates, hospital specialty, HRG (procedure code), emergency or elective status and inpatient or day-case episode. Portsmouth required a number of options to be explored. These included the bed-needs by existing hospital specialty (General Medicine, Orthopaedics, Paediatrics etc.) and additionally by re-designed hospital clinical groupings derived primarily by clinical procedure (for example hips, knees and neck of femur as one clinical grouping). There were nine such groups in total.

Apollo was used to create the necessary patient groupings under both scenarios. Groups were formed using specialty and/or HRG fields in the database. Table 7.5 shows an example group created within Apollo, with a number of key statistical indicators and fitted LoS distributions. Monthly and daily arrival profiles for each group were readily captured using episode start date and this information saved alongside the LoS distributions for use in PROMPT.

Table 7.5: Example hospital consolidation clinical grouping

Patient Group	**Status**	**Episodes (1998/99)**	**Mean LoS (1998/99)**	**Fitted Distribution**
Vascular Surgery	Emergency Elective	600 663	8.8 4.6	Weibull (7.23, 0.73) Lognormal (0.53, 1.44)
Vascular Medicine	Emergency Elective	470 44	2.7 2.9	Gamma (0.15, 17.98) Weibull (1.04, 0.42)
Poisoning	Emergency	892	1.7	Lognormal (0.01, 10.6)
Thoracic Procedures	Emergency Elective	93 31	2.0 1.7	Weibull (0.33, 0.32) Weibull (0.55, 0.41)
Thoracic Medicine	Emergency Elective	3,882 428	5.7 3.5	Weibull (5.10, 0.81) Lognormal (0.05, 1.58)

Evolved patient groups were loaded into the PROMPT model and the necessary admission and patient-flow rules defined. Bed-needs were captured for each specialty and then for each clinical grouping. Various sensitivity analyses were then examined by hospital managers to assess their impact on the overall costing for the PFI. Initially these included changing the demand and LoS. Sensitivity on LoS was aided by examining LoS over-time (five years worth of data was available) and forecasting likely LoS percentage increases or reductions. Access was also given to peer group hospitals LoS. These LoS figures came from ten similar size hospitals across the UK and allowed for comparison of Portsmouth LoS to the peer group average and best peer group performance (lowest LoS for a given procedure). Figure 7.4 shows how PROMPT was designed to allow end-users to easily change the shape of the LoS distribution using a percentage increase or decrease. This avoids the need to re-fit the distribution in Apollo.

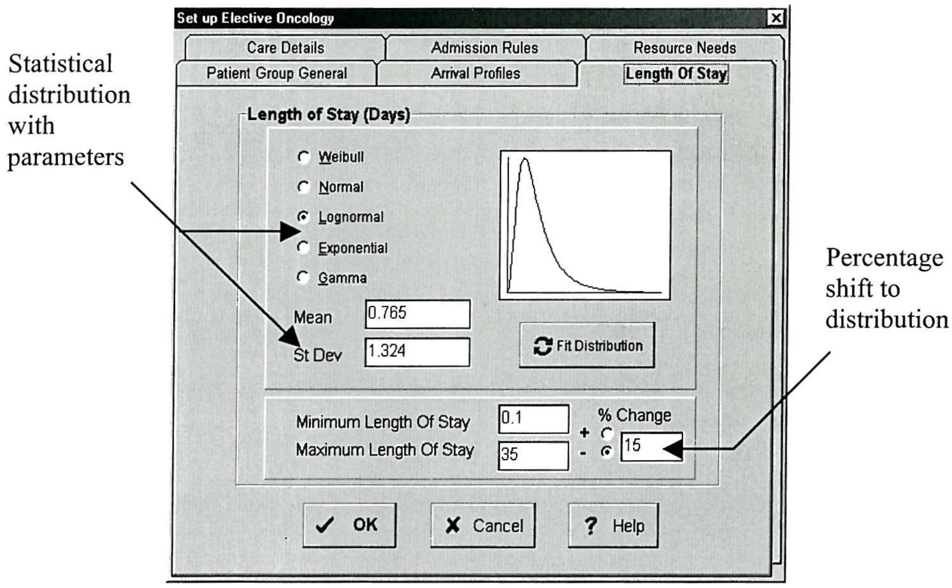


Figure 7.4: LoS sensitivity analysis by shifting the LoS distribution

An illustrative LoS shift of 15% (decrease) for a Portsmouth Oncology patient group is presented in Figure 7.5. The mean LoS was 5.12 days with a 95th percentile point of 16 days. The fitted distribution was Lognormal with mean 0.765 and standard deviation of 1.324. After the 15% shift the LoS is 4.35 days and 95th point at 13.6. The revised fitted distribution has a mean of 0.593 and standard deviation unchanged at 1.324.

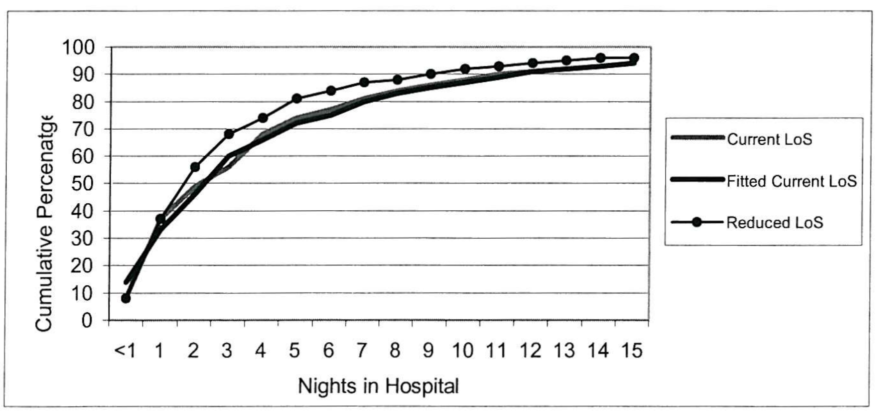


Figure 7.5: Illustrative percentage shift to a patient group LoS distribution

A matrix of bed-needs was created under a number of LoS scenarios. This allowed hospital managers and financial personnel to agree on achievable LoS and hence read-off the corresponding hospital bed-needs. A subset of the full hospital matrix is shown in Table 7.6 below.

Table 7.6: A matrix of hospital bed-needs by LoS scenario

Speciality				Changes in number of beds % LoS reduction/increase (4)		
	Deterministic (1)	PROMPT (2)	Best LoS (3)	5%	10%	20%
Cardiology	3	3	3	0	0	1
Dermatology	10	11	10	0	1	2
General Medicine	270	282	236	10	22	47
Thoracic Medicine	1	1	1	0	0	0
Haematology	8	9	7	0	1	2
Orthopaedics	149	159	148	7	14	30
A & E	9	9	9	0	1	2
Rheumatology	23	26	21	1	2	5
General Surgery	137	143	132	6	12	25
ENT	18	20	17	1	2	4
Ophthalmology	6	6	3	0	0	1
Oral Surgery	8	9	7	0	1	1

(1): Beds required for 85% occupancy using a spreadsheet deterministic calculation.
(2): Beds required for 85% occupancy using the PROMPT model.
(3): Beds required for 85% occupancy using PROMPT and best peer group LoS
(4): Change to the number of beds required under various LoS percentage shifts.

Of particular concern amongst hospital managers was the distribution between acute and post-acute beds. There is a growing acknowledgement within the medical profession that for many categories of patients within the hospital, the acute LoS (time in the hospital) is unnecessarily long. On discharge, many of these patients move to beds within local community hospitals (post-acute stay). Unfortunately due to existing limited post-acute capacities, many patients have to wait some time in hospital before transferring to more suitable (and possibly cheaper) post-acute care. The distribution therefore between acute and post-acute bed capacities plays an important role in the planning and management of NHS Trusts.

To understand and model the system, PROMPT was used to examine a number of patient-care scenarios. Working with specialty managers, we were able to identify groups of patients who would benefit from a post-acute phase and less time within the hospital itself (typically elderly patients). Furthermore using Apollo we were able to examine current practice in detail (LoS distributions for these groups) and challenge existing protocols of care. Using expert opinion, PROMPT was used to model various "discharge cut-off times" for different groups within the Trust. Essentially these times refer to the maximum permitted time that a patient from this group may spend in the hospital. If a patient is still in a bed at this time, they will be transferred to post-acute care. Thus we were able to monitor both acute and post-acute bed needs over time under various clinical and managerial patient-care decisions. The Trust was able to gain a system-wide view of both the Trust and community care needs and incorporate these into the developed PFI outline business case.

7.5 Case Study Four – General Surgery Theatre Needs

Like many other NHS Trusts, Royal Berkshire and Battle Hospitals needed to re-evaluate their theatre capacities in light of existing capacities, forecasted patient case-mix and a constant shift in the distribution between in-patients and day-case surgery. To guide this work, a multi-disciplinary working group was created and tasked with evaluating a variety of theatre options. Concerns included the numbers of theatre sessions by specialty, the daily distribution of these sessions and the scheduling of patients within the sessions. Essentially the task group was looking to make better use of existing capacities through improved theatre management. In addition they required a tool to assess a number of longer-term planning decisions.

After various discussions with the group, a number of objectives were defined. In essence the hospital must find efficient and effective policies that account for the complex nature of patient care. For example the policies must account for the daily number of available hospital beds before considering theatre needs. Clearly it would an oversight to examine the numbers of possible operations without accounting for the necessary bed capacities to accommodate this patient population. The objective is two-fold:

- ***Maximise theatre session utilisation (daily event)***
  Adopt a scheduling policy that will prevent session under-run and large over-run, giving greater utilisation (time theatre in use divided by time theatre open). Utilisation will depend on patient operation times.

- ***Optimise bed utilisation (across the week)***
  Adopt an admission policy that will allow maximum utilisation of beds. This will depend on post-operation LoS. For example, on which day(s) of the week should we admit the major operations (those who will subsequently stay a long time on the ward)? There is a need to examine outlier and deferral rates as a measure of policy benefit.

Combined, these two objectives should help to maximise throughput (number of patients we can see in a year) and help to flatten out occupancy levels over the week, thus avoiding extremely busy and slack periods in the bed-pool. In turn this should reduce hospital refusal rates and theatre cancellations.

PROMPT was used to examine all surgical specialties within the Trust. Each specialty brought its own complexities, such as differing case-mixes and theatre needs. As an illustration of the PROMPT model in use, the General Surgery case study is discussed below.

The results also account for non-surgical patients in the General Surgery bed-pool, such as medical outliers and those surgical patients not requiring theatre (accounting for 27% of all patients in General Surgery). These *without procedure* patients are additionally captured within PROMPT and are passed through the bed-pool to reflect true bed-needs accounting for all admissions. This is a necessary condition of a practical hospital bed and theatre capacity model.

Royal Berkshire and Battle Hospitals NHS Trust categorise all surgical patients into one of three classes: minor, intermediate and major patients. The categories reflect the severity of the operation (procedure) and consequently act as an indicator to likely workforce needs and patient post-operative LoS. Apollo was used to examine current General Surgery practice by class of patient (Table 7.7 and Figure 7.6).

Table 7.7: General Surgery patient statistics

Patient Status	Patient Category	Number of Operations	Av. Operation Time (mins)	Average Post-op LoS (days)
Day-Case	Minor	197	20	< 1
	Intermediate	608	32	< 1
	Major	85	44	< 1
	Overall	920	30	< 1
Elective	Minor	194	21	2.0
	Intermediate	879	32	2.0
	Major	1,469	89	4.8
	Unclassified	266	-	-
	Overall	2,808	63	3.4
Emergency	Minor	201	23	2.0
	Intermediate	199	46	2.7
	Major	920	81	6.0
	Overall	1,320	61	4.6
All Data	Overall	5,048	54	3.0

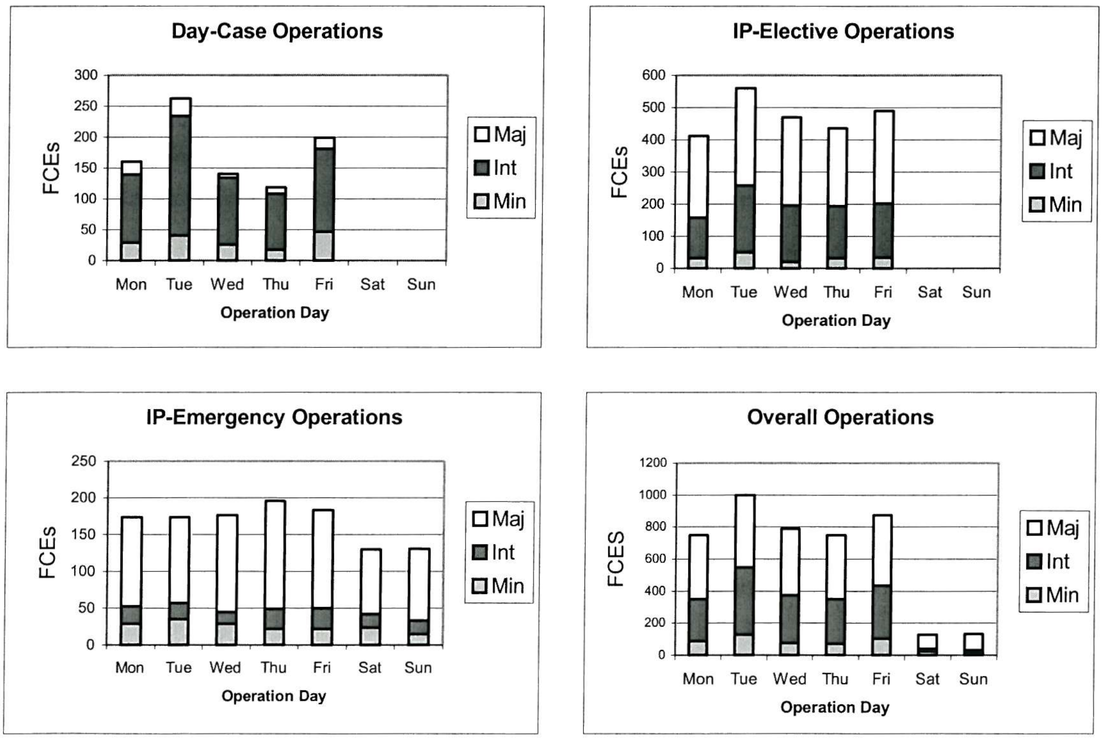


Figure 7.6: General Surgery operation day profiles

Figure 7.6 helps to illustrate the large daily variation in numbers of operations. These peaks and troughs throughout the week are largely influenced by historical theatre schedules and preferred working patterns of surgeons who have a significant say in when different types of operation are performed. The corresponding impact on bed-needs is complex and non-trivial, depending on the estimated post-operative LoS of each patient. Furthermore, because of restricted bed-capacities on the wards, current emergency demand should be acknowledged when planning and managing elective theatre schedules.

The nine with-procedure patient groups shown above, together with their LoS and operation time distributions and current admission profiles were used in PROMPT. Two care units were created: a surgical inpatient bed-pool and a day-case centre. Existing theatre scheduling rules and numbers and durations of daily theatre sessions were defined.

Various scheduling methods were experimented with in an attempt to better manage existing capacities and meet the stated objectives. The following conclusions concern only the management and planning of existing bed and theatre capacities. The model however has also been used to examine a number of options, including the increase of bed and theatre capacities.

Scheduling of general surgery theatre sessions

Schedule expected large operations first followed by the smaller operations later in the session i.e. major first, then intermediate and then minor. The larger operations generally have the most variability in operation times. By scheduling these first and smaller ones later, there is less chance of closing the session early and thus increasing throughput and utilisation.

Adopting LTF (Longest Time First) strategy:
Current session occupancy = 88%, Current refusal rate = 5.0%
Predicted LTF session occupancy = 85%, Predicted LTF refusal rate = 3.8%

Essentially, this shows that under a LTF policy, we can still achieve the same number of operations but reduce occupancy. Put another way, we can improve efficiency of our sessions by maintaining current throughput but on average saving session time under a different session schedule. This extra saved time could be used to operate on more patients, thus increasing throughput without the need for extra resources (i.e. beds and theatre time). In fact, this reduced occupancy equates approximately to an extra 60 hours a year operating time (current total operation times approximately 3,800 hours) plus a drop in the refusal rate. In turn, this equates to potential 1.6% increase in throughput (or around an extra 70 patients operated on per year) without the need for extra resources.

Day-case and inpatient sessions

Generally it is best to keep day-case patients in the same theatre as in-patients. As day-cases typically have shorter and less variable operation times, it makes logical sense to include these patients with other in-patient theatre sessions if we are primarily concerned with session utilisation. Where possible, some day-cases could be placed towards the end of the sessions to reduce potential under-run and wasted time (as LTF above). For example, the majority of day-cases could go towards the end of the morning sessions. This might well help keep the time that day-case ward needs to open down to a minimum as well. Practical considerations will dictate how late in the day a day-case patient may commence their operation.

Weekly profiling

So far we have only examined the effects of scheduling patients within a session. This has assumed that the day-to-day scheduling of patients remains unchanged. So, for example, do we admit major patients early or late on in the week, or does it not make any difference when these patients are admitted? This should help to "even-out" busy and slack times in the bed-pool, making optimal use of beds whilst helping to reduce refusal rates. We will now need to consider pre and post-operation LoS. Post-op LoS for each patient group is shown in Table 7.7. The data revealed that 6% of minor, 11% of intermediate and 47% of major patients needed to be admitted the day prior to their operation on clinical grounds. This profile, based on hour of arrival, was reflected in the model.

From the many possible schedules examined together with hospital personnel and the working group, and whilst respecting constraints such as surgeon's availability, the schedule shown in Table 7.8 was produced which gave a benefit of a flatter occupancy across the week, which helped the overall refusal rate drop from 5 to 3%.

Table 7.8: Proposed General Surgery weekly scheduling profile

Patient Group	**Operation Day**
Minor	Early in week (majority on Monday)
Intermediate	Early to Mid Week (majority Tuesday & Wednesday)
Major	Mid to Late Week (majority Thursday & Friday)

Figure 7.7 and Table 7.9 show the current and proposed occupancy rates by day of the week. Please note that the occupancy rate shown is the occupancy at the start of the day *before* admissions on that day.

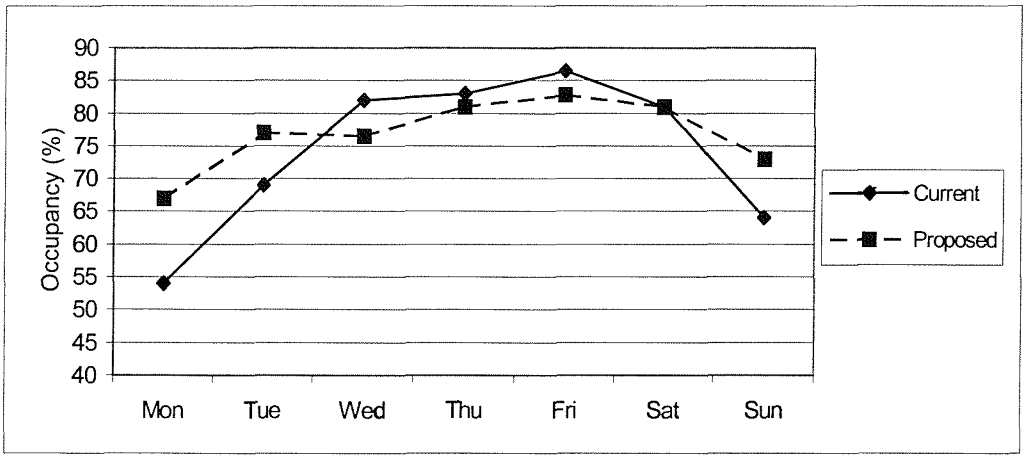


Figure 7.7: Comparison of General Surgery daily occupancy rates for current and proposed scenarios

The logic behind this schedule is that the major patients are now very likely to stay in the bed across the whole weekend (days 3, 4, or 5 of their stay). This is a time when there are no (or little) admissions, thus utilising beds when perhaps weekends might be less busy than early to mid week currently. The majority of these patients would then be discharged during Monday in time for the short LoS Minors and all discharged before Intermediates in the mid week. This is shown in Table 7.9 and Figure 7.7 by the flattening out of occupancy, and hence probability of a bed being free for admitting patients in that day, across the days. Essentially we attempt to reduce peaks and troughs in occupancy over the week and hence reduce the likelihood of all beds being full and refusing patients.

Table 7.9: General Surgery occupancy and likely bed availability for current and proposed scenarios

Start of Day	Occupancy in bed-pool (%)		Probability of a free bed (%)	
	Current	Proposed	Current	Proposed
Monday	54	64	46	36
Tuesday	69	77	31	23
Wednesday	82	76	18	24
Thursday	83	81	17	19
Friday	86	84	14	16
Saturday	81	81	19	19
Sunday	64	72	36	28

The model has highlighted a number of areas for potential improvement. Some of these policies equally apply to other specialties within the hospital and indeed to other NHS Trusts. The general message from this modelling work is that an operational modelling approach, through necessarily practical and detailed capacity tools, can be used to experiment with the system helping to improve existing planning and management of hospital theatre resources.

7.6 Case Study Five – A Study of Trauma and Orthopaedic Nursing Needs

The NHS is one of the largest employers in the world, employing over one million people, including 350,000 nursing staff and 140,000 administrative and clerical staff (Department of Health, 2000). The NHS needs to continually add to its complement of workers to meet the needs of increasing admissions and to keep the service running 24 hours a day, 365 days a year. Managers need to quantify the number and type of workforce required in order to successful staff a hospital. Shortages in necessary nurses, for example, can lead to the temporary closure of hospital wards. With beds in constant demand, managers can ill-afford to have to take such action.

To illustrate the workforce module, this case study outlines the role of the model in the planning of Trauma and Orthopaedic nursing needs within The Royal Berkshire and Battle NHS Trust. Apollo was used to derive the appropriate patient groups using a combination of statistics (CART) together with clinical knowledge based on defining patient groups with homogeneous hospital nursing needs. This process produced the following five distinct clinical groupings:

- Hips
- Knees
- Spinal (decompositions and fusions)
- Discectomies, minor hands/knees, head of femur, spinal lesions, dissolving discs
- Shoulders, other major joints

The workforce module builds on the foundations of the bed capacity model: patients from each group arrive, attempt to obtain a bed and stay for a LoS before discharge. Additional information is now required governing the care of patients during their stay in hospital.

For each patient group, necessary resource needs for the patient during their stay on a Trauma and Orthopaedic ward were defined. These needs were provided in the form of patient-to-resource ratios (or dependencies). Dependencies were defined for different stages during a patient's stay together with the expected percentage of time that the patient will stay in that state (as a percentage of their expected total LoS). For each state, dependency ratios were entered against each resource. Furthermore, the ratios have been defined for each nurse shift - early, late and night. This reflects the varying care-needs across the day.

An illustrative resource needs profile for a hip replacement patient is shown in Table 7.10. For this procedure clinicians, nurses and hospital managers defined three dependency states: High, Intermediate and Low. Nursing needs for two types of workforce resource are studied: Trained Nurse and Untrained Nurse. Ratio figures indicate the number of patients that one nurse may care for in a given shift.

Table 7.10: Resource needs profile for a hip replacement patient

Dependency State: High (5% of LoS)			
Resource / Shift	Early	Late	Night
Untrained Nurse	5	5	6
Trained Nurse	3	4	5

Dependency State: Intermediate (10% of LoS)			
Resource / Shift	Early	Late	Night
Untrained Nurse	3	4	6
Trained Nurse	3	4	6

Dependency State: Low (85% of LoS)			
Resource / Shift	Early	Late	Night
Untrained Nurse	3	4	5
Trained Nurse	5	5	6

Figure 7.8 illustrates how hospital configuration (beds and admission rules) and patient information (LoS and resource needs) are combined to form an integrated workforce planning and management approach using the PROMPT methodology.

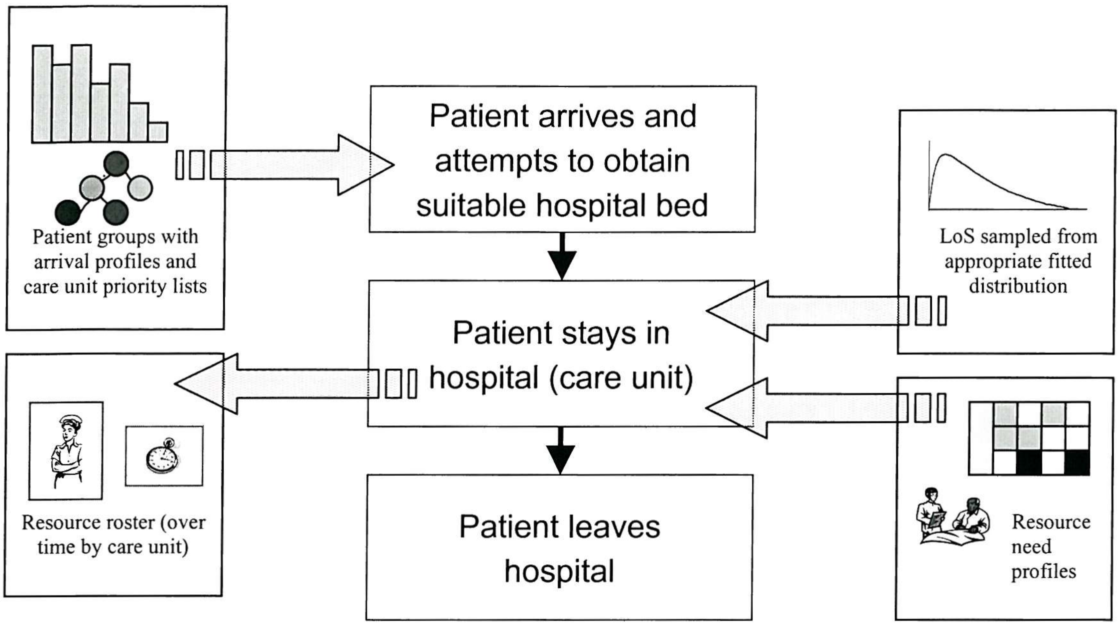


Figure 7.8: An integrated approach: modelling hospital workforce requirements

The above structure was adopted to predict likely nursing needs for Trauma and Orthopaedic wards in the Royal Berkshire and Battle hospital. The model enables managers to simulate patients through time whilst assigning corresponding care needs, allowing for the rapid production of workforce rosters for each resource by month, which are further broken down to day and shift levels. Because workforce needs will vary between simulation runs and during the same shift/day combination within a month, the mean number of resources required is displayed with 95% confidence intervals.

The model was used by the hospital to ascertain nursing needs for various wards and specialty bed-pools across the Trust. Because of the integrated PROMPT model, managers and human resource personnel were able to evaluate the impact on necessary staffing levels as a consequence of changes to a number of hospital variables, including:

- Changes to the number of ward beds.
- Changes in patient demand.
- Changes to the number of contracted local health authority elective episodes.
- Changes to the number of theatres and durations of sessions.

7.7 Chapter Summary

The work presented in the case study reports helps illustrate the wider experience of addressing real and practical needs of health service professionals using the developed generic framework and the PROMPT methodology. This experience played a key element in helping to structure and develop the Apollo and PROMPT models. The various projects undertaken within the participating Trusts and the spectrum of planning and management issues provided a rich source of input into the evolution of the prototype versions of the models.

Each case study details how the methodology was integrated into managerial processes of NHS Trusts and the way in which the simulation and its interface responded to user requirements. The case studies help illustrate the wide use of the model in the planning and management of hospital beds, theatres and workforce. It has been successfully used within the participating hospitals to help mangers and clinicians understand and quantify the consequences of planning and management policies (see Appendix H). The model has highlighted to the participating hospitals the need for incorporating the necessary detail when calculating resource requirements. Monthly, daily and hourly demand and meaningful statistical distributions that capture the inherent variability in LoS, workforce dependencies and operation times are important in the development and use of planning tools for hospital capacities. Healthcare resource allocations should obviously be made in light of both bed and theatre occupancies and refused admission rates. The relationship between beds, occupancy and refusals is complex and often overlooked. The developed model can be used to study a variety of hospital planning issues and it has demonstrated to the hospitals that it is possible to improve efficiency and effectiveness of the available limited resources.

Many challenges were encountered during the work and were very real in the sense that authentic solutions were being sought for genuine managerial and clinical concerns. Many critical issues became apparent and were brought into sharp contrast, such as data availability and quality, and the political dimension of the NHS. These challenges, together with the author's experience and perceived modelling opportunities are discussed in greater detail in Chapter 9.

Chapter 8 – Simulation Models for Critical Care Services

8.1 Chapter Introduction

Although a critical care unit (CCU) forms a part of the overall hospital system, to an extent this unit exhibits distinctive planning and management challenges that are rarely seen elsewhere in the hospital. The extreme costs of critical care (section 3.3.4), the relatively few beds available and the critical medical condition of the patients admitted intensify the planning and management issues. There is currently a great need to better plan and manage critical care beds at both a local and regional level (Department of Health, 2000). In response, two simulation models for the planning and management of critical care services have been developed within the evolved generic framework for modelling of healthcare resources (Chapter 4). The first simulation models the complex flow of patients through an individual unit (intensive and/or high-dependency care). The second simulation models a number of critical care units within a region. Both models have been built using Simul8 (Visual Thinking International), a standard *off-the-shelf* simulation package, and enhanced through the use of Visual Logic and an Excel front and back-end. The deliberate use of Simul8 allows for a comparison between standard software packages with an adopted programming approach such as the development of the TOCHSIM shell. A discussion is presented in Chapter 9. This chapter outlines the development, structure and validation of these two critical care models with illustrative case studies.

8.2 Towards a Critical Care Unit Model

Although a critical care unit (CCU) forms a part of the overall hospital system, to an extent this unit exhibits distinctive planning and management challenges that are rarely seen elsewhere in the hospital. The extreme costs of critical care, coupled with the relatively few beds available and the critical medical condition of the patients admitted intensify the planning and management issues. Critical care concerns the provision within a hospital of both intensive care and high dependency care. An Intensive Care Unit (ICU) bed is usually reserved for patients with threatened or established failure of one or more organs, particularly respiratory, cardiovascular or renal systems. High Dependency Unit (HDU) beds have been introduced as a step between intensive care and ward care. They reflect a need for more suitable levels of care for patients and as a means for reducing some of the costs of an ICU.

The provision of critical care has to meet the challenges of considerable uncertainty and variability in the needs of the patients, high costs and scarce resources. The demand for critical care beds arises from many sources as emergency or planned admissions. The vast majority of the demand for intensive care and high dependency is experienced as emergencies and electives respectively. Patient's lengths of stay and the large costs of treating patients are highly variable.

A recent Department of Health review of adult critical care services (Department of Health, 2000) concluded that there is a great need to better plan and manage CCU beds at both a local (individual) and regional level. The review reinforced and reiterated the beliefs and concerns amongst CCU managers and consultants participating with this research. During initial lengthy discussions with the working group, and building on previous work by IMH (Ridge *et al.*, 1998), a number of characteristics concerning the flow of patients through a CCU were evolved (section 4.3.3). There was an evident desire amongst the participants to model in detail an individual unit followed by a model of a number of co-operating units within a region. Accordingly these research aims complement the Department of Health's directives, although it should be noted that the work commenced before the publication of the Department of Health's report.

There are a number of critical variables within a CCU that must be effectively and efficiently managed on an hour-to-hour, day-to-day basis. In particular, the working group expressed the need to examine the following key questions:

1. How many CCU beds do we need? (local, regional and national levels)
2. What is a good mix of ICU and HDU beds? (distribution of beds within a CCU)
3. Are the internal rules good enough? (admission and discharge rules etc.)
4. Is the geographical distribution of CCU beds right? (across a region/the nation)
5. Can we pool resources from different CCUs? (more efficient use of resources)

Both the individual and regional models have been designed and built within the Simul8 package (version 5), an off-the shelf simulation tool that is now widely used in both business and educational establishments. It would be possible to develop these models within TOCHSIM, the simulation shell in which PROMPT (Chapter 6) has been constructed. It was felt however that the use of Simul8 would permit the comparison between standard software packages with adopted programming approaches. This is discussed in Chapter 9.

8.3 Simul8

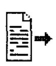
Simul8 (Visual Thinking Ltd.) was first used in industry in 1995. It is now used by thousands of users worldwide in enterprises such as Ford and Hewlett Packard. Much of its success can be attributed to its relative cheapness (compared to other simulation packages commercially available), its use of Windows™ 'point & click' technology to make Simul8 easy to use, and the ability for non-experienced simulation users to begin building simple models with very little training and in minimal time. Simul8 has built on the growing success of Visual Interactive Simulation (VIS) tools now used in many sectors of business and education. VIS technology is discussed in section 8.4.

Simul8 adopts the *process-based* approach to *discrete-event* modelling (section 3.6.2). It allows the user to create a visual model of the system being investigated by drawing objects directly on the screen. Typical objects may be queues or service points. The characteristics of the objects can be defined in terms of, for example, capacity or speed. The basic philosophy of Simul8 is to represent a system graphically, using *icons* to represent the various elements of the model. The flow of entities through the system is made by connecting these elements; the rules governing this flow can be made very complex. When running the simulation, the flow of work around the system is shown through animation on the screen so that the appropriateness of the model can be assessed. Once the structure of the model has been confirmed, a number of trials can be run and the performance of the system described statistically. Statistics of interest may include average waiting times, utilisation of facilities or resources, etc.

A Simul8 model consists of *objects* (items such as *storage bins* (queues) and *work centres*) on the screen with a default structure (routing) between them and *work items* which flow around the model. These work items are the entities that move around the system, for example patients in a hospital, products in a factory or invoices in an accounts department. Work items can have associated attributes (for example age, sex, diagnosis). Each individual work item can have different values for each of its attributes. Values of attributes may be changed and used by work centres. An important type of object is a *resource* which can be used at the work centres. For example, if work centres are machines, they might need recourses called people to operate them.

The basic Simul8 building blocks are summarised below (with default graphics, although these may be edited or replaced by images from a Simul8 graphics library). Corresponding illustrative CCU activities and objects are provided in brackets.

- ***Work Item*** – an object (entity) that moves around the system (*patient*).



Work Entry Point – a place where work items first appear in the system (*patient arrives*).

Storage Bin – a place where work to be done can wait until appropriate resources or work centres are available i.e. a queue (*patient waits for a bed*).

Work Centre –where work takes place on work items (*bed*).

Resource – items in the simulation model which are required at work centres in order for the work centre to work on an item (*nurse, doctor etc.*).

Work Complete – a place where work that is complete (or otherwise "finished") leaves the model (*patient leaves*).

In order to capture complex patient-flows through a CCU, extensive use was made of *Visual Logic* within the Simul8 software. The use of Visual Logic acts as an internal programming language allowing the modeller to incorporate rules governing the movement of entities and behaviour of objects within the simulation that are outside the scope of standard commands and parameters. For example, Visual Logic can be used to test for a number of conditions before deciding whether to admit an arriving patient. Figure 8.1 shows a Simul8 Visual Logic window for a work centre encapsulating the necessary code governing the rules of early discharge of patients on the CCU in order to free beds for arriving patients.

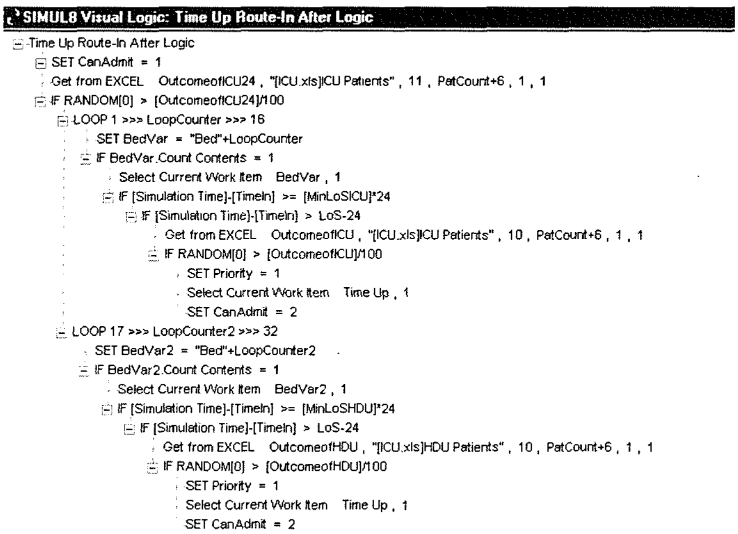


Figure 8.1: Visual Logic code

Simul8 may also be linked to MS-Excel using a Visual Basic connection. This has the advantage of using Excel as a user-friendly front and back-end to the simulation model. Simul8 currently has very poor facilities for displaying user-specified results and is particularly limited in allowing end-users to easily change parameter values without the need to reference objects and entities within the model. Standard Visual Logic commands can send and receive information to and from specified cells in the spreadsheet. Visual Basic code built within Excel, employing the use of a Simul8 library of functions, may be used to perform a number of operational activities, for example instructing Simul8 to start a run, end a run or simply change the parameter values of various Simul8 objects.

8.4 Visual Interactive Simulation (VIS)

Visual interactive simulation (VIS) originated with the discrete event work of Robert Hurrion in the late 1970's (Hurrion, 1978). Before that time, although simulation was seen by many as an attractive tool in Management Science and Operational Research, it was hampered by the complexity and lack of transparency inherent to many of the models used (Bell and O'Keefe, 1987). In addition, the limitations of computer technology had contributed to a general lack of accessibility.

Before VIS, simulation models were often characterised as *black-boxes* which typically could not easily be verified, validated or even understood by many who could benefit from them. Accessibility to such models was generally limited to technical specialists within the larger organisations who were able to afford the computer technology required for their operations. VIS represented an important development since it effectively broke the black-box image of simulation. Use of computer graphics and animation, for example, allowed the flow of entities through the system to be visually represented on screen. Numerous studies have since demonstrated the importance of a visual aid in the decision making process (Dickson *et al.*, 1986 and Chau and Bell, 1995).

Since the 1970's, VIS has benefited from the confluence of many developments in computing and other external factors including:

- Continuing increases in computer performance.
- Rapidly falling costs of hardware.
- Cheaper and improved colour displays (VDU's and LCD projectors)
- Increasing number of commercial VIS products and services –current VIS packages on the commercial market include Witness, Simul8 and Arena.
- Increasing recognition of computer tools in management.
- Greater accessibility due in part to a recent fall in VIS software costs – for example, Simul8 is now sold at $495 per license or a one-off site-license price of $999 for educational establishments. As a result, Simul8 is rapidly becoming a popular tool to use on Management Science and OR degree courses.

8.5 A Simulation Model for an Individual CCU

8.5.1 Capturing the flow of CCU patients

The complex characteristics of an individual CCU, coupled with desired user requirements, indicated a need for a sophisticated local CCU capacity planning and management tool. Working alongside managers and consultants from a number of critical care units, a detailed understanding of patient-flows was acquired. The corresponding activity flow diagram is presented in Figure 8.2. Patient-flows are appropriately generic as specified in the evolved operational modelling framework of healthcare resources (section 4.4). Suitable use of model parameters allows the simulation to be fine-tuned to reflect local individual CCU conditions.

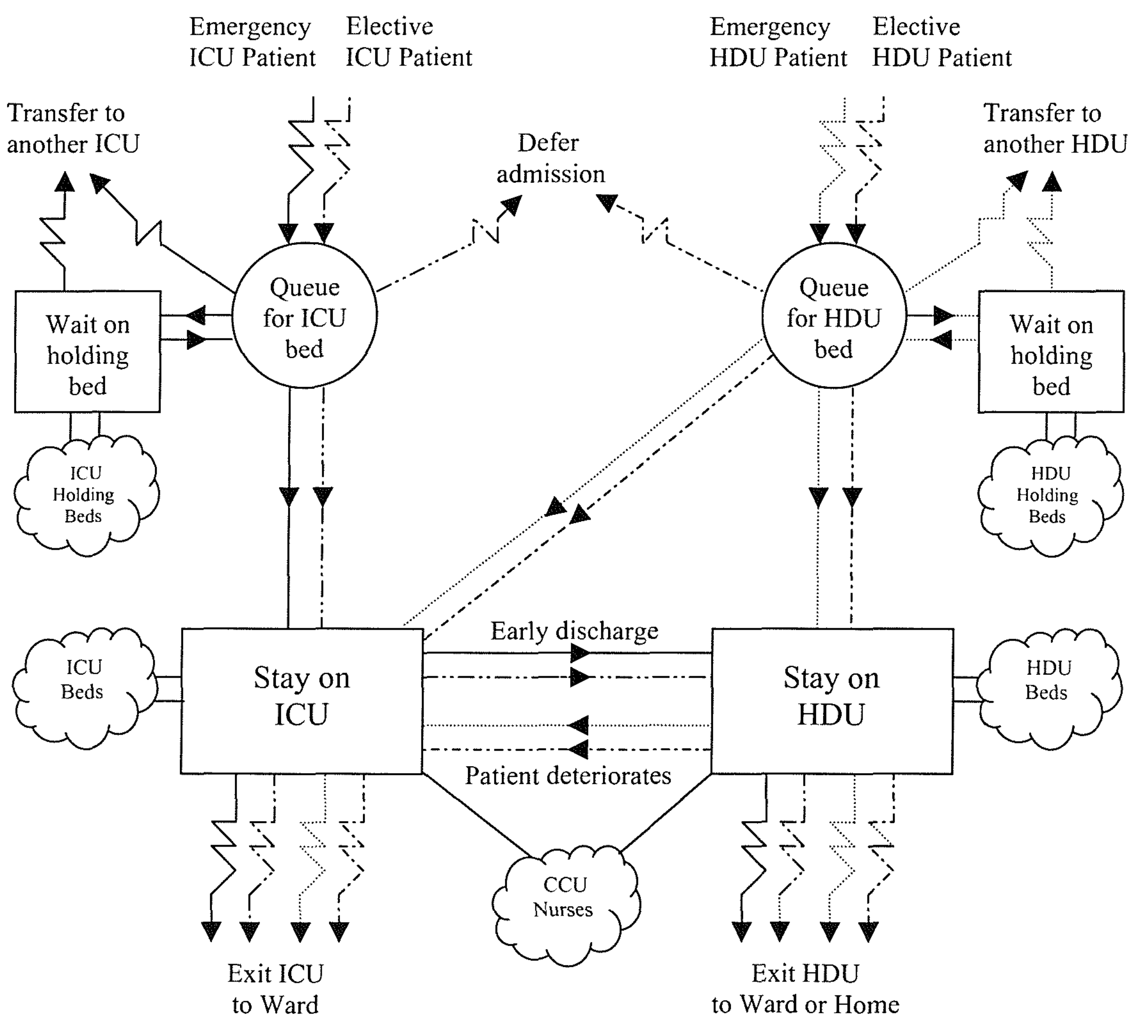


Figure 8.2: Activity flow diagram depicting patient-flows through a CCU

8.5.2 CCU configuration

The operational modelling approach used in the CCU model can help to evaluate the implications of various options for CCU patient care. *What if..?* scenarios may be examined by changing the following unit configuration parameters:

(A) For separate ICU and/or HDU, or combined CCU

- ***Beds*** – number of ICU and HDU beds either on separate units or combined CCU. Patients will stay in the bed for a sampled LoS unless they are early discharged (see below).

- ***Holding Beds*** – number of ICU, HDU or combined CCU holding beds (trolleys). A holding bed is used to accommodate arriving emergency patients when no main bed is available. During the time spent on the holding bed, every effort will be made to make available a bed by means of early discharge. If a bed becomes free then the waiting patient will be moved provided that they will survive their stay on the holding bed (a user-defined probability of death on the holding bed is necessary as this initial time on the unit is critical and holding beds typically do not provide the comprehensive facilities of a main bed on the unit. A separate probability of death once on the unit is also defined in the model).

- ***Maximum Time on Holding Bed*** – the maximum permitted time an emergency patient may spend on the holding bed whilst waiting for a bed to be made available. If at the end of this time no bed is free, the patient will be transferred out of the hospital.

- ***Number of Emergency-Only Beds*** – some beds may be reserved for emergency only patients (for example, one emergency-only bed on a six bedded unit: arriving elective patients will not be admitted if five or more beds are occupied).

- ***Number of Elective Deferrals*** – elective patients that cannot be admitted will be deferred. They will need to return at a later date (see below). Elective patients may only be deferred up to a permitted number of times before receiving upgraded status and high priority for admission.

- ***Deferral Waiting Time*** – The time between an elective being refused admission and re-attempting admission at a later date.

- ***Minimum LoS for Early Discharge*** – patients may be discharged from the unit provided they satisfy a number of criteria: they will survive; they are currently on their last day of LoS; they have stayed at least *X* days, where *X* represents a user-defined minimum LoS. The minimum LoS appears to vary between units (although is typically 2 days), hence the need to parameterise.

(B) For a combined CCU only

- ***Move to ICU from HDU*** – percentage chance that a high dependency patient will deteriorate and need intensive care.

- ***Move to HDU from ICU*** – percentage chance that early discharged intensive care patients will subsequently need to stay on HDU (alternative being to move directly to ward care).

The above unit configuration parameters are displayed on a worksheet in the Excel front-end so that they may be easily referenced and edited by end-users (Figure 8.3).

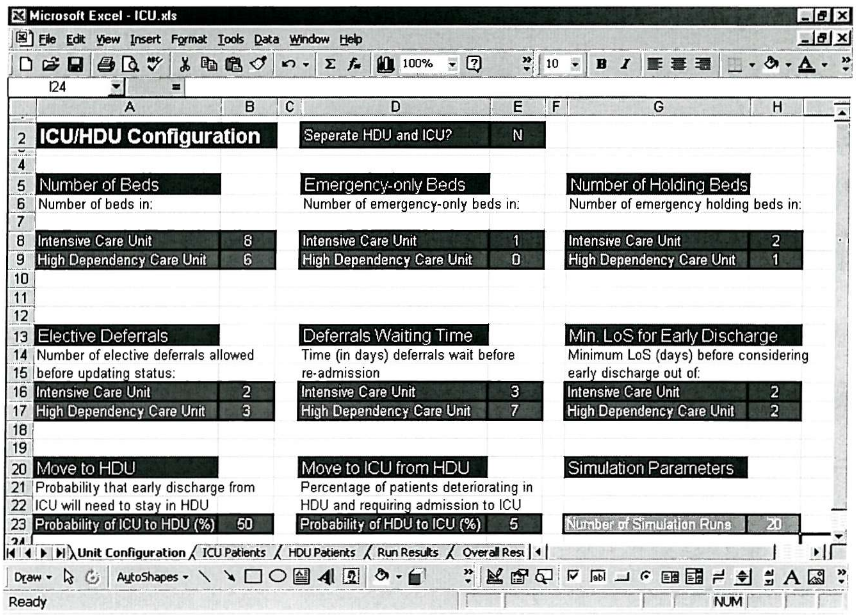
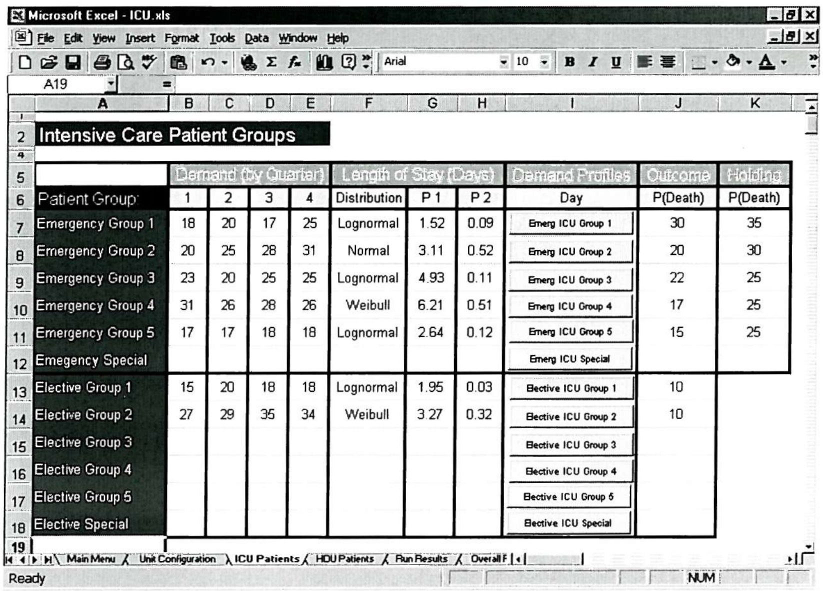


Figure 8.3: Excel front-end CCU configuration parameters

8.5.3 Patient profiling

In order to characterise the uncertainty and variability in CCU patient needs, the model allows up to six patient groups for each of ICU and HDU elective and emergency patients. The necessary statistical analysis is conducted in Apollo (Chapter 5), including the creation of these patient groups with demand profiles and LoS distributions derived for each group. Currently the CCU model reflects demand profiles through quarter yearly demand and daily profiles. Patient LoS is sampled from a number of statistical distributions (Lognormal, Weibull, Gamma and Normal). Two additional pieces of information are required: the probability of the patient dying during a stay on a holding bed (emergency patients only), and the probability of dying on the unit.

Intensive and high dependency patent information is included on separate worksheets within the Excel front-end. Figure 8.4 shows illustrative intensive care patient information with five emergency and two elective groups.



Intensive Care Patient Groups

	Demand (by Quarter)				Length of Stay (Days)			Demand Profiles	Outcome	Holding
Patient Group	1	2	3	4	Distribution	P 1	P 2	Day	P(Death)	P(Death)
Emergency Group 1	18	20	17	25	Lognormal	1.52	0.09	Emerg ICU Group 1	30	35
Emergency Group 2	20	25	28	31	Normal	3.11	0.52	Emerg ICU Group 2	20	30
Emergency Group 3	23	20	25	25	Lognormal	4.93	0.11	Emerg ICU Group 3	22	25
Emergency Group 4	31	26	28	26	Weibull	6.21	0.51	Emerg ICU Group 4	17	25
Emergency Group 5	17	17	18	18	Lognormal	2.64	0.12	Emerg ICU Group 5	15	25
Emegency Special								Emerg ICU Special		
Elective Group 1	15	20	18	18	Lognormal	1.95	0.03	Elective ICU Group 1	10	
Elective Group 2	27	29	35	34	Weibull	3.27	0.32	Elective ICU Group 2	10	
Elective Group 3								Elective ICU Group 3		
Elective Group 4								Elective ICU Group 4		
Elective Group 5								Elective ICU Group 5		
Elective Special								Elective ICU Special		

Figure 8.4: Patient group information

Entering and editing daily demand profiles for each group is facilitated by the use of daily variation graphs (Figure 8.5). The user may drag-up or down the bars on the graph as necessary until the desired demand profile is depicted. Based on the simulation time during the year, patient group inter-arrival times are calculated using the total yearly demand, current quarter yearly demand and relative daily demand (see section 6.5 for greater detail).

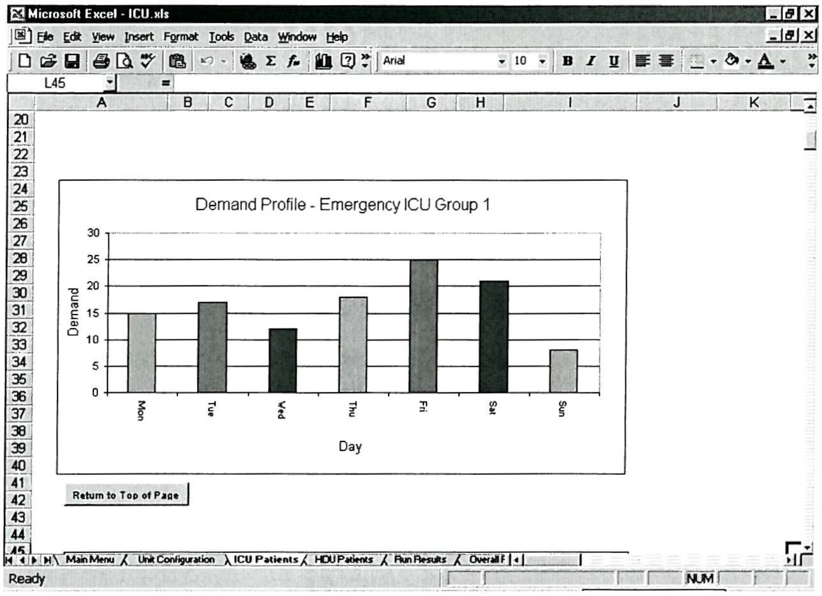


Figure 8.5: Excel graph capturing patient daily demand profile

8.5.4 Nursing constraints within a combined unit

Within a combined CCU, admissions are restricted by both the number of beds (accounting for emergency-only beds) and nursing constraints. Because of their critical nature, intensive care patients require a one-to-one ratio of nursing care. High dependency patients are currently given a one-to-two ratio of nursing care (one nurse can care for two patients).

Consider a combined CCU comprising of I intensive care beds, H high dependency beds and E emergency-only beds. At most the unit will have ($I + 0.5H$) nurses. Let i denote an intensive care patient and h a high dependency care patient staying on the unit. The CCU must satisfy the following conditions:

$$i + 0.5h \leq I + 0.5H \quad \text{(Nursing constraint)}$$

and

$$i + h \leq I + H \quad \text{(Bed constraint)}$$

Furthermore, for all arriving elective patients:

$$i + h \leq I + H - E \quad \text{(Emergency-only beds constraint)}$$

8.5.5 Running the Simul8 model

Having configured the CCU and provided the necessary patient information, the Simul8 model may be run for one year for a user-defined number of simulation runs. Extensive use of animation within the package allows the user to monitor the CCU status throughout the runtime. A number of statistics are displayed during each simulation run, including the number of patients who have survived and died, the number of patients on holding beds and numbers of deferred elective patients currently waiting to be re-admitted. Figure 8.6 illustrates the Simul8 model during runtime for separate intensive and high dependency care units with 8 and 6 beds respectively.

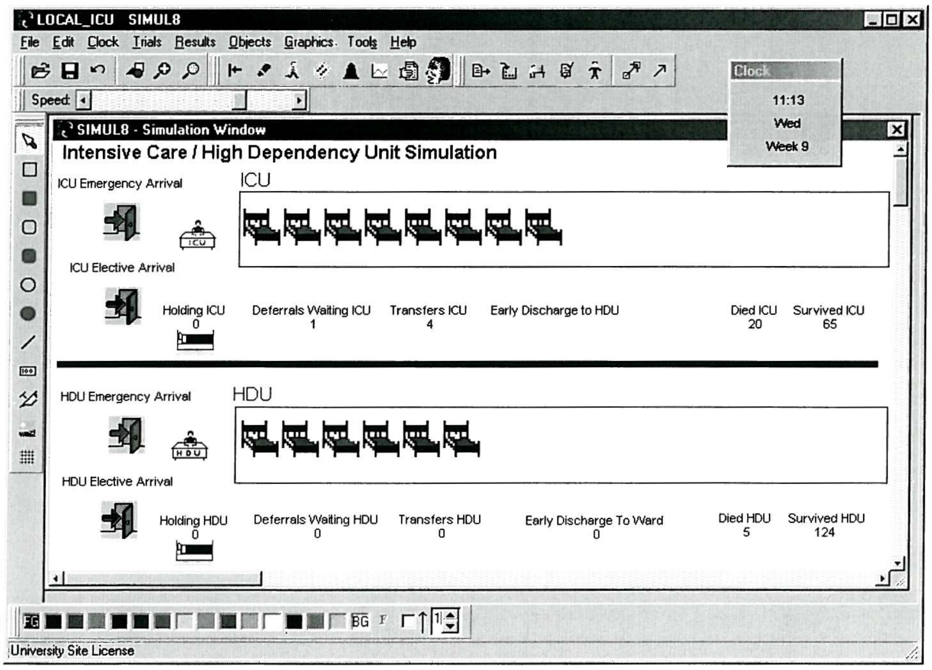


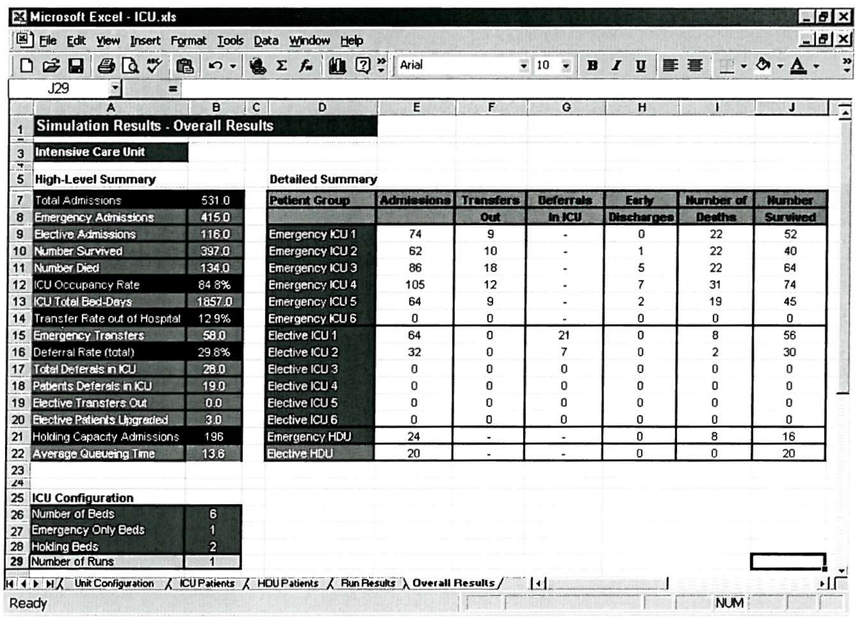
Figure 8.6: Individual CCU Simul8 model

8.5.6 Simulation results

A number of statistics are collected during each simulation run and sent back from Simul8 to the Excel back-end on competition of each run. Statistics are grouped at the intensive care and high dependency care levels, with a detailed breakdown for each patient group. Statistics include:

- Admissions, number survived and number died.
- Bed-days used and unit occupancy rates.
- Emergency transfers and elective deferrals.
- Holding bed admissions and average waiting time.
- Number of early discharges.
- Number of elective patients with upgraded status.

Figure 8.7 shows the Excel overall simulation results worksheet with illustrative output. A run results worksheet displays a results table for each simulation run.


Microsoft Excel - ICU.xls

Simulation Results - Overall Results

Intensive Care Unit

High-Level Summary

Total Admissions	531.0
Emergency Admissions	415.0
Elective Admissions	116.0
Number Survived	397.0
Number Died	134.0
ICU Occupancy Rate	84.8%
ICU Total Bed-Days	1857.0
Transfer Rate out of Hospital	12.9%
Emergency Transfers	58.0
Deferral Rate (total)	29.8%
Total Deferals in ICU	28.0
Patients Deferals in ICU	19.0
Elective Transfers Out	0.0
Elective Patients Upgraded	3.0
Holding Capacity Admissions	196
Average Queueing Time	13.6

Detailed Summary

Patient Group	Admissions	Transfers Out	Deferrals In ICU	Early Discharges	Number of Deaths	Number Survived
Emergency ICU 1	74	9	-	0	22	52
Emergency ICU 2	62	10	-	1	22	40
Emergency ICU 3	86	18	-	5	22	64
Emergency ICU 4	105	12	-	7	31	74
Emergency ICU 5	64	9	-	2	19	45
Emergency ICU 6	0	0	-	0	0	0
Elective ICU 1	64	0	21	0	8	56
Elective ICU 2	32	0	7	0	2	30
Elective ICU 3	0	0	0	0	0	0
Elective ICU 4	0	0	0	0	0	0
Elective ICU 5	0	0	0	0	0	0
Elective ICU 6	0	0	0	0	0	0
Emergency HDU	24	-	-	0	8	16
Elective HDU	20	-	-	0	0	20

ICU Configuration

Number of Beds	6
Emergency Only Beds	1
Holding Beds	2
Number of Runs	1

Unit Configuration / ICU Patients / HDU Patients / Run Results / Overall Results

Figure 8.7: Illustrative simulation results

8.6 Validating and Verifying the Simul8 Model

As a necessary part of simulation development, validation and verification techniques have been utilised to ensure that the CCU simulation is behaving in the desired manner and to increase confidence in the model operations. The working group contributed to the model development at all stages and ensured that the simulation model reflected the original conceptual schema. A number of validation and verification techniques, as described by Sargent (1991) and discussed in section 6.6, were adopted.

Simul8 represents the system graphically on screen and the flow of work around the system is animated so that the appropriateness of the model can be assessed and the behaviour of the system's objects over time examined. This is a major benefit of VIS tools in aiding the validation and verification process. By stepping through the simulation, moving entities (patients) through the CCU one-by-one, it was possible to check that the Visual Logic code governing the flow of patients was functioning in the desired way.

8.6.1 M / G / s / GD / s / ∞ *(blocked customers cleared) queueing model*

In addition to various verification techniques, including degenerate tests, extreme-condition tests and tracing, a simplified version of the Simul8 model has been compared to an analytical queueing model. In many queueing systems, an arrival who finds all servers (beds) occupied is, for all practical purposes, lost to the system. We call such a system a *blocked customers cleared*, or BCC, system. Assuming that inter-arrival times are exponential, such a system may be modelled as an $M / G / s / GD / s / \infty$ system. It would be too complex to formulate and solve the necessary analytical model to capture all of the complex rules governing real-life flows of patients (for example, incorporating early discharge rules, holding trolleys and movements between HDU and ICU). Hence simulation model parameters have been tuned to represent a simple single unit system described below:

Model description

- A single unit with admissions taken from a single queue.
- Arrivals are taken from the queue on a FCFS basis.
- Inter-arrival times are assumed to have negative exponential distributions.
- Once all beds are occupied, subsequent arrivals are transferred out of the hospital.
- All patients have a mean LoS of 4.0 days.
- There is a patient demand of 600 per year.

Analytical solution

Assume that at time 0 all beds are free. The probability that n out of the s beds are occupied is given by Erlang's Loss formula:

$$P(n) = \frac{\rho^n / n!}{1 + \rho/1! + \rho^2/2! + \dots + \rho^s/s!}$$

where

$$\rho = \frac{\lambda}{\mu}$$

such that

$$\sum_{i=0}^{s} P(i) = 1$$

It follows that in a unit with s beds, the percentage chance that a patient will be transferred occurs when all beds are occupied, i.e. P(s)*100%.

Queueing model results and simulation comparison

The queueing model has been used to find the transfer rate in a unit with varying numbers of beds. The analytical solution has then been compared to results from the corresponding simulation model (Table 8.1).

Table 8.1: $M / G / s / GD / s / \infty$ queueing model and Simul8 model results

s	% Transfers – Queueing Model	% Transfers – Simulation Model
5	39.9	38.5
6	30.4	29.8
7	22.2	21.3
8	15.4	15.5
9	10.1	10.4
10	6.3	5.7
11	3.6	3.2

The simulation model performs very well against the analytical solution. The Simul8 model was also configured to mimic the real-life complexity of the CCU in question, including the use of emergency-only beds, patient groupings and early discharges. Using recorded data from the unit, the observed transfer rate was calculated and compared to the results of the Simul8 model (15.9% and 15.5% observed and simulated transfer rates respectively). This is an additional and necessary verification process.

Validation, through the use of the generic framework and the adopted evolutionary model development (section 4.2), and verification methods such as the comparison with the above queueing model and observed data, help to increase confidence with the Simul8 model and ensure that it is sufficiently accurate and necessarily detailed. It becomes possible to use the model to examine a number of options for the care of patients in a CCU.

8.7 Model Applications

The individual CCU Simul8 model has been used by a number of different Critical Care Units across the UK as a decision-making aid for consultants and managers. It has been used to examine various configurations of care at a local CCU level. An illustrative list of model applications is provided below. This is not intended to be an exhaustive list of possible uses.

- ***Bed capacities*** – the current Government has indicated a desire for combined units to replace existing divisions between ICU and HDU. There is a need for each unit to model current bed needs and the consequences, where relevant, of moving towards a combined bed-pool.

- ***Nursing needs*** – a combined CCU is restricted by both bed capacities and nurse availability. It is important to quantify the required nurse-mix and to understand how various nursing levels impact of the unit's efficiency and effectiveness.

- ***Holding bed capacities*** – to appreciate the benefit of holding beds and to quantify how different numbers of holding beds affects transfer and deferral rates.

- ***The provision of emergency-only beds*** – to examine the relationship between the numbers of emergency-beds provided and the unit transfer, deferral and occupancy rates.

- ***CCU operating rules*** – examine in detail the consequences of changes to a number of unit operating rules, such as the impact of early-discharges and maximum time allowed on a holding bed.

- ***Re-configuration of care-pathways*** – in addition to modelling the redesign of separate ICU and HDU to a combined unit, the model may examine other configurations of care, for example allowing ICU patients to stay on HDU beds and preventing HDU patients from moving to ICU beds.

8.8 A Simulation Model for a Region of Co-operating CCUs

There is currently a great need to better plan and manage CCU beds at a regional level. The Department of Health's review of adult critical care services highlighted the lack of co-operation between units within a region. Improved planning and management of resources within a region of co-operating units has the potential to greatly benefit both CCU healthcare professionals and patients alike. For example, given that two neighbouring units both reserve one emergency-only bed, the consequences of just one "floating" emergency-only bed shared between the two units needs to be fully examined and understood. This scenario would likely cause a drop in the elective deferral rates (increased capacity for elective patients) but could unwittingly cause a small but significant rise in the emergency transfer rate. Such a critical issue demonstrates the power and value of an operational modelling approach. It is possible for managers and consultants to simulate various polices on a computer as opposed to potentially placing lives at risk by real-life experimentation with unit configurations and admission rules. Furthermore, often within a region of CCUs (such as those in a large city), some units are designated specialist centres for the treatment and care of patients with critical conditions such as major head injuries. It is therefore important to capture this detail in a model of co-operating units by simulating the transfer of specialist group patients from one unit to another.

Given the desired needs of the working group and the real-life complexities surrounding the flow of patients within a region of CCUs, it was decided to build on the foundations of the individual Simul8 model and evolve a model for up to six units within a given area. The regional model follows the same basis as the individual unit model, namely that the same configuration parameters must be defined for each of the six units together with the necessary information on emergency and elective patient group demand profiles, LoS distributions and survival probabilities.

In order to incorporate the role of the specialist centres, information for a number of specialist patient groups are entered on to a worksheet in the Excel front-end. Figure 8.8 illustrates a patient group for head injuries and includes the expected annual number of patients, LoS distribution and a preference matrix of which unit they should be sent to (more than one CCU may admit major head injuries and so the user may express an ordered preference to where patients should be sent).

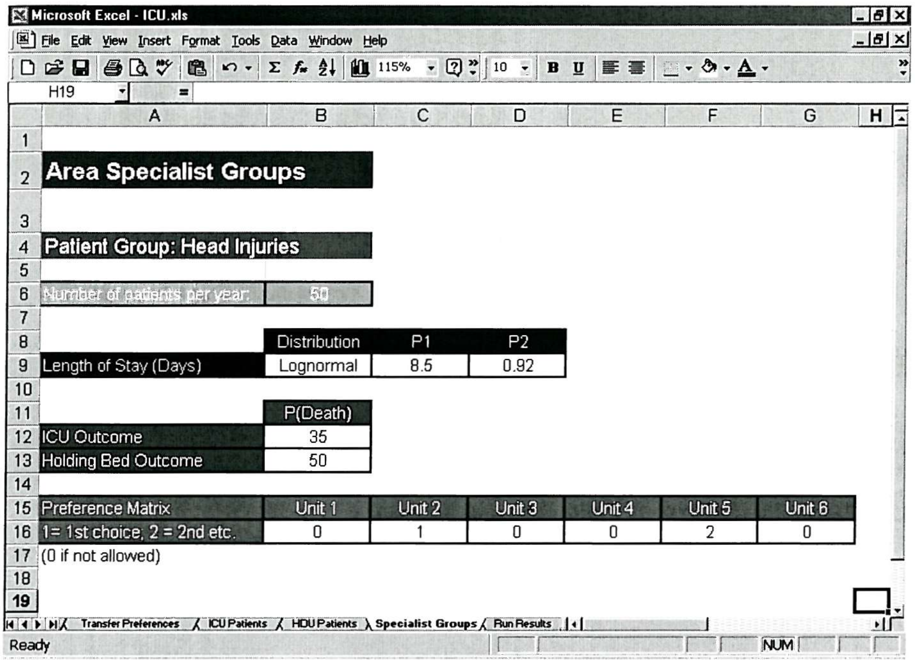


Figure 8.8: Specialist patient group worksheet

Within the individual CCU model, transferred patients will be sent to another hospital. In practice this will be to the nearest CCU with an available bed. The movement of transfers within an area is included in the regional model through a transfer preference matrix. These preferences will typically reflect distance, or other factors, between transferring units. Preferences are numerically ordered (first choice, second choice etc.) with transferred patients attempting to find an available bed in a unit starting with the highest priority (preference). An Excel worksheet captures the user-specified preference matrix (Figure 8.9).

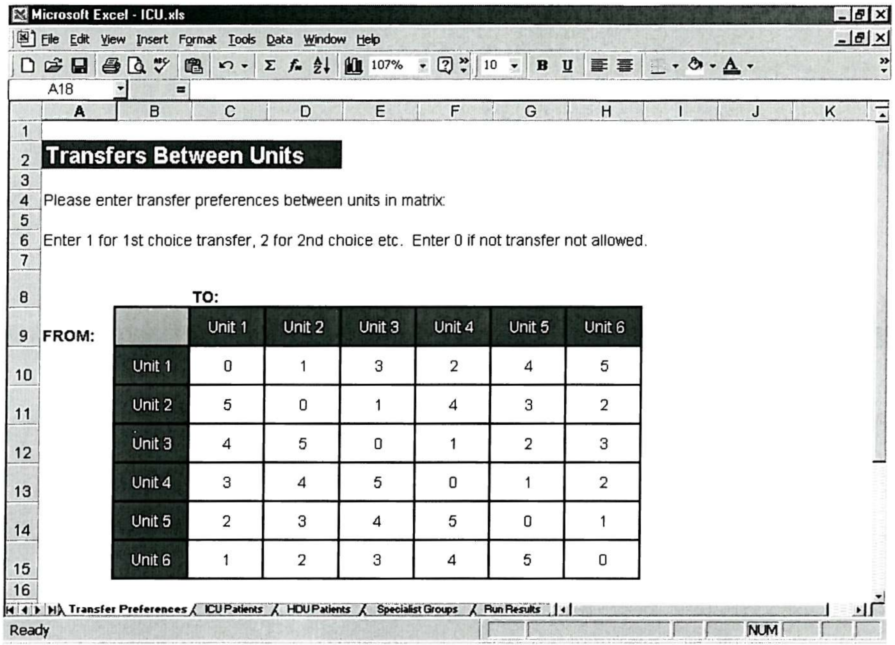


Figure 8.9: A preference matrix governing the movement of transfers between units

If no suitable available bed can be found in the area, the patient will be transferred to a hospital outside of the region. This statistic is displayed on screen when running the model (Figure 8.10). Other statistics, as indicated in the individual model, are shown on a run-by-run basis in the Excel back-end.

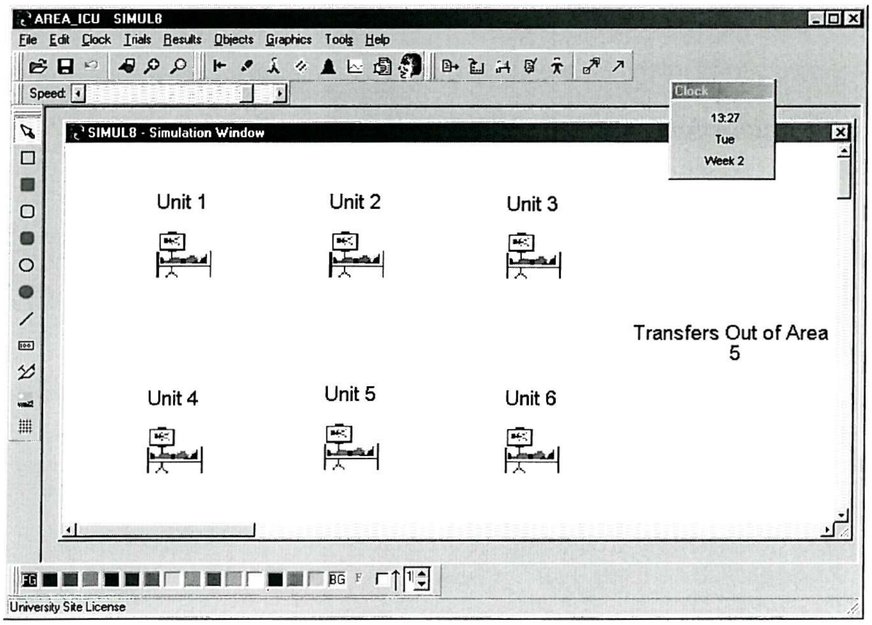


Figure 8.10: Regional CCU Simul8 model

8.9 Advocating Change Through Improved Co-operation

The regional CCU Simul8 model has been used to highlight a number of ways in which CCUs could improve patient-care through improved co-operation amongst a number of regional units. Using "hands-on" workshops and demonstrations, the model has helped consultants and managers to appreciate the value of co-operation and to fully understand and quantify the benefits. The regional model has shown that careful planning and management of a number of CCU variables can significantly impact on the effectiveness and efficiency of a number of co-operating units. These include, but are not limited too:

➢ ***Improved planning***

- *Unit re-configuration* – planning and provision of beds for each unit whilst acknowledging patient case-mix across the region and specialist centres. Consequences of a mixture of combined and separate units and the effect of changes in bed and nurse capacities on the care of patients within the region as a whole.
- *Transfer preferences* – evaluation of changes to the transfer preferences of units and the impact on the necessary number of CCU resources.

➢ ***Improved management***

- *Shared bed capacities* – greater flexibility in the provision of beds between units provides a responsive environment in which bed supply more closely matches bed needs over time. For example, provision of one emergency-only bed between two units as opposed to one emergency-only bed per unit. This will help eliminate redundant capacity and reduce deferrals.
- *Flexible admission/discharge rules* – greater flexibility achieved through improved communication between units can help reduce daily transfer and deferral rates and correspondingly increase the overall occupancy rate. For example the early discharge or transfer of patients from one unit to another.

8.10 Chapter Summary

The extreme costs of critical care, coupled with the relatively few beds available and the critical medical condition of the patients admitted intensify the planning and management issues. A recent Department of Health review concluded that there is a great need to better plan and manage CCU services at both a local (individual) and regional level. In planning for CCU capacities we need information for answering the following types of question:

- What are the effects of changing the number of intensive care and high dependency beds in a particular unit?
- What are the effects of changes in the casemix of the patients?
- What are the effects of changes in the lengths of stay of the patients?
- What are the advantages, and disadvantages, of formal co-operative arrangements for the care of patients in a group of units?

To help meet this need and provide the necessary detailed information, two detailed models for the planning and management of CCU services have been developed within the Simul8 package, an off-the-shelf simulation tool. The first model considers resources within individual units whilst the second models co-operating regional units. The developed models take individual patients through time as they pass through the complex care-pathways within the CCU.

Both simulations have been utilised in a number of ways by the working group (a team of CCU managers and consultants who have guided the work) to aid them in evaluating the implications of various options for patient care. With these models it is possible to have any arrival patterns, for example, winter pressures, and any casemix of patients. Thus the models can be tuned to capture the important, different, features of the various critical care units, the current practice, and possible changes in the current practice, can be reflected in the models. The modelling process has highlighted different approaches to help improve patient care by taking appropriate action at the individual unit level and through improved co-operating between regional units.

Chapter 9 – Moving Forward: Challenges and Opportunities

9.1 Chapter Introduction

This chapter draws together the research themes of previous chapters and aims to provide a synthesis of the research, practical work and issues raised. In many respects the work may be regarded as a feasibility study into the potentials of operational modelling for the planning and management of healthcare resources. During the period of research, working alongside a number of healthcare personnel from the participating hospitals, a number of central themes have emerged. The purpose of this chapter is to explore these issues representing both challenges and opportunities, in order to provide a basis for tentative conclusions about the current state of modelling for healthcare. Against the backcloth of the literature review, and based on personal experiences of working with the hospitals and the lessons learned from the research work itself, an attempt is made to present a framework for the successful design and implementation of operational models in a healthcare environment.

9.2 Identifying Critical Issues and Challenges

During the course of the case study work, a number of issues defining the primary challenges faced by the modeller in this field have emerged. It is fair to say that their resolution determines the likely success or failure of healthcare modelling in general. This section explores these key challenges under a series of subject headings.

9.2.1 Complexity

Anyone who has experienced the richness of interaction and activity of a district hospital, for example, will readily empathise with the inherent complexity of the system. You cannot fail to be impressed by the elaborate organisational structure and systems entailed in its operation. Complexity exists at all levels of the health service, from the society of disparate individuals who comprise the staff and patients, to the organisational and strategic relationships that exist both within and between the service providers.

The field in which the NHS is operating is in constant turmoil. It is absurd to assume that the delivery and planning of healthcare will remain static in this country. The political and socio-economic elements of the healthcare system give rise to the need for structural reform. In the UK there is little doubt that the complexity of healthcare management has been compounded by government led changes. Some of the multi-dimensional factors contributing to the general complexity of the system include:

- Demographic change
- Social change
- Organisational change
- Political change
- Strategic change
- Technological change
- Clinical change
- Inherent variation and uncertainty in treating individuals

Often a change at one level of an organisation can impact in an unforeseen way on other levels of operation. For example, changes to the number of beds will impact on the number of necessary theatre sessions. Thus organisational and strategic changes cannot be considered in isolation from the complex network of inter-and intra-relating hospital services.

In such circumstances, it is important to ask whether quantitative operational tools, such as the simulation models presented in this thesis, can play a useful role in the context of healthcare management. It is evident that a common current practice is to plan and manage hospital capacities through a simple deterministic spreadsheet calculation using average patient flows, average needs, average length-of-stay, average duration of surgical operations etc. Mathematically speaking, a hospital corresponds to a complex stochastic system so that the common deterministic approach for planning and managing the system can be expected to be inadequate. It has been shown that the deterministic approach will typically underestimate hospital requirements. With great advances being made in computing power and technology, it is fair to conclude that the need to create simplistic unrealistic models is unnecessary and outdated. The mathematical modelling approach of OR, and in particular simulation methodology, is ideal for dealing with complexity, uncertainty, variability, constraints, and scarce resources. Appropriate models can avoid the dangers of planning on the basis of average values only. This research has identified the benefits and opportunities of the development, solution, and validation of sufficiently detailed stochastic models for planning and managing hospital capacities.

9.2.2 Diversity

Diversity exists at all levels in the health service and forms a central aspect of the NHS culture. It is a reflection of a number of factors including the diversity between individuals at work within the NHS, between the management of the service providers, between the information systems employed and between the populations that each institution seeks to serve.

The majority of NHS management is borne by the administration of local service providers themselves. For example, a hospital will plan and manage patient care at a local level, independently of other hospitals in the UK. Current Labour Government policy appears to be promoting this cause, with policies directed at giving greater managerial flexibility at a local Trust level. It may be argued that this helps to ensure that past inefficiencies of the NHS are avoided by giving increased responsibilities and incentives, together with the penalties that come with poor performance, to effectively privatised parts of the healthcare system. The result however is that healthcare institutions, although conforming to national standards, often have their own idiosyncratic methods and operations. It is amazing to contemplate that between hospitals, even separated in distance by a few miles, the differences are clearly evident. For example, there are different ranges of specialities, different arrangement of wards and different ways of collecting and analysing data within the Trusts.

From a modelling perspective, this represents a clear challenge. One approach could be to focus research at the level of individual service providers and to limit findings to particular institutions. This approach seems to have been adopted by the vast number of papers reviewed during the literature search as detailed in Chapter 3. The major downside to this approach is the large overheads entailed in producing institutional specific models with irrelevance to a wider NHS context. Modelling tools that are targeted for specific areas of healthcare application are limited by the focused scope of the models. Unfortunately many of the proposed models suffer from this fate of over-specificity to one scenario of care.

More generic models, that may be applied to a number of healthcare providers and used to model a number of patient-care scenarios, bring obvious benefits. The development of the generic framework for the modelling of hospital resources exemplifies one variant of the generic approach. Whilst acknowledging the challenges of diversity and complexity, models within this framework incorporate generic processes with the ability to fine-tune the model to reflect local conditions. Hence we obtain one operational model for many healthcare settings as opposed a hospital-specific suite of programs. Unfortunately there is little evidence of their widespread adoption at the management levels in the NHS. Hopefully over time, funding bodies such as NHS Executive and MRC will come to appreciate and fund more projects aimed at developing operational models to deliver research findings applicable to the widest possible healthcare audience. Such models represent excellent value for money in terms of grant expenditure versus advancement and transfer of knowledge across the NHS.

9.2.3 Credibility

An operational modelling approach is likely to be unfamiliar to the vast majority of healthcare managers and clinicians in the NHS. It is often consequently treated with caution and slight contempt. The adopted evolutionary development approach taken in this research goes some way to easing the fears of the client. This approach requires a constant dialogue with the end-users (hospital consultants and mangers). Models may then be created and enhanced alongside the potential users. It is fair to say that confidence in the model is only gained through verification and validation techniques. This ensures that the model is right (behaves as required) and that it is the right model (is modelling the true real-life system). This should be considered as a stepwise process; an ongoing dynamic in which both the modeller's and client's confidence is gradually increased. Hopefully there will come a time when the model is trusted enough by the client to accurately represent the real system to a degree that is sufficient for the purposes at hand.

Validation and verification must be recognised as fundamental to the modelling and development of healthcare resources. Given the political sensitivities that often accompany the decision making process in healthcare management (section 9.2.5), this becomes especially important. Managers and clinicians appear only to gain confidence and trust after the model has been used and is seen to be validated by mimicking observed results over time.

9.2.4 Data availability and quality

The old adage *"garbage in; garbage out"* is nowhere more relevant than in the context of modelling and simulation. A model is only as good as the data that informs its development. Within the healthcare domain, the range of data required varies enormously according to the nature of the model and its purpose. Confidence in the data however, is prerequisite to a model's credibility; a necessary condition (but not itself sufficient).

Although there is undoubtedly an enormous quantity of data collected in one form or another within the health services, it is less certain that this always forms the basis for useful decision support. More recently NHS Trusts have responded with the creation of departments given grand names such as *Hospital Strategic Analysis* or *Trust Information Team*. Now more than ever there is a need for *evidence-based* hospital planning and management by utilising the vast data supply to deduce meaningful hospital information for managers. One of the main factors inhibiting the efficient use however is the multiple classes of data that are collected and the wide range of interest groups and objectives that need to be served. Financial, contractual, clinical, personnel, operational, audit and strategic needs all place different demands on the data collection exercise. Partly in response to this need, the systems used for collection vary widely, ranging from manual systems such as nurse diaries and operating theatre logbooks, through to specialised departmental databases and hospital information systems.

In hospitals, the primary source of patient information is the Patient Administration System (PAS) or Patient Management System (PMS). This is used to routinely collect information on every patient who passes through the hospital. Typical fields include the patient name, address, date of birth, admission date, discharge date and diagnosis (HRG). Much of this data formed a central role in the case studies (Chapter 7). With such a large source of data, often in the region of 70,000 inpatient records of information per year for a typical medium sized NHS hospital, it is so bewildering and frustrating that the NHS often seems starved of information with limited use of evidence-based management at the Trust level.

The issue of data quality is clearly central to building confidence in the modelling process. Accuracy unfortunately can rarely be assumed. The recording of clinical data, for example, is an area where problems of classification are particularly acute. Likewise, the recording of medical activity, especially within hospitals, is fraught with problems. The use of the FCE (Finished Consultant Episode) as the standard activity measure for instance, has widely been criticised although an alternative is not forthcoming. There is a strong suspicion that FCE measurements can be too easily manipulated for contractual purposes, for example, by amplifying the total number of FCEs by transferring the same patient around different hospital specialities.

In many respects the problems of attaining high data quality in healthcare is a reflection of the inherent difficulties and complexities as discussed in previous sections of this chapter. Clinical coding for example is beset with difficulties and often coding clerks may have little clinical knowledge. Databases themselves are often incomplete with missing values and erroneous data entries in many of the fields. For example, it was common during the course of this work to find within provided datasets negative ages, clinical codes that do not exist and patients who were apparently discharged before they were admitted. Data quality represents a major challenge to the modeller. Extreme care must be exercised when handling and analysing hospital data.

9.2.5 Politics

The political dimension of healthcare plays a major role in the management and decision-making processes within the organisation. It seemingly permeates every level of the service and becomes quickly obvious to anyone involved in the management and planning of resources. The NHS is one of the best loved in principle, most vilified in debate and least understood parts of the welfare provision of this country. Funded largely from national taxes, central government has an intense interest in its well being, the resources it consumes, and the service it provides. This political element significantly impacts the way decisions are made and how they are presented. At local level public opinion often manifests itself as opposition to changes to its service. At the national level government policy is frequently a subject of much debate.

It is important that modellers understand the politics of the environment in which they are working. It is unlikely that operational models will be perceived in a neutral light given the pervasive nature of the political dimension in the NHS. Precautions must be taken to safeguard against distortion and misuse of the models, for example by managers trying to use the model to prove their own beliefs.

The political element was never more evident than when the model was being used to examine the way in which the operational effectiveness and efficiency of the system could be improved by changing the working culture of hospital consultants. There still persists in today's NHS a *them and us* working culture, with consultants often in conflict with specialty managers attempting to bring managerial changes to the running of wards. Hospital managers often saw the simulation as a tool to potentially influence change. It should be stressed that operational modelling tools have a large role to play in that they can quantify the impact of change and help in objective, as opposed to subjective, decision-making. There is however a need to be aware of any specific sub-group agendas so that the modelling process is not railroaded by external political factors.

9.3 Identifying Opportunities

The previous section has highlighted potential challenges that may impede an operational modelling approach for the planning and management of healthcare. The work however has also identified a number of opportunities for simulation modelling to make a significant impact on the way in which healthcare is delivered within the NHS.

9.3.1 Application

There is a vast range of areas in which simulation could be applied in healthcare. Even within a specific *area* of research, there are a number of *ways* in which simulation could be utilised. Some of these themes will be picked up within the further research discussion in Chapter 10. Below is a list of potential simulation projects within a hospital capacity modelling context. This list was evolved from discussions with participating NHS Trust managers.

- Planning bed numbers
- Workforce rostering
- Planning operating theatre needs
- Scheduling operations
- Management of ward beds on a day-to-day basis (*diary-planner*)
- Costing resources
- Contingency planning
- Management of waiting lists

- Resource planning for new configurations of care
- Planning and management of outpatient resources
- Outpatient clinic scheduling
- Resource planning for a health authority (movement of patients within a region)

In addition to the many *areas* of application, it is possible to define various *types* of its use. Model utilisation will often dictate which form of simulation is necessary. For example, VIS could be used when it is desirable to visualise the movement of patients; perhaps suitable for bed planning but unnecessary for a waiting list management simulation tool. Potential types include operational simulation (what-if scenario tools), educational simulation tools (to facilitate understanding of a particular system) and real-time simulations (decision support tools).

In addition to what may be termed *patient-flow* simulation models, which capture the movement of patients through healthcare organisations (hospitals, outpatient clinics, GP surgeries etc.), are *patient-progress* models which capture the movement of patients through the natural history of a disease or medical complication. The Institute of Modelling for Healthcare (IMH) has a wealth of experience in building models for this purpose. These including simulation tools to examine the care of patients with asthma and HIV/AIDS, and screening models for colorectal cancer and breast cancer. There is an endless list of possible applications for the modelling of diseases.

9.3.2 Emerging technologies

Advances in simulation software and technology are only now beginning to be applied in new areas of healthcare. This has been aided by dramatic falls in the cost of computing hardware. Access to powerful simulation tools is no longer restricted to large businesses with large research budgets. Instead, for as little as under £300, hospitals may purchase easy to use simulation packages for their own purposes, provided of course they have suitably qualified personnel to work them.

More recent advances in simulation include 'Virtual Reality' (VR) applications which are now becoming more widely used in healthcare, primarily to allow consultants to rehearse complex surgery on a three-dimensional computer-generated human body. Other applications of VR include 3D graphics built into simulation software to allow the user to model a hospital building (for example the WITNESS software by Lanner Group).

Real-time systems also have a major role to play in the future of healthcare management. For example, the use of decision support systems (DSS) to assist in the daily running of a hospital. Such tools represent a large opportunity to the OR community. Although there has been some evidence of their use in recent use (Rojas and Martinez, 1998), there is however a great need for similar and more advanced models for use in a wider healthcare setting.

9.3.3 Promoting the cause

Key to the success of the future role of operational models is the ability to extend the user base; to encourage and promote the methodology. This may be achieved through successful projects being given the acknowledgment they deserve in publications that provide the widest possible readership to healthcare professionals (e.g. BMJ). Academics tend to limit publications to more technical and subject specific audiences through mathematical, operational research and computing journals. There is a real need for the transfer of knowledge to potential healthcare "clients". Experience has shown that many managers are simply not aware of an operational approach and in particular that of simulation. It is only through discussions and presentations of existing models that they are able to understand and appreciate the potential benefits.

9.4 Selecting an Appropriate Simulation Tool

The research work as described in this thesis has utilised different simulation approaches for the modelling of healthcare resources. The deliberate choice of different techniques allows for a discussion on the advantages and disadvantages of each approach. In general, careful consideration should be given to selecting the appropriate simulation tool for the project in hand. The choice will likely depend on a number of factors, such as:

- Client preference
- Cost
- Time
- Size of simulation (estimated number of simulated entities) and complexity
- Available computer hardware (to develop on and that of the end user to run on) and software, and existing client IT infrastructure and support
- Expertise/skills of the modeller
- Speed of simulation (runtime)
- The importance of a visual representation (VIS)
- Whether the model will be handed over to client
- Ease of use by end user

In essence, the modeller must choose between building the simulation from scratch themselves (a *programming* approach) as opposed to making use of existing simulation tools (*off-the-shelf* approach). Furthermore, if the off-the-shelf approach is chosen, it may be possible for the modeller to adapt the model in the desired manner by incorporate some programming elements outside of the software package itself (e.g. combining Visual Basic with Simul8 via an Excel front and back-end).

The PROMPT model (Chapter 6) was developed using the programming approach. This was built within the TOCHSIM simulation shell using Object Orientated Coding in a Delphi software environment. The CCU simulation models (Chapter 8) however have been built using the commercial software package Simul8 and enhanced through Excel, Visual Logic and Visual Basic. Based on the author's experience of creating a number of healthcare simulation tools, the purpose of this section is to highlight the key benefits and limitations of both the programming and off-the-shelf approaches. These points are intended to act as issues to consider when proceeding with a simulation-modelling project. Clearly there is no right or wrong answer to the question of which approach to select. Instead, the final choice of tool will likely depend on the modeller's personal preferences, model utilisation, the factors bulleted above and the issues raised below.

9.4.1 The programming approach

✔ ***Benefits***

- The model can be design to meet the exact specification
- It may be tailor-made for the end user
- It does not require the client to buy commercial software – only the programmed executable is required to run the model
- The simulation may be readily designed to read in data from / output data to existing client systems
- Provided models are constructed using a suitable simulation shell (e.g. Three-phase or event-based approach), run-time speeds are typically very fast
- There are no constraints to the possible size of simulation (numbers of entities, resources, activities etc.)

! ***Concerns***

- Takes a long time to code and build models (more suited for mid to long-term projects rather than for quick solutions)
- VIS elements may take a long time to design and code
- Requires modeller with skills in computer programming

9.4.2 The off-the-shelf approach

✔ ***Benefits***

- Rapid time to design prototypes and final model
- Requires little programming skills as most packages use Windows 'point and click' technology (create the model by selecting appropriate simulation objects from the menu toolbar)
- Possible to build the model 'in front' of end user (design and draw processes directly onto the computer screen interactively with end user)
- Can enhance model by linking to other packages (e.g. MS-Excel which most clients are familiar with)

! *Concerns*

- Model may not meet exact user requirements (for example, limited functionality of objects within software may not mimic exact real-life processes)
- Difficult to give a tailor-made appearance (for example no main menu, problems with customised data input and summary sheets, limited number of graphs permitted within software package etc.)
- Requires the client to buy the commercial software (issues with cost, license problems and number of machines that will run the software)

9.5 Towards a Framework for Successful Operational Modelling for Healthcare

Healthcare modelling projects will typically be varied in their nature, according to the needs of the client and the hospital system under consideration. Based on the author's experiences of working with a number of healthcare clients, some general conclusions and issues however have emerged. The following list has consequently been evolved and is intended to provide points for discussion within a framework for the successful implementation of healthcare operational models. Each of these issues should be given the due attention they deserve, as they will ultimately determine the likely success or failure of the project. This framework is intended for use when embarking on a new study.

The thirteen discussion points are described below. A framework has been evolved to show the natural order and causal relationships between these issues. Each healthcare project may be divided into three high-level components: *Pre-model*, *Model* and *Post-model* stages. The presented framework illustrates the positioning of the thirteen issues within the modelling project as a whole, together with inter-dependences and natural orderings.

1. *Form a steering group*

Identify appropriate hospital personnel to form a group to steer the direction of the work. This group should meet at agreed regular intervals and should discuss project management issues. Members should be made up from relevant areas of the hospital who should act as key contact points for information or provide access to other staff. Appoint one member of the steering group to act as Chairperson.

2. *Conduct a feasibility study*

A critical part of any project: scope out the project and create a project brief to be agreed by the steering group. Identify what is required and crucially what are the project deliverables and time scales. Check for any potential obstacles such as lack of necessary data or access to key hospital staff. If deemed feasible, a comprehensive feasibility study lays an excellent foundation for a successful project.

3. *Spend time collecting the necessary data and information*

A considerable amount of time should be spent working on-site at the hospital and liaising with hospital "process-owners". Much of this contact may be informal and unstructured, although time should also be spent sitting on any existing and appropriate working groups and within process re-design teams. These meetings provide a rich source of insight into management processes at different levels and consequently add to the understanding of hospital. A range of structured methods should also used to elicit feedback as necessary. Such techniques could include questionnaires, structured interviews and more soft-system OR methodologies for brainstorming and cognitive mapping activities. Raw data may be obtained from the relevant hospital data sources (such as Patient Management Systems).

4. *Pay attention to data quality*

Having obtained the necessary data and information, never assume completeness or quality. Check issues that might have biased the contents (such as a sudden surge in day-case surgery to meet government targets). Carry out necessary preliminary data quality checks and analysis to identify any outliers, data entry errors or anomalies.

5. *Think carefully about the level of detail*

One tension that exists within healthcare modelling is the extent to which it is possible to abstract the processes of care and still retain a level of accuracy within a model capable of producing useful results. A trade-off needs to be recognised between the need to simplify processes and the need to retain detail to accurately reflect the real system.

6. *Select the appropriate tools*

Initially spend necessary time to decide on the most appropriate tool(s). The OR practitioner should call upon their *toolkit* of OR techniques and select those which best meet the needs of the problem with the end user in mind. For example this could involve simulation models (with careful selection of the best methodology as discussed in section 9.4) coupled with suitable data analysis packages and databases. Bear in mind the desired end product which will influence the appropriate tools. For example, if a graphic representation of the system is required then this will narrow the range of potential approaches.

7. *Design for wide use*

Develop models with flexibility in mind. For example, where possible parameterise variables so that they may be readily changed by end-users at the hospital and indeed to reflect conditions at other hospitals (potential clients). This level of flexibility avoids the misgivings of institutional specific models with irrelevance to a wider NHS context.

8. *Involve end-users at all times*

Maintain strong links with the client and make the necessary visits to the hospital to meet key personnel. Prototype models should be developed *alongside* end-users and not independently from them. A successful project involves a powerful combination of OR practitioners and medical experts. Be aware that if at the end of the work you are planning to propose change, there has to be a top-down emotional case for change as well as a bottom-up rational case for change. Hospital politics are likely to be entangled within this process of change (see later point).

9. Build credibility

Spend time getting to know key personnel and understand their role within the organisation. Presentation of prototype models to targeted individuals will build credibility over time if the models can be seen to be meeting expressed user needs and reflecting real-life processes.

10. Acknowledge the politics

Seek and acknowledge any political sub-agendas. Maintain a non-biased objective view throughout the work. Try to keep a safe distance from political issues but understand and accept political relationships within the organisation. Use these to your own benefit.

11. Allocate resources

A successful project will need the necessary resources allocated to the project both within the hospital and back at base (practitioners organisation). These are not only in terms of human resources but should include computing hardware, software and other resources as relevant.

12. Foster relationships

Keeping the client happy with good working relationships and a successful outcome will likely lead on to future projects with the same client or useful endorsements when approaching other healthcare organisations.

13. Promote the results

Ensure that successful projects are promoted both within the hospital and to a wider healthcare audience. This may be achieved through presentations of the work at healthcare meetings, conferences and through publications.

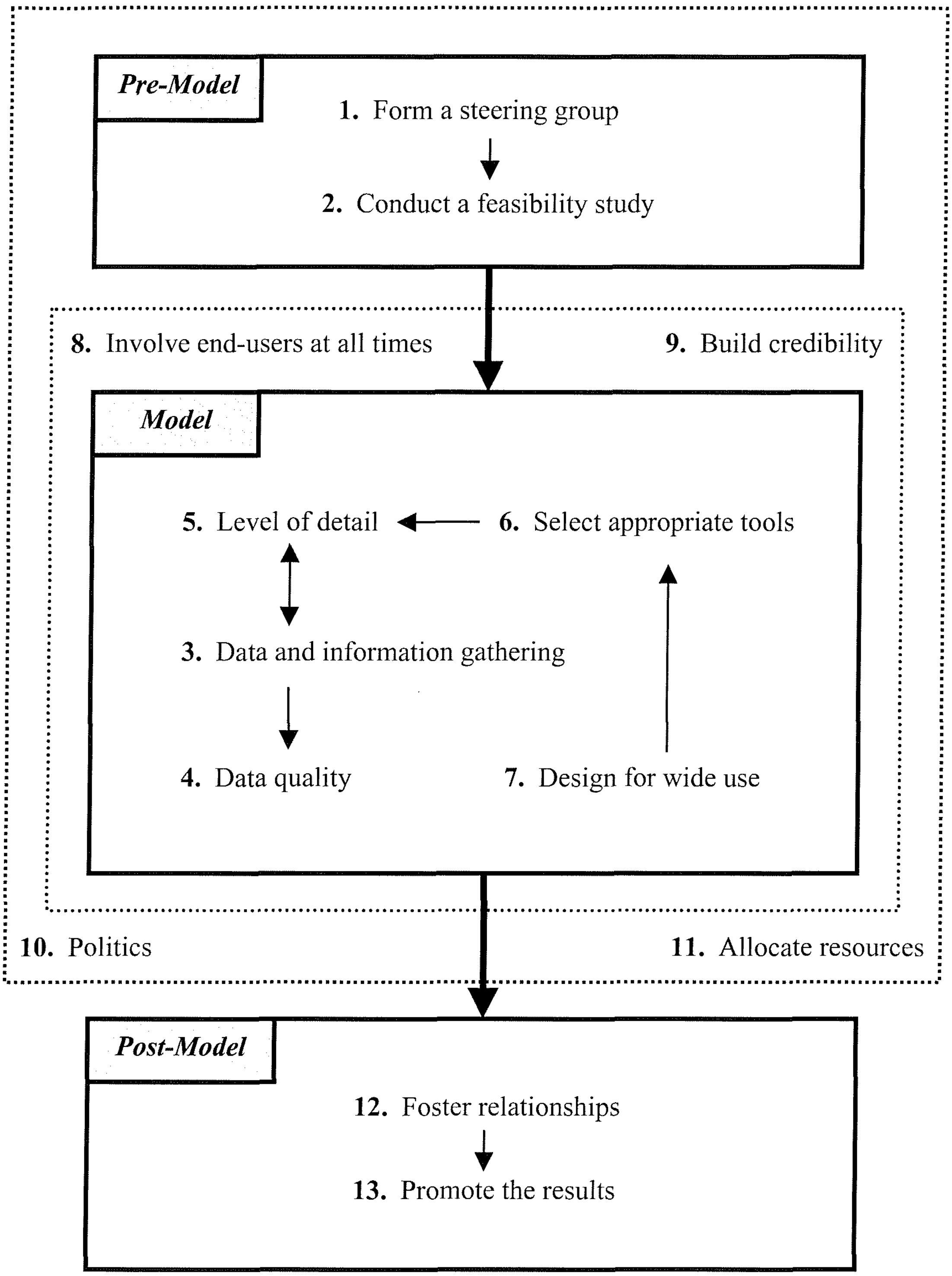


Figure 9.1: A causal diagram for successful implementation of healthcare models

9.6 Chapter Summary

Healthcare modelling is beset with many challenges. It is important that careful consideration is given to these issues, ranging from politics through to data quality and availability. In order for a healthcare operational modelling project to succeed, a number of points should be borne in mind both before commencing and during the work with the healthcare organisation. A framework towards successful implementation has been evolved through the author's own experiences. This shows that many of the inherent challenges may be overcome by adopting suitable working methods and standards. Most of these reflect issues that are unique to, or intensified in, the healthcare environment.

With greater emphasis now on efficient and effective healthcare to meet the needs of our nation, these are surely exciting times for healthcare OR. There is so much to be done and endless possibilities and opportunities. It is crucial however that the necessary steps are taken to ensure that the future generation of work meets the true needs of managers and clinicians. In particular, careful thought should be given to selecting the appropriate operational tools. Successful projects should be promoted to a wider healthcare audience.

Chapter 10 – Conclusions and Further Research

10.1 Chapter Introduction

The provision of healthcare services is perhaps one of the largest and most complex industries worldwide. As one of the essential necessities to sustain life, it faces the consequences of increasing demand in times of limited financial resources and competing social needs.

The current policy in the United Kingdom focuses on appointing responsibilities and decision making to regional and district health authorities rather than the centralisation of power. It is expected that in this way the needs of the population in each region will be successfully met, thus leading to the improvement of the NHS performance. Providing the appropriate medical care involves decision-making in terms of planning and management of the healthcare services and the resources contain within. However, such decisions occur in a larger context that includes ethical, economical, social and legal considerations. Issues surrounding planning and management of healthcare services are varied and include:

- The need for appropriate level of care.
- Supply of adequate resources, including beds, operating theatres and workforce.
- Measurement of resource consumption.
- The need of effective monitoring of the provision of care.

To help meet these needs, the research as described in this thesis has explored many issues and has proposed various frameworks and models for use within the healthcare profession. The purpose of this chapter is to review the ground covered and to outline the future direction of research in this field.

10.2 Thesis Review

This thesis concerns the planning and management of healthcare resources. The purpose of this section is to discuss whether the research objectives, as discussed in section 1.2, have been met and to summarise the main conclusions from the research work.

10.2.1 Simulation methodology as a tool for decision support

Within the healthcare environment, it is clearly apparent that there is a great need for detailed quantitative management tools to aid decision-making. This has been evident through the expressed needs of steering groups and through the magnitude of possible model applications, some of which have been described through case studies (Chapter 7). To many in the profession, the use of operational tools such as the adoption of a simulation methodology represents a new and untested concept.

A common current practice is to plan and manage hospital capacities through a simple deterministic spreadsheet approach using average patient-flows, average needs, average length-of-stay and average duration of surgical operations. Patient-flows, patient needs and utilisation of hospital capacities involve complexity, uncertainty, variability, constraints and scarce resources. Mathematically speaking, a hospital corresponds to a complex stochastic system so that the common deterministic approach for planning and managing the system can be expected to be inadequate.

Unlike deterministic models, stochastic models provide a more accurate and realistic model by incorporating uncertainty and variability through the use of probabilities and random variables. In particular it has been shown that such models contain desirable properties for the modelling of healthcare resources, in which the time between transitions of states can occur after any positive time spent in a state and where this transition time can depend on the transition that is made.

The analytical solution of stochastic models, as necessary for the modelling of hospital capacities, presents a formidable challenge. Unless restricting assumptions are applied, many stochastic models are impossible to solve analytically. Numerical methods are needed for solving realistic stochastic models. A computer simulation need not make such stringent simplifying assumptions and is an ideal tool for modelling hospital capacities. It is hoped that this thesis has demonstrated the power of computer simulation in support of management decision-making processes for a variety of healthcare resources.

10.2.2 Classification techniques

The research has explored the use of classification techniques for the creation of patient groupings. Necessary patient groupings may then be fed into developed simulation models and individual patients from each group passed through the particular healthcare system of concern. In order to capture the uncertainty and variability amongst the patient population, a number of classification techniques have been considered and evaluated for their relative performances and practical usefulness. Research has shown that there is not necessarily a single *best* classification tool but instead the best technique will depend on the features of the dataset to be analysed. The research has made a start in investigating what these features are with particular emphasis on healthcare data.

A survey of healthcare staff has however revealed that tree-based tools, such as CART, do have a greater practical appeal than that of the other tested techniques. This is a measure of the extent to which the CART algorithm produces comprehensible results that are generally easier to interpret by medical staff than the results of other algorithms and on the time it took for hospital staff to understand the technique, prepare the data and actually perform the analysis to produce correct and meaningful results.

A statistical package, Apollo, has been developed as part of the evolved generic framework for modelling healthcare resources. Apollo incorporates the CART tree-based algorithm, that assists in the production of clinically and statistically meaningful healthcare groupings. For example, these could be patient groupings based on LoS, operation times or survival rates. Derived groupings may be automatically saved and fed in to developed simulation models within the framework.

10.2.3 Evolving a generic framework

A generic framework has been evolved in the light of perceived user-needs and real-life hospital processes. Developed models for hospital resources and critical care units should be designed within this framework. A specially designed statistical analysis program Apollo (as described above) has been developed to enable the creation of statistically and clinically meaningful patient groups and to obtain information about particular flows over time. This automated rapid classification of patient groups forms a key differentiator between this approach and other attempts to produce practical capacity planning and management tools. Developed simulation models within this framework take individual patients through time as they pass through the chosen healthcare system. These models take uncertainty, variability and complexity into account properly. Figure 10.1 illustrates the inter-linking components of the proposed approach.

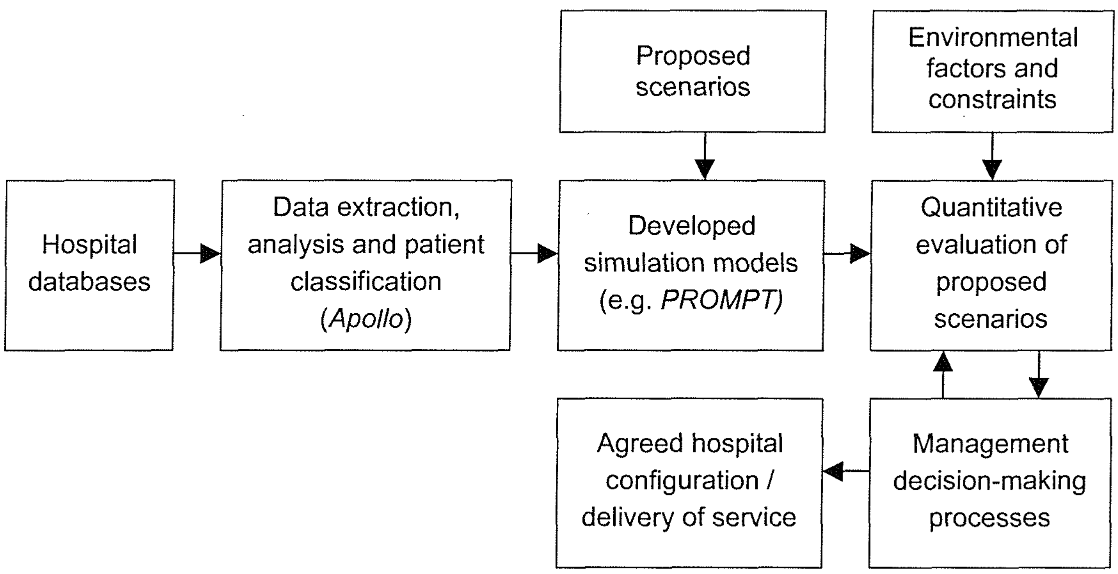


Figure 10.1: Framework components

The framework was critical to the success of developed simulation models. Integrated models for the planning and management of hospital beds, theatres and workforce (PROMPT) and for critical care services have been developed within the evolved framework. Each model is suitably generic to allow for the models to be readily used by other hospitals, helping to promote the benefits of detailed and flexible computer simulation tools.

10.2.4 Towards successful implementation

In many respects the work may be regarded as a feasibility study into the potentials of operational modelling for the planning and management of healthcare resources. During the period of research, working alongside a number of healthcare personnel from the participating hospitals, a number of central themes emerged representing both opportunities and challenges to the operational modeller. It is fair to say that the resolution of a number of issues determines the likely success or failure of healthcare modelling in general. To this end, a framework towards the successful implementation of modelling tools in healthcare is proposed. Each of the issues raised within the framework should be given the due attention they deserve.

The deliberate choice of different simulation tools throughout the research has allowed for a discussion on the advantages and disadvantages of each approach. In essence, the modeller must choose between building the simulation from scratch themselves (a *programming* approach) as opposed to making use of existing simulation tools (*off-the-shelf* approach). In general, careful consideration should be given to selecting the appropriate simulation tool for the project in hand. The choice will likely depend on a number of factors.

10.3 The Direction of Future Research

Working closely with the participating NHS Trusts has revealed a number of opportunities for operational modelling to make further significant impacts on the way in which healthcare is delivered within the NHS. At present the use of operational tools within the profession is limited and successful applications are often not given the wide publicity they deserve. There is considerable scope for further work. Some proposals are described in the subsequent sections of this chapter.

10.3.1 Modelling the wider environment

Critical issues within the hospital setting have highlighted the need to research and model a wider environment incorporating the population structure that the hospital serves and the delicate balance of links to other healthcare institutions within the region. For example, the number of *bed-blockers* (elderly patients who should not be in hospital but in a more suitable care setting), and the great stress that such patients place on bed capacities, points towards the need for an integrated region-wide model examining the flows of patients between hospital, elderly day-centres, nursing homes and home-care facilities. This potentially powerful model could help to structure mid to long-term plans for the provision and distribution of elderly care places within society. With an ageing population, the country can ill-afford to misjudge future elderly needs. The government, with the aid of necessarily detailed, integrated and valid operational models, should begin to plan to avoid placing the healthcare system into further crisis and to avoid potentially placing lives at risk.

The PROMPT model illustrates the benefit of simulation tools for examining the link between hospital beds, theatres and workforce needs. The PROMPT model and similar capacity models do however need to be enhanced to incorporate information on the population that it seeks to serve. The developed models are currently independent of community needs and in practice planning should be a regional issue accounting for the demographic and socio-economic composition of the surrounding population and resulting healthcare needs. For example, current projections on the forecasted number of patients expected to require hospital care are conducted outside of the PROMPT model and then fed in as a user input for scenario analysis. Future work should concentrate on developing a total integrated package linking capacity models to the necessary forecasting tools. This tool would link the demographic structure with other important factors, such as the health of the surrounding population, to predict referral rates and feed them directly into the simulation model. In this way, intervention policies and health improvement programmes within the region may be monitored, their consequences captured in the model and their impact on hospital resources measured.

10.3.2 Hospital data sources and databases

It was necessary to access a number of hospital data sources during the period of research. Perhaps no other system within the NHS displays the levels of diversity as hospital data sources. The problem arises from the huge assortment of data collection and storage methods that are in use at both hospital-wide and departmental levels across the UK. Based on personal experiences of working with various NHS Trusts, current data storage methods may be categorised under the following three headings:

- **Paper-based records** – despite the availability of cheap and reliable computer systems, some hospital departments still only keep paper records of patients. Accessibility and analysis of this data is a painful process.

- **Out-sourced databases** – databases by external software consultancy groups. Built and implemented often at great financial expense to the hospital, most are out of the control of the hospital themselves. So, for example, if an extra field or an extra report is required, the hospital must pay for the consultancy group to carry out the necessary modifications. It is possible that the data legally belongs to the consultancy group who demand extra money to give access to the raw data for analysis. All too often there is mismatch between hospital requirements and the consultancy deliverables.

- **In-house databases** – depends on the skills of the personnel within the hospital IT department. All too often though, due to the failure to attract the right calibre of IT staff (salaries are ridiculously low compared to other IT jobs), in-house databases are limited in number and quality.

Some of the current critical issues surrounding the majority of existing databases in use across UK hospitals appear to include:

- **Lacking of ease of use** – many systems are still in DOS and are complicated to work with and difficult to use.

- **Lacking validation** – most do not include self-validating rules so that the quality is prone to missing values or erroneous entries by the data entry clerk.

- **Out of the hospital control** – healthcare user requirements evolve over time, but without access to the source code, it is not possible for the hospital to re-configure different components of the database without reference to the software developers.

- **Lacking integration** – hospitals may have various data collection sources, both at the hospital-level and for further research needs at a departmental level. Typically however there is no integration between different levels of databases and correspondingly there is often a need to duplicate data entry.

- **Small user-base** – there arose a situation during the research where the only member of staff who was able to use a particular database had left the hospital and consequently no data had been collected until a replacement was found and trained.

In summarising the current situation, there is clearly a great need for improved databases for use within the NHS. Such databases should be in the control of the hospitals themselves, have the ability to add/edit fields as necessary, be self-validating and be able to print a number of standard and ad-hoc reports as required. The combination of databases, with built in statistical and classification tools such as Apollo, and simulation models like PROMPT would be a powerful integrated hospital management tool.

10.3.3 Diary planning tools

In response to the need for improved research linking databases, classification tools and simulation models, a *diary planner* tool for use across a wide range of hospital departments is proposed. Initial discussions with healthcare managers and consultants have been well received with an enthusiastic response. A diary planner could greatly aid the effective monitoring of the provision of care and help improve the day-to-day management of hospital wards whilst still allowing for longer-term planning of hospital capacities by building on the foundations of the developed simulation models.

The proposed model links a patient management system (patient database) with a graphical front-end so that staff can readily see the current status of the ward or hospital. For example, the diary planner could tell nurses and consultants information on which patient is in which bed and how long they may be expected to remain in the bed. This prediction could be based on the results of an integrated classification tool (such as CART or a neural network) that learns and updates itself overtime. Based on medical and socio-economic factors, patient LoS may be obtained from evolved patient groupings and displayed next to the corresponding bed on the front-end display.

By linking the current diary (case-mix on the ward) with information on the likely demand for beds over the forthcoming user-defined *planning horizon* (e.g. the next two weeks), an integrated simulation model could simulate the status of the ward over time and report back forecasted occupancy rates and refusals by shift of each day. A particular benefit of this system would be in the ability to see when in the planning horizon is the best time to schedule elective bookings. For example, the simulation model could report back the probability of there being *n* beds free for a given time block in the future and a management decision regarding patient booking(s) made based on this information. This would help avoid cancelling a large number of elective patients and generally improve the day-to-day efficiency of the ward and hospital. Figure 10.2 shows illustrative screen-shots of how a diary planning tool might look.

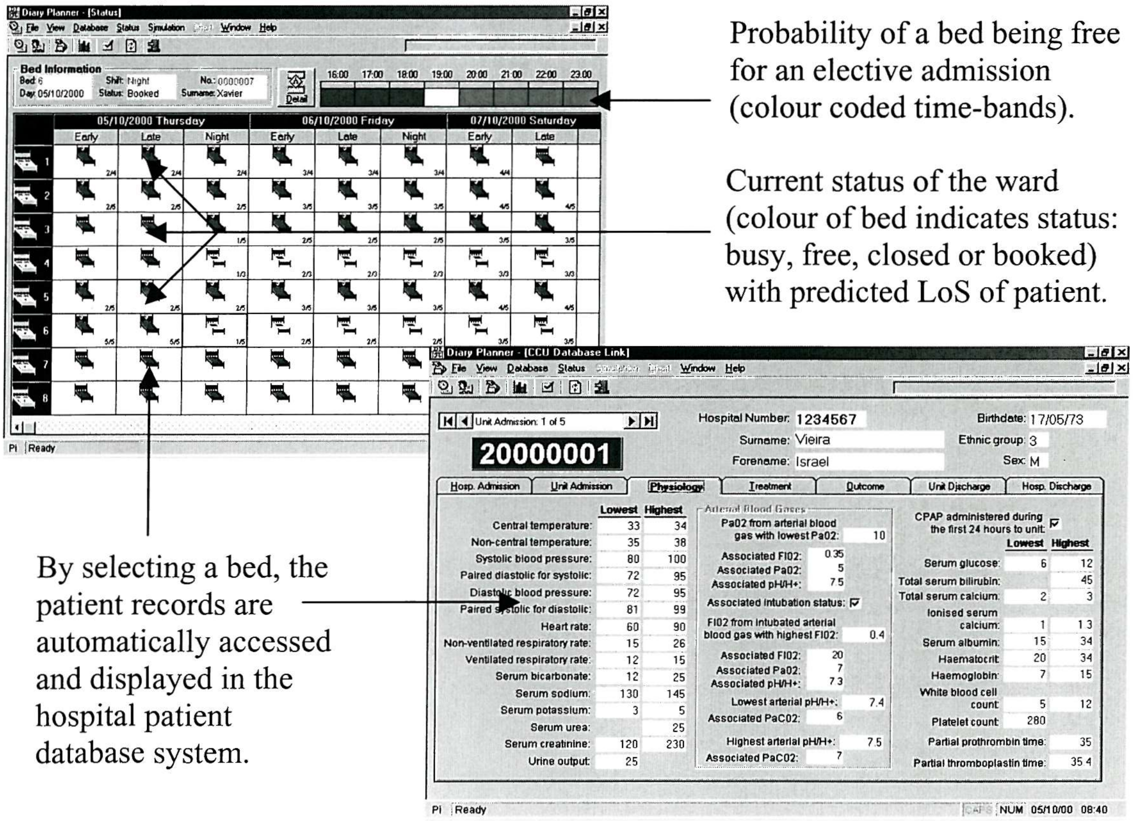


Figure 10.2: Illustrative screen-shots of a proposed diary planner tool

10.4 Chapter Summary

The research work and developed operational tools have delivered a number of benefits to the participating NHS Trusts (see Appendix H). There is a vast range of areas in which simulation could be applied in healthcare and this research has identified a number of opportunities for simulation modelling to make a significant impact on the way in which healthcare is delivered. It is essential that the modeller is aware of the many potential pitfalls and limitations of the healthcare modelling approach. A number of issues defining the primary challenges faced by the modeller in this field have emerged. The success or failure of the modelling approach is dependent on the resolution of these challenges.

Although significant progress has been undoubtedly achieved, it is vital to continually research and develop methodologies for improving the performance and delivery of our healthcare services across the face of the globe. There is a need to promote the evolved framework and developed models, such as those described within this thesis, to a wider healthcare audience. The flexibility of the models should aid this process and help to contribute towards the continued need for improved planning and management of resources. Whilst acknowledging that resources are limited and budgets are tight, the work has shown that there are a number of controllable variables and planning and management methodologies that can help to make life generally more comfortable for hospital staff and to improve the overall quality of patient care.

References

Abernathy, B J *et al.* (1973), "A three-stage manpower planning and scheduling model – a service sector example". *Operations Research.* 21: 693-711.

Abramowitz, M *et al.* (1970), *Handbook of Mathematical Functions.* Dover Publications, New York.

Aharonson-Daniel, L, Paul, R J and Hedley, A J (1996), "Management of queues in out-patient departments: the use of computer simulation". *Journal of Management in Medicine.* 10: 50-58

Anderson, J A (1995), *An Introduction to Neural Networks.* MIT Press, Cambridge, Mass.

Appleby, J (1992), *Financing Health Care in the 1990's.* Open University Press, Buckingham.

Arnold, M (1992), *Outpatient Appointments.* MSc dissertation, University of Southampton.

Ashwood, C and Woodland, M (1990), *SSADM: A Practical Approach.* McGraw Hill, Maidenhead.

Bagust, A, Place, M and Posnett, J W (1999), "Dynamics of bed use in accommodating emergency admissions: stochastic simulation model". *BMJ.* 319: 155-158.

Bailey, N T J (1952), "Study of queues and appointment systems in hospital outpatient departments with special reference to waiting times". *Journal of the Royal Statistical Society.* Series B, 14: 185-198.

Ball, R and Roberts, E (1998), "The relevance of DEA in the public sector". In *Proceedings of the Young Operational Research Conference* (Wadsack, K M editor). Operational Research Society. 15-34.

Baum, E B and Haussler, D (1989), "What size net gives a valid generalisation?". *Neural Computation.* 1: 151-160.

Bell, P C and O'Keefe, R M (1987), "Visual interactive simulation – history, recent developments , and major issues". *Simulation.* 40: 615-624.

Benneyan, J C (1997), "An introduction to using computer simulation in healthcare: patient wait case study". *Journal of the Society for Health Systems.* 5: 1-15.

Benneyan, J C, Horowitz, M I and Terceiro, M B (1994), "Using computer simulation to help reduce patient waiting times". In *Proceedings of the 1994 Annual HIMSS Conference.* 1: 323-342.

Berchtold, G *et al.* (1994), "Simulation modelling as a tool to evaluate alternative configurations of clinical laboratories". *Simulation.* 62: 108-121.

Birchall, J P, Cox, T F and Wong, H (1983), "Computer modelling of ENT outpatients". *Clinical Otolaryngology.* 411-415.

Blatchford, O and Capewell, S (1997), "Emergency medical admissions: taking stock and planning for winter". *BMJ.* 315: 1322-1323.

Blewitt, F *et al.* (1972), "Computer simulation model for multi-specialty ward". *Operations Research Quarterly.* 23: 139-149.

BMA (1998), *Waiting List Prioritisation Scoring Systems.* Health Policy and Economic Research Unit, Discussion Paper 6, December 1998.

Brahami, M and Worthington, D (1991), "Queuing models for out-patient appointment system – a case study". *Journal of the Operational Research Society.* 42: 733-746.

Brailsford, S C *et al.* (1996), "Practical models for the care of HIV and AIDS patients". *International Journal of STD and AIDS.* 7: 91-97.

Breiman, L *et al.* (1984), *Classification and Regression Trees.* Chapman & Hall, London.

Buist, M (1994), "Intensive care resource utilisation". *Anaesthesia and Intensive Care.* 22: 46-60.

Butler, T W, Karwan, K R and Swigart, J R (1992), "Multi-level strategic evaluations of hospital plans and decisions". *Journal of the Operational Research Society*. 43: 665-675.

Callan, R (1999), *The Essence of Neural Networks*. Prentice-Hall, London.

Capewell, S (1996), "The continuing rise in emergency admissions". *BMJ*. 312: 991-992.

Carter, M *et al.* (1992), "Simulation, scheduling and operating rooms". In *Proceedings of the 1992 Simulation in Health Care and Social Services Conference*. 28-30.

Ceric, V (1990), "A simulation study of an automated guided vehicle system in a Yugoslav hospital". *Journal of the Operational Research Society*. 41: 299-310.

Charnes, A *et al.* (1995), *Data Envelopment Analysis: theory, methodology and applications*. Kluwer, New York.

Charnes, A, Cooper, V W and Rhodes, E (1978), "Measuring the efficiency of decision making units". *European Journal of Operational Research*. 2: 429-444.

Chau, P Y K and Bell, P C (1995), "Designing effective simulation-based decision support systems: An empirical assessment of three types of decision support systems". *Journal of the Operational Research Society*. 46: 315-331.

Checkland, P B (1981), *Systems Thinking, Systems Practice*. Wiley, Chichester.

Checkland, P B (1984), "Systems thinking in management: the development of soft systems methodology and its application for social science". In *Self-organisation and the Management of Social Systems*. (Ulrich, H and Probst, G J B editors). Springer-Verlag, Berlin.

Checkland, P B and Scholes, J (1990), *Soft Systems Methodology in Action*. Wiley, Chichester.

Cheng, R C H and Holland, W (1997), "Sensitivity of computer simulation experiments to errors in input data". *Journal of Statistical Computation and Simulation*. 57: 219-241.

Cheng, R C H and Kleijnen, J P C (1999), "Improved design of queueing simulation experiments with highly heteroscedastic responses". *Operations Research.* 47: 762-777.

Cheng, R C H and Lamb, J D (2000), "Making efficient simulation experiments interactively with a desktop simulation package". *Journal of the Operational Society.* 51: 501-507.

Clark, D E *et al.* (1994), "Optimal location for a helicopter in a rural trauma system: prediction using discrete-event computer simulation". In *Proceedings of the 18th Annual Symposium on Computer Applications in Medical Care.* (Ozbolt, J G editor). McGraw Hill, New York. 878-883.

Clark, P and Boswell, R (1991), "Rule induction with CN2: Some recent improvements". In *Proceedings of ESWL '91* (Porto, Portugal) 151-163.

Cohen, J W (1956), "Certain delay problems for full availability trunk groups loaded by two traffic sources". *Communication News.* 16: 105-113.

Cohen, M A, Hershey, J C and Weiss, E N (1980), "Analysis of capacity decisions for progressive patient care hospital facilities". *Health Services Research.* 15: 145-160.

Croft, P, Barlow, J and Edwards R (1997), "A CliniQueue solution – managing outpatient referral lists with a generic computer simulation model". *OR Insight.* 10: 22-27.

Dankbar, G, Shellum, J and Bennet, K E (1992), "The use of simulation to evaluate automated equipment for a clinical processing laboratory". In *Proceedings of the 1992 Winter Simulation Conference.* (Swain, J J editor). 1065-1070.

Davies, R (1985), "An Assessment of models of a health system", *Journal of the Operational Society.* 38: 679-687.

Davies, R, Johnson, D and Farrow, S (1975), "Planning patient care with a Markov model". *Operational Research Quarterly.* 26: 599-607.

Davies, R and O'Keefe, R (1989), *Simulation Modelling with Pascal.* Prentice Hall, London.

Day, V (2000), *Comprehensive Critical Care: A review of adult critical care services.* Department of Health, London.

Dennett, E R *et al.* (1998), "Generic surgical priority criteria scoring systems". *New Zealand Medical Journal.* 111: 163-166.

Dennett, E R *et al.* (1998), "Priority access criteria for elective cholecystectomy: a comparison of three scoring methods". *New Zealand Medical Journal.* 111: 231-233.

Department of Health (1990), *The Patient's Charter.* HSMO, London.

Department of Health (1991), *The Health of a Nation.* HSMO, London.

Department of Health (1995), *A Code of Openness in the NHS.* HSMO, London.

Department of Health (1996), *Jurisdiction of the Health Service Commission.* HSMO, London.

Department of Health (1997), *The New NHS Executive Summary.* HSMO, London.

Department of Health (1999), *General and Personal Medical Service Statistics.* HSMO, London.

Department of Health (2000), *Comprehensive Critical Care: A Review of Adult Critical Care Services.* HSMO, London.

Department of Health (2000), *Creating a 21st Century NHS.* HSMO, London.

Department of Health (2000), *Hospital Episode Statistics, England 1998-1999.* HSMO, London.

Dickson, G W, Desanctis, G and McBridge, D J (1986), "Understanding the effectiveness of computer graphics for decision support: a cumulative experimental approach". *Communications of the ACM.* 29: 40-47.

Dowsland, K A (1998), "Nurse scheduling with tabu search and strategic oscillation". *European Journal of Operational Research.* 106: 393-407.

Dowsland, K A and Thompson, J M (2000), "Solving a nurse scheduling problem with knapsacks, networks and tabu search". *Journal of the Operational Research Society*. 51: 825-833.

Draper, N R and Smith, H (1966), *Applied Regression Analysis*. Wiley, New York.

Dubey, S F (1967), "Some percentile estimators for Weibull parameters". *Technometrics*. 9: 119-129.

Dumas, M B (1984), "Simulation modelling for hospital bed planning". *Simulation*. 43: 69-78.

Dumas, M B (1985), "Hospital bed utilization: An implemented simulation approach to adjusting and maintaining appropriate levels". *Health Services Research*. 20: 42-61.

Dundas, A and Meechand, D (1986), "Better use of surgical beds". *Health Service Journal*. 96: 4999, 665.

Eliasz, A W (1992), *Theory and Practice of Coloured Petri Nets using Design/CPN*. Micromatch Ltd., England.

Eliasz, A W, Manning, B and Chattaway, C (1993), "The use of CASE tools in static and dynamic modelling of healthcare delivery processes with special reference to community healthcare nursing". In *The Eleventh Congress for European Medical Informatics*. (Reichert *et al.* editors). 732-734. Freund Publishing, London.

Erlang A K (1917), "Solution of some problems in the theory of probabilities of significance in automatic telephone exchanges". *Elektrotkeknikeren*. 13.

Estes, C L and Swan, J H (1993), *The Long Term Care Crisis*. Sage Publications, London.

Fahlman, S (1999), "Distributed connectionist systems for AI: Prospects and problems". In *Concepts and characteristics of knowledge-based systems*, proceedings of IFIP 10/WG 10.1 workshop , Mount Fuji, Japan. 22-35.

Farrow, S C, Fisher, D J H and Johnson, D B (1971), "Statistical approach to planning an integrated haemodialysis/transplantation programme". *British Medical Journal*. 2: 671-676.

Fetter, R B and Thompson, J D (1966), "Patient's waiting time and doctor's isle time in the outpatient setting". *Health Service Research.* 1: 66-90.

Forrester, J W (1961), *Industrial Dynamics.* John Wiley, New York.

Forrester, J W (1968), *Principles of Systems.* Wright-Allen, Cambridge, Mass.

Freeman, P (1992), *Designing Appointment Systems for ENT Outpatient Clinics.* MSc dissertation, University of Southampton.

Fries, B E (1976), "Bibliography of Operations Research in health-Care systems". *Operations Research.* 24: 801-814.

Fries, B E (1979), "Bibliography of Operations Research in health-Care systems". *Operations Research.* 27: 408-419.

Gamlin, H (1997), *Planning and Management of ENT Outpatient Clinics.* MSc dissertation, University of Southampton.

Garbe, C *et al.* (1995), "Primary Cutaneous Melanoma – Identification of prognostic groups and estimation of individual prognosis for 5,093 patients". *Cancer.* 75: 2484-2491.

Gerchak, Y, Gupta, D and Henig, M (1996), "Reservation planning for elective surgery under uncertain demand for emergency surgery". *Management Science.* 42: 321-334.

Gibson (1998), "A model way to plan for surgical needs". *Health Service Journal.* 98: 5085, 6.

Goldman, J, Knappenberger, H A and Eller, J C (1968), "Evaluating bed allocation policy with computer simulation". *Health Service Research.* 3: 119-129.

Gordon, T *et al.* (1998), "Surgical unit time utilisation review: resource utilisation and management implications". *Journal of Medical Systems.* 12: 169-179.

Gove D, Hewett D and Shahani, A K (1995), "Towards a model for hospital case-load decision support". *IMA Journal of Mathematics Applied in Medicine and Biology.* 12: 329-338.

Graeber, S (1995), "Object-orientated modelling of hospital information systems". In *MEDINFO '95 Proceedings*. (Greenes, R A *et al.* editors). IMIA, Amsterdam.

Gudex, C *et al.* (1990), "Prioritising waiting lists". *Health Trends*. 22:3.

Hackney, B A, Casey, K L and Narasismhan, S L (1984), "Maximising resources: efficient scheduling of the OR". *AORN Journal*. 39: 1174-1180.

Hadorn, D C and Holmes, A C (1997), "The New Zealand priority criteria project". *BMJ*. 314: 135-138.

Ham, C (1992), Editors Introduction in *Planned Markets and Public Competition: Strategic Reform in Northern European Health Systems*. (Saltman, R D and Von Otter, C editors). Open University Press, Buckingham.

Hamilton, B A and Breslawski, S (1994), "Operating room scheduling: factors to consider". *AORN Journal*. 59: 665-680.

Hannan, E (1984), "A Markov sensitivity model for examining the impact of cost allocations in hospitals". *Journal of the Operational Research Society*. 35: 117-129.

Hanson, K H (1982), Computer assisted operating room scheduling". *Journal of Medical Systems*. 6: 311-314.

Harrison, S (1991), "Working of the markets, purchaser/provider separation in English health care". *International Journal of Health Services*. 12: 625-635.

Hawkins, J D *et al.* (1992), *TOCHSIM manual*. Southampton Centre for Operational Research, Faculty of Mathematical Studies, University of Southampton.

Haykin, S (1999), *Neural Networks: A Comprehensive Foundation*. Prentice-Hall, London.

Herbert, S I and Willis, A W (1992), "What is general medical practice?". In *MEDINFO '92*. (Lun *et al.* editors). Elsevier, Amsterdam. 1172.

Hirata, C M and Paul, R J (1996), "Object-Oriented programming architecture for simulation modelling". *International Journal in Computer Simulation*. 6: 269-287.

Hollingsworth, B and Parkin, D (1995), "The efficiency of Scottish acute hospitals; an application of data envelopment analysis". *IMA Journal of Mathematics Applied in Medicine and Biology.* 12: 161-173.

Hornberger, J C, Habraken, H and Bloch, D A (1995), "Minimum data needed on patient preferences for acute, efficient medical decision-making". *Medical Care.* 33: 297-310.

Howard, R A (1971), *Dynamic Probabilistic Systems: Markov Models.* Wiley, New York.

Howard, R A (1971), *Dynamic Probabilistic Systems: Semi-Markov and Decision Processes.* Wiley, New York.

Huang, X M, Taylor, C and Stewart, I (1995), "Computer simulation of an A&E department". *OR Insight.* 8: 18-23.

Huebner, D and Miller, L (1996), "Business process re-engineering of an outpatient clinic using simulation". In *Proceedings of the 1996 Annual HIMSS Conference.* 87-100.

Hurrion, R D (1978), "An investigation of visual interactive simulation methods using the job-shop scheduling problem". *Journal of the Operational Research Society.* 29: 1085-1093.

Intensive Care Society (1997), *Standards for Intensive Care units.* ICS, London.

Jenkins, A M (1985), "Prototyping: a methodology for the design and development of application systems". *Spectrum Magazine.* Society of Information Management, April Edition.

Jones, S K (1997), *Mathematical Modelling for Early Detection and Treatment of Cancer.* PhD Thesis, University of Southampton.

Kalton, A G *et al.* (1997), "Using simulation to improve the operational efficiency of a multi-disciplinary clinic". *Journal of the Society for Health Systems.* 5: 43-62.

Kao, E P C and Tung, G G (1981), "Aggregate nursing requirement planning in a public health care delivery system". *Socio-Economic Planning Science.* 15: 119-127.

Kao, E P C and Tung, G G (1981), "Bed allocation in a public health care delivery service". *Management Science.* 27: 507-520.

Kay, J W and Titterington, D M (1999), *Statistics and Neural Networks: Advances at the Interface.* Oxford University Press, Oxford.

Kay, S (1993), "Object-orientated analysis of patient registration". In *Conference Proceedings – Current Perspectives in Healthcare Computing.* (Richards, B editor). BJHC Books, Weybridge.

Keller, T F and Laughman, D J (1973), "An application of queuing theory to a congestion problem in an outpatient clinic". *Decision Sciences.* 4: 379-394.

Kilmer, R A, Smith, A E and Shuman, L J (1997), "An emergency department simulation and a neural network metamodel". *Journal of the Society for Health Systems.* 5: 63-79.

King, R D, Feng, C and Sutherland, A (1995), "Statlog: Comparison of classification algorithms on large real-world problems". *Applied Artificial Intelligence.* 9: 289-333.

Kirkland, A *et al.* (1995), "Simulating an emergency department "Is as much fun as…"". In *Proceedings of the 1995 Winter Simulation Conference.* (Alexopoulos, C *et al.* editors). 1039-1042.

Klassen, K J and Rohleder, T R (1996), "Scheduling outpatient appointments in a dynamic environment". *Journal of Operations Management.* 14: 83-101.

Kleinsorge, I K and Harney, D (1992), " Management of nursing homes using Data Envelopment Analysis". *Socio-Economic Planning Service.* 26: 57-71.

Knighton, J, Will, S J and Brampton, W J (1994), "A statistical model to predict intensive care unit bed availability". *Clinical Intensive Care.* 5:146.

Kwak, N, Kuzdrall, P and Owen, V T (1975), "Simulating the use of space in a hospital surgical suite". *Simulation.* 25: 147-151.

Lane, D C (1992), "Modelling as learning: a consultancy methodology for enhancing learning in management teams". *European Journal of the Operational Research Society.* 59: 64-84.

Lane, D C (1999), *Systems Dynamics Modelling of Patient Flows Through Acute Hospitals.* Report for the NHS Executive, London.

Lane, D C, Monefeldt, C and Rosenhead, J V (2000), "Looking in the wrong place for healthcare improvements: A system dynamics study of an accident and emergency department". *Journal of the Operational Research Society.* 51: 518-531.

Law, A M and Kelton, W D (1991), *Simulation Modeling and Analysis.* McGraw Hill, New York.

Lehaney, B and Paul, R J (1994), "Using soft systems methodology to develop a simulation of out-patient services". *Journal of the Royal Society for Health.* 114: 248-251.

Lehaney, B and Paul, R J (1996), "The use of soft systems methodology in the development of a simulation of out-patient services at Watford General Hospital". *Journal of the Operational Research Society.* 47: 864-870.

Lehaney, B, Clarke, S A and Paul, R J (1999), "A case of an intervention in an outpatients department". *Journal of the Operational Research Society.* 50: 877-891.

Levy, J L, Watford, B A and Owen, V T (1989), "Simulation analysis of an outpatient services facility". *Journal of the Society for Health Systems.* 1: 35-49.

Liebman, S J, Young, J P and Bellmore, M (1972), "Allocation of nursing personnel in an extended care facility". *Health Services Research.* 7: 209-220.

Liu, C-M, Wang, K-M and Guh, Y Y (1991), "A Markov chain model for medical record analysis". *Journal of the Operational Research Society.* 42: 357-364

Lowery, J (1992), "Simulation of a hospital's surgical suite and critical care area". In *Proceedings of the 1996 Winter Simulation Conference.* (Swain, J J *et al.* editors). 1071-1078.

Luck, G M *et al.* (1971), *Patients, Hospitals, and Operational Research.* Tavistock Publications, London.

Luckman, J (1971), "Planning surgery: a computer simulation". *Health Report No. 2.* Institute of Operational Research, London.

Magerlein, J M and Martin, J B (1978), "Surgical demand scheduling: a review". *Health Services Journal.* 13: 418-433.

Maier-Rothe, C and Wolfe, H B (1973), "Cyclical scheduling and allocation of nursing staff". *Socio-Economic Planning Science.* 7: 471-487.

Manly, B F J (1998), *Multivariate Statistical Methods.* Chapman & Hall, London.

Marca, D A and McGowen, C (1988), *SADT: Structured Analysis and Design Technique.* McGraw Hill, New York.

McGuire, F (1997), "Using simulation to reduce length of stay in emergency departments". *Journal of the Society for Health Systems.* 5: 81-90.

McQuarrie, D G (1981), "Limits to efficient operating room scheduling. Lessons from computer use models". *Archives of Surgery.* 116: 1065-1071.

Metcalfe, A and McPherson, K (1994), *Study of Provision of Intensive Care in England, 1993.* Revised report for Department of Health.

Mehrotha, K *et al.* (1997), *Element of Artificial Neural Networks.* MIT Press, Cambridge, Mass.

Meyer, D A (1993), "Use of object-orientated approach for the management of waiting lists in ophthalmology". In *Conference Proceedings of The Eleventh International Conference for European Medical Informatics.* (Reichert *et al.* editors). 735-738. Freund Publishing, London.

Miller, H E, Pierskala, W P and Rath, G R (1976), "Nurse scheduling using mathematical programming". *Operations Research.* 24: 857-870.

Moores, B (1987), "Parking problem? No problem". *Health Service Journal.* 97: 5054, 672.

Mukherjee, A (1991), "A simulation model for management of operations in the pharmacy of a hospital". *Simulation.* 2: 91-103.

Mullen, P M (1998), "Priority setting and waiting lists: rationing or demand management". Paper presented at OR40, *Conference of the Operational Research Society.* Lancaster, September 1998.

Muller, N and Muller, H (1998), "Decision support system for on site medical regulation during mass casualty disasters". In *Simulation: Past, Present and Future – Proceedings of the 12th European Simulation Multiconference.* (Zobel, R and Moeller editors). 366-369.

Murphy, D, Ducket, S and Sigal, E (1985), "Evaluating surgical block schedules using computer simulation". In *Proceedings of the 1985 Winter Simulation Conference.*

Nelder, J A and Mead, R (1965), "A simplex method for function minimisation". *Computer Journal.* 7: 308-313.

O'Donnell, D J (1976), "Theatre utilisation analysis". *Medical Journal of Australia.* 2: 650-651.

O'Neil, R (1971), "Algorithm AS47: Function minimisation using a simplex procedure". *Applied Statistics.* 20: 338-345.

Oren, T I (1974), "A bibliography of bibliographies on modelling, simulation and gaming". *Simulation.* 23: 90-95 and 23: 115-116.

Oren, T I (1976), "Annotated bibliography of simulation". *Simulation Council.* Inc, La Jolla, CA.

Orman, A J (1995), *Models for Scheduling a Multifunctional Phased Radar System.* Ph.D. Thesis, University of Southampton.

Ozcan, Y A and McCue, M J (1996), "Development of a finicial performance index for hospitals: DEA approach". *Journal of the Operational Research Society.* 47: 573-582.

Ozcan, Y A *et al.* (1998), "Provider experience and technical efficiency in the treatment of stroke patients: DEA approach". *Journal of the Operational Research Society. Journal of the Operational Research Society.* 49: 573-582.

Paul, R J (1995), "Outpatient clinics – The CLINSIM simulation package". *OR Insight.* 8: 24-27.

Pidd, M (1992), *Computer-Based Simulation for Management Science.* Wiley, Chichester.

Pidd, M (1998), *Computer Simulation in Management Science.* Wiley, Chichester.

Press, W H *et al.* (1992), *Numerical Recipes in C: The Art of Scientific Computing.* Cambridge University Press, Second Edition.

Remme, J, Habbema, J D F and Hermans, J (1980), "A simulative comparison of linear, quadratic and kernel discrimination". *Journal of Statistics and Computer Simulation.* 11: 87-106.

Richamond, B M , Vescuso, P and Peterson, S (1990), *iThink Software Manuals.* High Performance Systems, Hanover, NH.

Ridge, J C *et al.* (1998), "Capacity planning for intensive care units". *European Journal of the Operational Research Society.* 105: 346-355.

Ridley, S *et al.* (1998), "Classification trees: a possible method for iso-resource grouping in intensive care". *Anaesthesia.* 53: 833-840.

Rivett, G (1998), *Cradle to Grave.* King's Fund, London.

Roberts, N *et al.* (1994), *Introduction to Computer Simulation – A Systems Dynamics Modelling Approach.* Productivity Press, Portland, Oregon.

Roginski, C (1995), "Applying SODA in the NHS". *OR Insight.* 8: 28-32.

Rojas, F and Martinez, A I (1998), "From process models to enactment and simulation". In *Simulation: Past, Present and Future – Proceedings of the 12th European Simulation Multiconference.* (Zobel, R and Moeller editors). 844-848.

Rose, M B and Davies, D C (1984), "Scheduling in the operating theatre". *Annals of the Royal College of Surgeons of England.* 66: 372-374.

Rosenbloom, E S and Goertzen, N F (1987), "Cyclic nurse scheduling". *European Journal of Operational Research.* 31: 19-23.

Royston, G (2000), *Modernising Government with Operational Research.* Overview Magazine January 2000 (Cunliffe, P editor). Public Sector Information Group, Stockport.

Ruth, R J (1981), "A mixed integer programming model for regional planning of a hospital inpatient service". *Management Science*. 27: 521-533.

Saaty, T L (1961), *Elements of Queuing Theory*. McGraw Hill, New York.

Sachdeva, R C and Guntupalli, K K (1999), "International comparisons of outcomes in intensive care units". *Critical Care Medicine*. 27: 2032-2033.

Saltman, R D and Von Otter, C (1992), *Planned Markets and Public Competition: Strategic Reform in Northern European Health Systems*. Open University Press, Buckingham.

Sargent, R G (1991), "Simulation modelling verification and validation". In *Proceedings of the 1991 Winter Simulation Conference*. (Nelson, B L *et al.* editors). 37-47.

Schroyer, D (1997), "Simulation supports ambulatory surgery facility design decisions". In *Proceedings of the 1997 Annual HISMSS Conference*. 4: 95-108.

Shahani, A K (1981), "Reasonable averages that give wrong answers". *Teaching Statistics*. 3: 50-54.

Shahani, A K *et al.* (1994), " Towards an operational model for the prevention and treatment of asthma attacks". *Journal of the Operational Research Society*. 45: 916-926.

Shonick, W and Jackson, J R (1973), "An improved stochastic model for occupancy-related random variables in general-acute hospitals". *Operations Research*. 21: 952-965.

Shuman, L J, Dixon-Spears Jr., J and Young, J P (1975), *Operations Research in Health Care: A Critical Analysis*. John Hopkins University Press.

Sier D, Tobin P and McGurk C (1997), "Scheduling surgical procedures". *Journal of the Operational Research Society*. 48: 884-891.

Silfred, S P and Benton, W C (1992), "Workforce staffing and scheduling: hospital nursing specific models". *European Journal of Operational Research*. 60: 223-246.

Silfred, S P and Benton, W C (1994), "A decision model for shift scheduling of nurses". *European Journal of Operational Research.* 74: 519-527.

Smith, M E *et al.* (1992), "Case-mix groups for hospital-based home care". *Medical Care.* 30: 1-16.

Smith-Daniels, V L (1989), "Capacity management in health care services: review and future research directions". *Decision Sciences.* 19: 889-919.

Stout, W A (1995), "Modelling a hospital main cafeteria". In *Proceedings of the 1995 Winter Simulation Conference.* (Alexopoulous, C *et al.* editors). IEEE Piscataway, New Jersey. 1060-1065.

Taylor, S and Kuljis, J (1998), "Simulation in health care management: modelling an outpatient clinic". *OR Insight.* 11: 7-11.

Tocher, K D (1963), *The Art of Simulation.* English Universities Press, London.

Vassilacopoulos, G (1985), "A simulation model for bed allocation to hospital inpatient departments". *Simulation.* 45: 233-241.

Vetter, N (1995), *The Hospital: From Centre of Excellence to Community Support.* Chapman and Hall, London.

Vissers, J (1979), "Selecting a suitable appointment system in an outpatient setting". *Medical Care.* 12: 1207-1220.

Weiss, E and McClain, J (1987), Administrative day in acute care facilities: a queuing analytic approach". *Journal of the Operational Research Society.* 35: 35-44.

Williams, W, Covert, R and Steele, J (1967), "Simulation modelling of a teaching hospital outpatient clinic". *Journal of the American Hospital Association.* 41: 71-75.

Wolfe, H and Young, J P (1963), "Staffing the nursing unit, part I: controlled variable staffing". *Nursing Research.* 14: 236-243.

Wolfe, H and Young, J P (1963), "Staffing the nursing unit, part II: the multiple assignment technique". *Nursing Research.* 14: 299-303.

Wolstenholme, E A (1993), "A case study in community care using systems thinking". *Journal of the Operational Research Society.* 44: 925-934.

World Bank (1993), *World Development Report 1993. Investing in Health.* Oxford University Press, Oxford.

World Health Organisation (2000), *The World Health Report 2000. Health Systems: Improving Performance.* WHO, Geneva, Switzerland.

Worthington, D (1987), "Queuing models for hospital waiting lists". *Journal of the Operational Research Society.* 38: 413-422.

Worthington, D (1991), "Hospital waiting list management models". *Journal of the Operational Research Society.* 42: 833-843.

Worthington, D (1995), "Use of 'what if?' models to aid strategic management of hospital waiting lists". Presented at ORS One-Day Conference – *Strategic Simulation for Purchasing and Providing.* 18th May 1995.

Wright, D (1987), "The application of a surgical bed simulation model". *European Journal of Operational Research.* 29: 56-60.

Wright, I H *et al.* (1996), "Statistical modelling to predict elective surgery time: comparison with a computer scheduling system and surgeon provided estimates". *Anaesthesiology.* 85: 1235-1245.

Xu, L, Krzyzak, A and Oja, E (1991), "Neural nets for dual subspace pattern recognition method". *International Journal of Neural Systems.* 2: 169-184.

Yates, J (1982), *Hospital Beds: A Problem for Diagnosis and Management?* William Heinemann, London.

Young, J (1962), *A Queuing Theory Approach to the Control of Hospital Inpatient Census.* Operations Research Division, John Hopkins University.

Zeigler, B P (1979), "Modelling and simulation methodology: state of the art and promising directions". In *Simulation of Systems.* (Dekker, L editor). North-Holland, Amsterdam.

Publications and Presentations

The combination of academic and hospital supervisors has enabled this research to be of practical use and benefit to the healthcare profession (Appendix H). As a consequence of the generic framework and flexibility of the developed operational models as discussed in this thesis, they have since been successfully used by a variety of Trusts across the UK and have been presented at a number of domestic and international conferences. The thesis has resulted in a number of publications, which are listed below, together with details of conferences at which the author presented the research.

2002 Harper, P R & Shahani, A K (2002), "Modelling for the Planning and Management of Bed Capacities in Hospitals". *Journal of the Operational Research Society*. 53: 11-18.

Harper, P R (2002), "A Framework for Operational Modelling of Hospital Resources". To appear in *Health Care Management Science*.

2001 Shahani, A K, Harper P R, Vieira, I T and Costa, A X, "Planning for Critical Care Capacities". IMH pre-print, Faculty of Mathematical Studies, University of Southampton.

"Planning and Management of Hospital Capacities: An Integrated Computer Model for Beds, Theatres and Workforce". Presented at *A Celebration of Innovation in Health Care Conference* (University of Reading).

2000 Harper, P R, Dale, J, Foden, D, de Senna, V and Shahani, A K (2000), "Operational Modelling for the Planning and Management of Capacities in Hospitals". In *Proceedings of The 7th International Conference on System Science in Health Care* (Budapest) 113-117.

Harper, P R (2000), "A Decision Support Hospital Simulation Model". In *Proceedings of The International Conference on Health Sciences Simulation* (San Diego) 27-31

1999 "Modelling for the Planning and Management of Hospital Beds". Presented at IFORS (Beijing, August 1999).

Workshop Leader at the Health Care Development Group Annual Conference (Southampton, March 1999). Workshop on Stochastic Modelling and Statistical Analysis.

1998 Healthcare Stream Organiser at the Young OR Conference (Guildford, April 1998). Presented "Towards an Integrated Hospital Capacity Model: Beds, Theatres and Workforce" and at the annual conference of the OR Society (Lancaster, September 1998).

Appendix A – Steering Committees

In was important to delineate a coherent research and development approach consistent with the stated objectives of this thesis and the needs of the participating NHS Trusts. To meet this need, an evolutionary development methodology was adopted (Chapter 4). This involved constant dialogue with end-users and its success was largely dependent on relevant hospital personnel contributing in many ways to the development and structure of the prototype models.

This was achieved through the creation of steering committees to guide the work at both Reading and Portsmouth. Members were made up from relevant areas of the hospital and acted as key contact points for information and provided access to other staff. One member of each steering group was designated as the Chairperson. These groups met at agreed regular intervals and discussed project management issues.

A.1 Royal Berkshire and Battle Hospital NHS Trust

Heather Bunce (Clinical Service Unit Manager – Adult Medicine)
Jana Dale (Hospital Process Redesign)
David Foden (Director of Change Management & Human Resources)
Paul Harper (Institute of Modelling for Healthcare)
Rodney Jones (Senior Analyst, Information Management)
Sharon Kearns (Chairperson; Director of Operations)
Eva Morgan (Hospital Process Redesign)
Chris Newman (Director of Womens Services & Child Medicine)
Andrew Pengelly (Medical Executive Director)
Arjan Shahani (Director, Institute of Modelling for Healthcare)

A.2 Portsmouth Hospitals NHS Trust

Ginette Alexander (Corporate Accountant)

Bill Broadribb (Executive Director of Finance)

Andy Burrows (Deputy Director of Finance)

Ann Carter (Senior Project Manager)

Brian Goodridge (Head of Corporate Information & Applications)

Paul Harper (Institute of Modelling for Healthcare)

Nicola Hartley (Director of Change Program)

Peter Howlett (Chairperson; Executive Director of Development)

Sue Millard (Head of PFI and Strategic Projects Accountant)

Arjan Shahani (Director, Institute of Modelling for Healthcare)

Liz Steel (Senior Project Manager)

Graham Taylor (Information Manager)

Michelle Wheeler (Capital Projects Manager)

Appendix B – Stochastic Modelling

B.1 Deterministic and Stochastic Models

A common current practice is to plan and manage hospital capacities through a simple deterministic spreadsheet approach using average patient-flows, average needs, average length-of-stay and average duration of surgical operations. Patient-flows, patient needs and utilisation of hospital capacities involve complexity, uncertainty, variability, constraints and scarce resources, as discussed in the previous section. Mathematically speaking, a hospital corresponds to a complex stochastic system so that the common deterministic approach for planning and managing the system can be expected to be inadequate.

Overcoming variability by using average values, which although intuitively attractive when considering the large numbers of patients in the system, has the effect of significantly biasing results in an environment of non-linearity and variability. Consider a system in which the output, Y, depends on the input X, where Y = f(X) and X is a random variable as illustrated in Figure B.1.

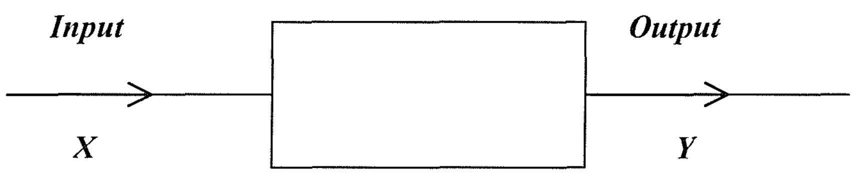


Figure B.1: A simple system

Let $E(X) = \mu$. A deterministic model that attempts to deal with variability using average values, so that $E(Y) \approx f(\mu)$, may seem intuitively simple for large systems. However in many practical cases this approach is likely to be inappropriate.

If the Taylor series expansion of $f(X)$ is examined we have:

$$f(X) = f(\mu) + (X-\mu)f'(\mu) + \frac{(X-\mu)^2}{2!}f''(\mu) + \ldots$$

which leads to

$$E(Y) = E(f(X)) = f(\mu) + \frac{Var(X)}{2!}f''(\mu) + \ldots$$

Hence, if f is non-linear and the variability in X is large, the approximation $E(Y) = f(\mu)$ would be a very poor one. The deterministic approach of using average values leads to significant bias and inaccurate results, typically under-estimating the true resource needs.

Unlike deterministic models, stochastic models provide a more accurate and realistic model by incorporating uncertainty and variability through the use of probabilities and random variables. A stochastic process is a sequence of random variables $\mathbf{X} = \{X_t\}$ where the parameter t usually denotes the time domain. Both t and X_t may be continuous or discrete. The values taken by the random variables X_t are called *states* and the domain of **X** is called the *state space*. A *realisation* (or simple path) of the stochastic process **X** is a sequence of states $\{x_0, x_1, x_2, \ldots\}$, where x_t is the value taken by the random variable X_t. For the modelling work of hospital systems, the necessary stochastic descriptions involve discrete states and continuous time. Two stochastic processes have to be considered, namely *Markov* and *semi-Markov* processes.

B.2 Markov and Semi-Markov Processes

Markov processes are attractive because of analytical ease. They are probabilistic models, which form a special category of stochastic processes. The fundamental property of the Markov process is its *memoryless property* i.e. the probability that the random variable X_t takes a particular value x_t depends only on x_{t-1} and not on the previous values $x_{t-2}, x_{t-3}, \ldots, x_0$. Notationally, this may be written:

$$\mathrm{P}\{X_t = x_t | X_{t-1} = x_{t-1}, X_{t-2} = x_{t-2}, \ldots, X_0 = x_0\} = \mathrm{P}\{X_t = x_t | X_{t-1} = x_{t-1}\}$$

Under this Markov hypothesis, the main limitation is that the transitional probability of a patient changing status is taken to be independent of previous events, thus the likelihood of a patient remaining in a specific state remains the same from one time unit to the next.

A semi-Markov process is a more general class of processes where the time between transitions of states can occur after any positive time spent in a state, and where this transition time can depend on the transition that is made. Each successive state occupancies are governed by the transition probabilities of a Markov process and the stay in any state is described by a random variable that can take on any positive, and not necessarily, an integer value. As an example, patients may pass through various states in a critical care unit (holding trolley to ICU bed to HDU bed to Ward bed) and the time spent in each state described by a random variable.

Let p_{ij} be the probability that a semi-Markov process that entered state i on its last transition will enter state j on its next transition. The transition probabilities p_{ij} must satisfy the same equations as the transition probabilities for a Markov process.

$$p_{ij} \geq 0 \qquad i = 1, 2, \ldots, N;\ j = 1, 2, \ldots, N$$

and

$$\sum_{j=1}^{N} p_{ij} = 1 \qquad i = 1, 2, \ldots, N$$

where N is the total number of states in the system.

Howard (1971) describes the process as follows: Whenever a process enters a state i, we can imagine that it determines the next state j to which it will move according to state i's transition probabilities $p_{i1}, p_{i2}, ..., p_{iN}$. However, after j has been selected, but before making this transition from state i to state j, the process "holds" for a time t_{ij} in state i. The holding times are positive, integer-valued random variables each governed by a probability density function $g_{ij}(t)$ called the holding time density function for a transition from state i to state j. The cumulative function $G_{ij}(t)$ is the probability that the transition from state i to state j will fall between 0 and t.

Returning to the previous critical care unit example, each patient will now have different probabilities of passing from one state to the next (for example, going from ICU bed to HDU bed to Ward bed, or ICU bed straight to Ward bed), and the times spent in each state (length of stay) are sampled from appropriate distributions. If lengths of stay are negative exponential variates, we obtain the Markov process. In general though, length of stay will follow other statistical distributions, such as Lognormal or Weibull.

As an illustrative example of the semi-Markov process, consider a simplified model with state space $S = \{0, 1, 2, 3\}$ and probabilities with the following probability matrix:

$$P = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

The model is given in Figure B.2.

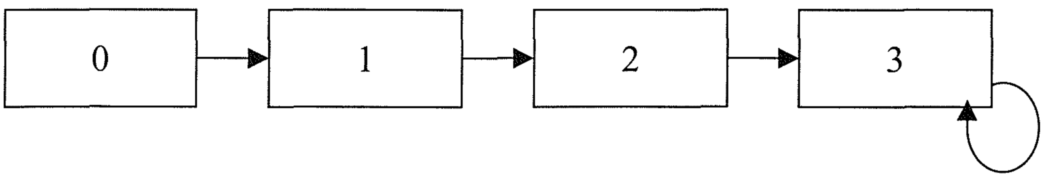


Figure B.2: A simple model

With this simplification a modification can be made to the general notation of the semi-Markov process. Define:

$$f_j(t) = g_{ij}(t), \qquad i, j \in S, t \geq 0$$

and

$$F_j(t) = G_{ij}(t), \qquad i, j \in S, t \geq 0$$

As an example, the probabilities of being in the first three states are as follows:

Probability in state 0 at time t $\quad 1 - \int_0^t f_1(s)ds$

Probability in state 1 at time t $\quad \int_0^t f_1(s)ds - \int_0^t f_1(s)F_2(x-s)ds = \int_0^t f_1(s)(1 - F_2(x-s))ds$

Probability in state 2 at time t $\quad \int_0^t \int_s^t f_1(s)f_2(x-s)dx[1 - F_3(t-x)]dxds$

The expected time of transitions may be calculated as:

$$E(T_1) = \int_0^\infty t f_1(t)dt$$

$$E(T_2) = \int_0^\infty t \int_0^t f_1(s)f_2(t-s)dsdt$$

$$E(T_3) = \int_0^\infty t \int_0^t \int_0^s f_1(x)f_2(s-x)dxf_3(t-s)dsdt$$

For the probability of being in the state of model with state space $S = \{0, 1, 2, \ldots, n\}$ the following association can be observed:

$$P_{State[i]}(t) = R_{(i-1)}(t) - R_i(t)$$

where

$$R_i(t) = \int_0^t r_{(i-1)}(x)\left(\int_0^{t-x} f_i(s)ds\right)dx \quad \text{and} \quad r_i(t) = \frac{dR_i(t)}{dt}, i \in S$$

This association is calculated using the differentiation rule:

$$R_i(t) = \int_{A(t)}^{B(t)} u(t,x)dx$$

$$\Rightarrow r_i(t) = \frac{dR_i(t)}{dt} = \int_{A(t)}^{B(t)} \frac{\partial u(t,x)}{\partial t} dx + u(t,B(t))\frac{dB(t)}{dt} - u(t,A(t))\frac{dA(t)}{dt}$$

Appendix C – Simulation Modelling

Models of various types are frequently used in Operational Research. Essentially they are all representations of a system of interest and are used to investigate possible effects of different scenarios, for example a new policy or the redesign of the system. The simplest models are probably physical (iconic) scale models to experiment with. Such models have limited use as they lack flexibility (are highly project specific) and are static (they are unable to show how various factors interact dynamically). Mathematical models, however allow a system to be represented in terms of logical and quantitative relationships and are then manipulated and changed to see how models react.

Once a mathematical model has been built it must be examined to see how it can answer the questions of interest about the system it is supposed to represent. If the model is simple enough, it may be possible to work with its relationships and quantities to obtain an exact, analytical solution. The analytical solution of stochastic models, as necessary for the modelling of hospital capacities, presents a formidable challenge. Unless restricting assumptions are applied, many stochastic models are impossible to solve analytically. Numerical methods are needed for solving realistic stochastic models. A computer simulation need not, however, make such stringent simplifying assumptions and is an ideal tool for modelling hospital capacities. The aim of the simulation is the solution of a model which mimics the behaviour of a real-life system over time. A single run of the model corresponds to one realisation of the system under particular defined circumstances. Figure C.1 summarises the possible ways to study a healthcare system. See Pidd (1998), Davies and O'Keefe (1989) and Law and Kelton (1991) for a more exhaustive review of simulation modelling.

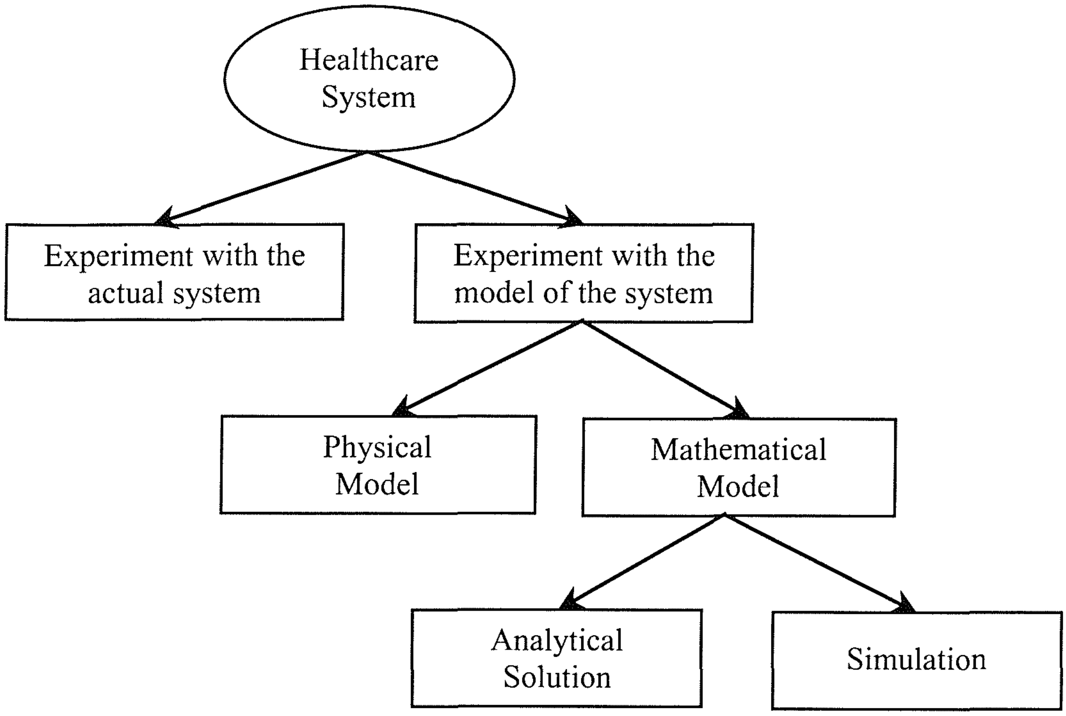


Figure C.1: Ways to study a healthcare system

Simulations are particularly useful for solving the complex models for the flow of patients through hospital because they can:

- Incorporate and recognise complex patient-flows and hospital rules governing these flows.
- Capture the variability of patients in the community (for example demand patterns, and length of stay distributions).
- Be easily understood and used by health managers and planners.
- Model large periods of time very quickly.
- Reduce the risk of reliance on potentially erroneous or unavailable data by incorporating expert opinion.
- Monitor use of resources over time, which aids the planning and management for hospital resource decision-making.
- Model a variety of different scenarios with minimum additional effort.

C.1 Discrete- event Simulation

The modelling of the individual patients through a hospital system is best accomplished with a discrete-event computer simulation. Discrete-event modelling observes the system whenever a change of state in the system occurs. The time at which the change takes place is termed an *event*, the objects which move in the system are known as *entities*, and an *activity* is defined to be the set of actions taken by the entities at each event. Finally, the *simulation clock* measures the passing of time. In the context of modelling patient-flows through hospital, entities are individual patients and the events constitute a patient changing from one state to another, corresponding to movement of patients around the hospital (for example moving from an inpatient bed to queueing for theatre or moving from an ICU bed to an HDU bed).

Two differing approaches to writing discrete-event simulation programs exist.

- ***The event-based approach***. The programmer must decide what constitutes a change of state (event) and the changes that result from each event. Events are scheduled by being placed on the *simulation calendar* (a time-ordered list of events). The simulation clock moves to the head of the simulation calendar (time scan) and all events due at that time are subsequently executed (event execution). To summarise the advantages and disadvantages of the event-based approach:

  Programs run very efficiently and quickly ✓

  Programs are difficult to write ✗

- ***The activity-based approach***. The programmer must list all the conditions necessary for each activity to take place. The program then repeatedly tests for these conditions at every instant of time, and if satisfied, the respective activity occurs.

  Programs are easy to write and modify ✓

  Programs run very slowly ✗

C.2 The Three- phase Approach

The pioneering three-phase approach was developed by Tocher (1963) and combines the simplicity of the activity-based approach with the efficiency of the event-based approach. Two different types of events exist:

- ***Bound events (or B-events)***, which may be scheduled once placed in time-order on to the simulation calendar.
- ***Conditional events (or C-events)***, whose executions depend on the fulfilment of certain necessary conditions.

The three-phase approach is named after the adopted three-phase structure:

Time Phase. Advance the simulation clock to the time of the B-event at the top of the calendar (i.e. the next event)

B-Phase. Execute all B-events that are due to happen at this time

C-Phase. Test all the conditions for all of the C-events and execute any that are now satisfied.

The three-phases are repeated until either the calendar is empty, or until the *duration* (specified time to run the simulation) has been reached.

The Operational Research group at the University of Southampton has developed a three-phase simulation shell called TOCHSIM, which was originally written in Borland's Turbo Pascal version 7.0. TOCHSIM consists of skeleton procedures for handling the queues used to hold the calendar of events and the entities in each state. Procedures to sample from various probability distributions are also included. Designing graphics in TOCHSIM (menus, input screens, output screens, graphs etc.) can however be a laborious process. As a consequence, Delphi software, which incorporates Pascal code and easy to design graphics, is now widely used within the research group.

C.3 Historical Simulation Developments

The emergence of computer simulation is generally traced back to post war developments in OR and numerical modelling. Subsequent developments in computing hardware and the evolution of high-level simulation languages has further popularised the application of computer simulation to a range of real-life modelling problems (Zeigler, 1979). SIMULA is one of the earliest simulation languages, and is credited with the introduction of object-oriented authoring concepts into computer science. Other early examples of simulation programming languages include GASP, GPSS, SIMSCRIPT and SLAM.

The establishment of several specialist groups and societies for computer simulation also marked the post-war period. One prominent establishment was the Society for Computer Simulation (SCS). The SCS has grown consistently since its birth.

A body of literature on the topic began to emerge during the 1960's, including important source books (Tocher, 1963 and Forrester, 1961 as examples) and a range of other publications (see Oren, 1974, 1976). At this time, simulation also found its way into university curricula as a subject in its own right.

The 1970's and beyond has witnessed enormous advances for computer simulation. Increasing computer power and growing accessibility in terms ease of use and reduced costs is largely attributable to this growth. Simulation packages such as Witness, ProModel, Simul8 and Arena have attributed to this success. The range of applications has consequently grown to a point where simulation is now routinely utilised for a vast range of operations in a wide spectrum of industries.

Appendix D – Classification Tools

Chapter Five describes how a number of classification tools were compared using various evaluation criteria. A summary of the findings was presented. This appendix contains detailed results from each of the adopted techniques in the study. It is not possible to provide results for each classification tool for each study. Instead only one study is shown for each tool with the intention of illustrating typical model inputs, outputs and issues.

D.1 Discriminant Analysis

DA is used to classify cases into the values of a categorical dependent. For studies 1 and 3, it was therefore necessary to divide the continuous dependent variable into a number of groups. Using percentile points, the distribution of values (LoS) was divided into ten groups (1-10%, 11-20% etc.). DA was then used to predict membership to each of the 10 groups. For studies 2,4 and 5, the dependent variable was in the necessary categorical (group) form.

Study 5 is used to illustrate the results from DA. In this study we are predicting the onset of diabetic retinopathy using 13 explanatory variables. Diabetic patients may suffer from a number of long-term complications. Retinopathy is a complication with the eyes, which can eventually lead to blindness. It can be treated successfully if detected in time. The dependent variable was a nominal 0-1 variable defining retinopathy. The list of independent variables included sex, height, systolic blood pressure, diastolic blood pressure, number of diabetic years, glucose level (HbA), type of diabetes, creatinine level, cholesterol level and age of onset of diabetes.

Presented tables are taken from SPSS (version 10) output.

Descriptive Statistics

Variable	Description	Min	Max	Mean	Std. Dev
S1	Sex of patient (M/F)	0	1	Cat	Cat
DTYPE	Diabetes type (IDDM/NIDDM)	0	1	Cat	Cat
TOESCORE	Toescore	6.48	14.30	10.5	1.8
HEIGHT	Height of patient	131.0	200.0	167.8	9.7
SBPMEAN	Systolic blood pressure	90.0	220.0	143.7	19.8
DBPMEAN	Diastolic blood pressure	51.8	116.7	80.2	9.5
HBAMEAN	Glucose level	4.8	14.9	8.7	1.5
CHOLMEAN	Cholesterol level	2.1	12.4	5.8	1.1
CRETMEAN	Creatinine level	50.5	644.6	98.3	35.7
BMIMEAN	Body Mass Index	18.3	49.0	28.1	5.0
DIAB_YEA	Years as a diabetic	0.0	66.5	9.5	11.1
AGE	Age of patient	10.4	93.3	61.2	16.5
AGE_DIAG	Age of onset of diabetes	0.4	90.6	51.7	20.2
RET	Retinopathy (Dept. Var)	0	1	Cat	Cat

Standardized Canonical Discriminant Function

	Function
	1
S1	.225
DTYP	.254
TOESCOR	.362
HEIGH	-.292
SBPMEA	.286
DBPMEA	.003
HBAMEA	.215
CHOLMEA	-.083
CRETMEA	-.038
BMIMEA	-.110
DIAB_YE	.853
AGE	-.146

Classification a,b

			Predicted Membership		
		RE	0	1	Total
Origin	Count	0	756	225	981
		1	236	501	737
	%	0	77.1	22.9	100.0
		1	32.0	68.0	100.0
Cross-validated a	Count	0	755	226	981
		1	242	495	737
	%	0	77.0	23.0	100.0
		1	32.8	67.2	100.0

a. 73.2% of original grouped cases correctly

b. 72.8% of cross-validated grouped cases correctly

D.2 Regression Models

Regression analysis is concerned with investigating the relationship between several variables in the presence of random error. In particular we build a model in which the dependent variable is expressed as a linear combination of the independent or explanatory variables.

The technique is illustrated using study 3 where we predict LoS on the ward based on routinely collected data from a hospital patient management. A number of socio-economic and medical variables are routinely collected. Four variables (age, intent, status and sex) were selected for this study. The level of variance in LoS is high (variance of 56.6 with a mean LoS of 4.3) and the number of records large (17,974). SPSS (version 10) output is presented. The best result was obtained after a logarithmic transformation was applied to LoS, although a number of other transformations were evaluated.

Descriptive Statistics

Variable	Description	Min	Max	Mean	Std. Dev
X1	Sex of patient (M/F)	0	1	Cat	Cat
AGE	Age of patient (M/F)	0	98	56.6	18.2
I1	Intent (Day-case/Inpatient)	0	1	10.5	1.8
S1	Status (Emergency/Elective)	0	1	Cat	Cat
LOS	Hospital LoS (Dept. Var)	1	220	4.3	7.5

The following additional and interaction terms were defined:

Variable	Interaction
AGE2	AGE * AGE
F1	AGE * I1
F2	AGE * X1
F3	AGE * S1
F4	I1 * S1
F5	I1 * X1
F6	I1 * S1
F7	AGE * I1 * X1

The results from CART gave a helpful insight into possible interaction terms. Age, intent (I1) and status (S1) were important explanatory variables. Only a three-way interaction term (excluding sex) was used in the model. The results from the regression analysis are consistent with the results from CART.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.593[a]	.352	.352	.7705
2	.628[b]	.394	.394	.7449
3	.630[c]	.396	.396	.7439
4	.630[d]	.397	.397	.7436

a. Predictors: (Constant), F1

b. Predictors: (Constant), F1, F4

c. Predictors: (Constant), F1, F4, I1

d. Predictors: (Constant), F1, F4, I1, F3

ANOVA[e]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	2896.515	1	2896.515	4879.140	.000[a]
	Residual	5328.033	8975	.594		
	Total	8224.548	8976			
2	Regression	3244.575	2	1622.287	2923.390	.000[b]
	Residual	4979.974	8974	.555		
	Total	8224.548	8976			
3	Regression	3259.545	3	1086.515	1963.604	.000[c]
	Residual	4965.003	8973	.553		
	Total	8224.548	8976			
4	Regression	3263.621	4	815.905	1475.592	.000[d]
	Residual	4960.927	8972	.553		
	Total	8224.548	8976			

a. Predictors: (Constant), F1

b. Predictors: (Constant), F1, F4

c. Predictors: (Constant), F1, F4, I1

d. Predictors: (Constant), F1, F4, I1, F3

e. Dependent Variable: LOGLOS

$$LOGLOS = (0.3940 * I1) + (0.0084 * AGE * I1) + (0.0039 * AGE * S1) + (0.2270 * I1 * S1)$$

Correlations

		LOS	LOSPRED
LOS	Pearson	1.000	**.620***
	Sig. (2-tailed)		.000
	N	8987	8987

*. Correlation is significant at the 0.01 level (2-tailed).

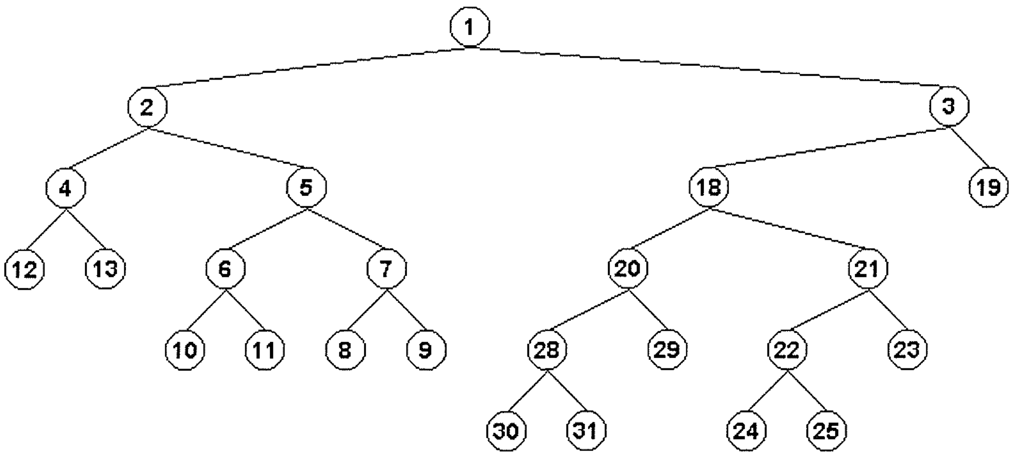
D.3 Regression Trees (CART)

The CART method produces a tree that, by answering a series of yes/no questions, can be used to classify each patient. An algorithm is used to split the original dataset into sub-populations of increasing purity (decreasing variance or deviance). CART is demonstrated using study 4. Given a number of explanatory variables, we wish to predict the probability of a pregnant woman having a complicated delivery. Complications include the need to induce the baby, caesarean section and stillbirth. A successful classification could help to flag women at high-risk for whom we might offer more dedicated care throughout labour and during delivery.

Apollo (Chapter 5) was used to train and then test the data. There were 2,402 records (1,201 in both the training and testing datasets) and 16 variables of interest (8 continuous and 8 categorical).

Descriptive Statistics

Variable	Description	Min	Max	Mean	Std. Dev
SMOKING	If woman smokes (Y/N)	0	1	Cat	Cat
EPILEPSY	History of epilepsy (Y/N)	0	1	Cat	Cat
HYPERTEN	History of hypertension (Y/N)	0	1	Cat	Cat
PARITY	Previous number of children	0	8	0.9	1.0
CAESAREAN	Previous number of caesareans	0	2	0.02	0.21
HEIGHT	Height of woman	1.4	1.9	1.6	0.06
WEIGHTMO	Weight of woman	39.0	146.0	66.7	13.2
DIABETES	Is diabetic (Y/N)	0	1	Cat	Cat
BP	Blood pressure	40.0	116.0	72.7	9.4
AGEMO	Age of woman	18.3	49.0	28.1	5.0
RACEMO	Race of mother (2 categories)	0	1	Cat	Cat
GEST_WKS	Gestation (in weeks)	23.0	42.0	39.3	1.9
B_INDEX	Body Mass Index of woman	15.8	57.0	24.8	4.7
NO_BABES	Number of expected babies	1	2	1.02	0.14
S1	Sex of baby (M/F)	0	1	Cat	Cat
SPONTDEL	Spontaneous delivery (Dept. Var)	0	1	Cat	Cat



Terminal Node Summary (numbers shown for testing dataset)

Node	Path	No. of records	SpontDel Y (%)	SpontDel N (%)
1	All Data	1,201	75	25
12	Parity = 0; Gestation <= 3	66	58	42
13	Parity = 0; Gestation > 37	172	66	34
10	Parity = 0; Gestation <= 40; Age <= 26	92	63	37
11	Parity = 0; Gestation <= 40; Age > 26	85	49	51
8	Parity = 0; Gestation > 40; Age <= 25	45	51	49
9	Parity = 0; Gestation > 40; Age > 25	85	46	54
30	Parity > 0; Caesarean = 0; BMI <= 20; Sex = M	51	92	8
31	Parity > 0; Caesarean = 0; BMI <= 20; Sex = F	50	98	2
29	Parity > 0; Caesarean = 0; Weight <= 64; BMI > 20	213	93	7
24	Parity > 0; Caesarean = 0; Weight <= 81; Gest <= 39	95	94	6
25	Parity > 0; Caesarean = 0; Weight <= 81; Gest > 39	126	95	5
23	Parity > 0; Caesarean = 0; Weight > 81; Gest <= 39	69	86	14
19	Parity > 0; Caesarean > 0	52	55	45

Classification

			Predicted Membership		
			Y	N	Total
Original	Count	Y	814	81	895
		N	173	123	296
	%	Y	90.9	9.1	100.0
		N	58.8	41.2	100.0

Overall classification rate of 78.0%

D.4 Artificial Neural Networks

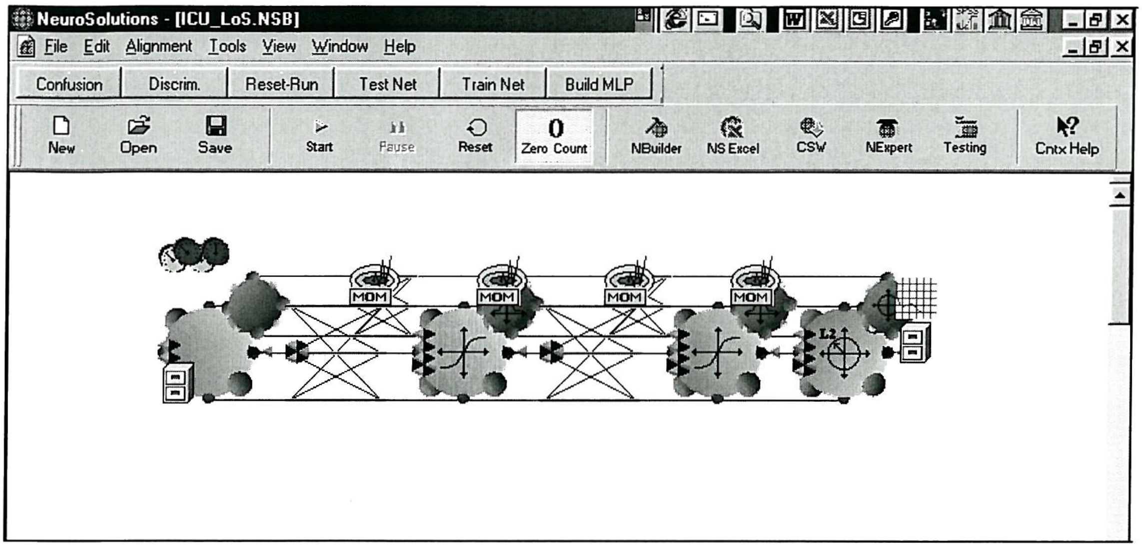
NeuroSolutions (version 3) was used to train, cross-validate and test each network. NeuroSolutions makes use of Windows™ 'point & click' technology to make the software easy to use, enabling non-experienced users to begin building and testing models with little training and in minimal time. Caution however should be exercised when selecting suitable network initial conditions, such as step-sizes and numbers of hidden layers. Only with experience of using the software is it possible to get a feel for appropriate model configurations.

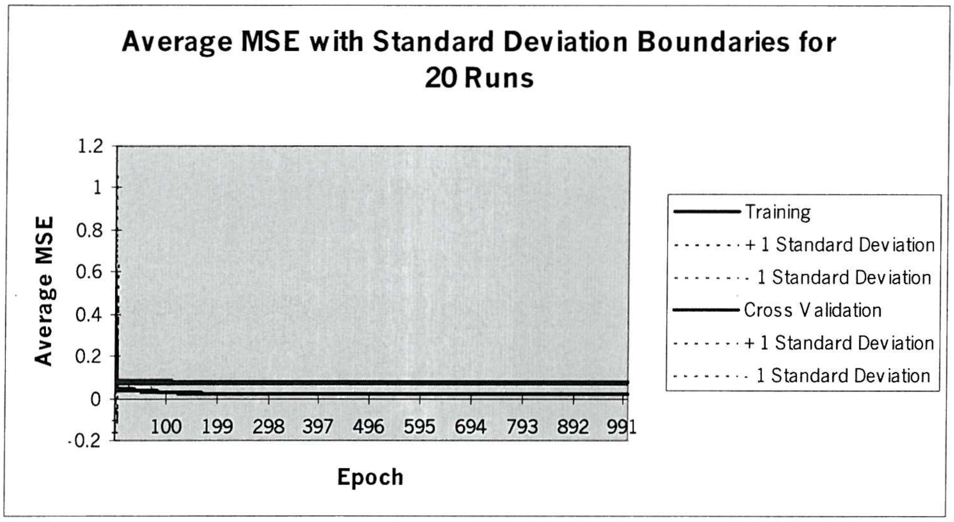
We illustrate the Neural Network approach in predicting ICU LoS (study 1). Routinely collected variables of interest include the patient's age, sex, outcome, source (A/E, HDU, Theatre or Ward), admission status (elective or emergency) and hospital speciality (ENT, General Surgery, Medicine, Orthopaedics, Thoracic Medicine, Trauma or Vascular Surgery). Categorical variables with multiple classes (*g*) were re-assigned to (*g*-1) binary groups (0-1).

Descriptive Statistics

Variable	Description	Min	Max	Mean	Std. Dev
AGE	Age	1	97	60.7	19.8
X1	Sex Male (Y/N); Else Female	0	1	Cat	Cat
DAYS_ICU	LoS (Dept. Var)	0.1	25.2	2.4	3.4
S1	ENT (Y/N)	0	1	Cat	Cat
S2	Gen Surgery (Y/N)	0	1	Cat	Cat
S3	Medicine (Y/N)	0	1	Cat	Cat
S4	Orthopaedics (Y/N)	0	1	Cat	Cat
S5	Thoracic (Y/N)	0	1	Cat	Cat
S6	Trauma (Y/N); if N to all then Vascular	0	1	Cat	Cat
A1	Elective (Y/N); Else Emergency	0	1	Cat	Cat
P1	A/E (Y/N)	0	1	Cat	Cat
P2	HDU (Y/N)	0	1	Cat	Cat
P3	Theatre (Y/N); if N to all then Ward	0	1	Cat	Cat
O1	Outcome Alive (Y/N); Else Died	0	1	Cat	Cat

The dataset contained 582 records. Half was used to train the network and half to test the model. 10% of the training data was used for cross-validation. Missing values were removed as requested by the software. The best accuracy was achieved using a multiplayer perceptron network with one hidden layer, a supervised learning control with 1,000 epochs (iterations over the training set), a step-size of 0.7 and a momentum rate of 0.5.





Performance	***DAYS_ICU***
MSE	11.6452
NMSE	0.9286
MAE	1.8410
Min Abs Error	0.0152
Max Abs Error	23.6978
r	**0.3280**

Appendix E – Capacity Planning Questionnaires

In was necessary to obtain information on the movement of patients through the hospital, so that the developed PROMPT model could be fine-tuned to reflect local hospital real-life conditions. Each speciality manager supplied the necessary information on admission rules, deferral times and theatre information, whilst the bed-managers constructed patient priority listings. The bed and theatre questionnaires used at The Royal Berkshire and Battle Hospitals NHS Trust are shown below. Similar questionnaires were adopted by The Portsmouth Hospitals NHS Trust.

E.1 Bed Capacity Questionnaire

Bed capacity questionnaire sent to all Specialty Managers.

Bed Capacity Planning Questionnaire

The Institute of Modelling for Healthcare (IMH), part of the University of Southampton, are working with the Hospital Process Redesign (HPR) team to help answer questions on bed capacity planning.

In order to fully understand patient flows through the hospital, I would very much appreciate your time in completing the attached short questionnaire. This questionnaire is being completed for every specialty within the hospital. It is vital that each specialty completes and returns this form so that I can incorporate the necessary detail into my work.

If you have any questions about the form, please contact either Jana Dale (HPR) or myself. Please return the questionnaire as soon as possible to HPR.

Thank you for your time and effort.

Paul Harper
Institute of Modelling for Healthcare
University of Southampton.

Bed Capacity Planning Questionnaire

Specialty: ____________________

Completed by: ____________________ **Ext:**________ **Date:** ________

1. Patient Flows

Into which wards are patients admitted for your specialty? Please indicate order of priority if applicable (e.g. 1.West Ward 2. Greenlands)

[**Note:** If ward lists and priorities are different for certain patient subgroups (e.g. Planned vs Emergency, Diagnosis), please provide lists on a separate page for each group identified]

Ward	Priority

2. Planned Patient Deferrals

If a planned patient has to be deferred, please indicate below the typical time a patient must wait before re-admittance. Additionally, state whether deferral time is usually fixed (e.g. 2 weeks) or whether it may vary (e.g. between 1 and 3 weeks). Remember that although times may significantly alter, we are only after a "typical" time.

Deferral Waiting Time:

- Fixed time(weeks) [ ]

or

- Between [ ] and [ ] weeks

3. Emergency Status

Do you enforce a procedure by where a planned patient is "upgraded" to emergency status if they have be deferred on a number of previous occasions? e.g. If a planned patient has been deferred on 3 separate occasions, they will now been treated with emergency status and given priority. If applicable, please state number of deferrals before emergency status below.

Number of Deferrals: [ ]

4. Comments

Please write any other relevant comments below.

Thank you for taking time to complete this form

Please return to:

HPR
The Royal Berkshire Hospital
Craven Road
Reading.

E.2 Operating Theatre Questionnaire

Theatre capacity questionnaire sent to all Specialty Managers.

Theatre Capacity Planning Questionnaire

Specialty: ______________________________

Completed by: __________________________ **Ext:**________ **Date:** ________

Current Theatre Sessions

Day	Session Number	Start Time	Duration (or End Time)

Current list of Surgeons

Surgeon's name:

Total number of in-patient beds across all specialty wards:

What happens to day-cases? Do they use in-patient beds? Do they share in-patient lists?

What happens to emergency patients? Do they use the same sessions as in-patients or a different session/theatre?

Any other relevant comments about current practice?

Any scenarios you have in mind that you would like examined? e.g. weekend theatre (and for which patients), changing the elective weekday admission profile by operation category etc.

Appendix F – Statistical Distributions

The developed simulation programs as described within this thesis allow the user to choose between five statistical distributions for fitting the necessary transition times, for example to LoS and operation times. These distributions are described in further detail below.

F.1 Exponential Distribution

This is an important distribution in reliability work, as it has the same central limiting relationship to life statistics as the normal distribution has to non-life statistics. It describes the constant failure rate situation. The pdf with mean life λ is given by:

$$f(x) = \lambda \exp(-\lambda x)$$

The mean and standard deviation of the distribution is $1/\lambda$. Historically the exponential distribution was the first widely used lifetime distribution partly because of the availability of simple statistical methods.

F.2 Normal Distribution

The normal pdf is given by:

$$f(x) = \frac{1}{\sigma(2\pi)^{\frac{1}{2}}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2 \right]$$

where μ is the location parameter, equal to the mean. The mode and the median are coincident with the mean, as the pdf is symmetrical. σ is the scale parameter, equal to the standard deviation.

The normal distribution is a close fit to most applications in quality control and some reliability observations, such as size of machined parts and the lives of items subject to wear out failures, as well as natural phenomena such as heights of adults and strengths of materials.

The cumulative probability function cannot be solved analytically. A number of polynomial and rational approximations are given in Abromowitz *et al.* (1970). The polynomial approximation, which is used in the estimation of parameters, is given by:

$$P(Z \le z) = \begin{cases} 1 - \dfrac{1}{2}\left(1 + c_1 z + c_2 z^2 + c_3 z^3 + c_4 z^4\right)^{-4} + \varepsilon(z) & z > 0 \\ \dfrac{1}{2}\left(1 + c_1 z + c_2 z^2 + c_3 z^3 + c_4 z^4\right)^{-4} + \varepsilon(z) & z < 0 \end{cases}$$

where $|\varepsilon(z)| < 2.5 \times 10^{-4}$

and $z = \left(\dfrac{\mu - x}{\sigma}\right)$, $c_1 = 0.196854$, $c_2 = 0.115194$, $c_3 = 0.000344$, $c_4 = 0.019527$.

F.3 Lognormal Distribution

The lognormal distribution is a more versatile distribution than the normal as it has a range of shapes, and therefore is often a better fit to reliability data, such as for populations with wear out characteristics. Additionally, it does not have the normal distribution's disadvantage of extending below zero to -∞. The lognormal pdf is:

$$f(x) = \frac{1}{\sigma x (2\pi)^{1/2}} \exp\left[-\frac{1}{2}\left(\frac{\ln x - \mu}{\sigma}\right)^2\right]$$

The mean and the standard deviation of the lognormal distribution are given by:

$$Mean = \exp\left(\mu + \frac{\sigma^2}{2}\right), \text{ Standard Deviation } = \left[\exp\left(2\mu + 2\sigma^2\right) - \exp\left(2\mu + \sigma^2\right)\right]^{1/2}$$

where μ and σ are the mean and standard deviation of the ln of the data.

An unattractive property of the lognormal distribution is that the hazard rate increases to a maximum, and then decreases, approaching 0 as x $\rightarrow \infty$.

The cumulative probability function also cannot be solved analytically. The normal polynomial approximation is used in estimating parameters with $z = \left(\frac{\ln x - \mu}{\sigma}\right)$.

F.4 Gamma Distribution

The gamma distribution describes, in reliability terms, the situation when partial failure can exist, i.e. when a given number of partial failures events must occur before an item fails, or the ath failure when time to failure is exponentially distributed. The pdf of the gamma distribution is defined by:

$$f(x) = \frac{\lambda}{\Gamma(a)}(\lambda x)^{a-1}\exp(-\lambda x)$$

where λ is the failure rate and a the number of partial failures per complete failure, or events to generate a failure. $\Gamma(a)$ is the gamma function:

$$\Gamma(a) = \int_0^{\infty} x^{a-1}\exp(-x)dx$$

The mean and standard deviation of the gamma distribution are $\mu = \frac{a}{\lambda}$ and $\sigma = \frac{a^{1/2}}{\lambda}$, respectively. The exponential distribution is a special case of the gamma distribution, when $a = 1$.

The cumulative probability function cannot be solved analytically. The approximation stated here is given in Press *et al.* (1992). Firstly we define the incomplete gamma function as:

$$P(a,x) \equiv \frac{\gamma(a,x)}{\Gamma(a)} \equiv 1 - \frac{\Gamma(a,x)}{\Gamma(a)} \equiv \frac{1}{\Gamma(a)}\int_x^{\infty} e^{-t}t^{a-1}dt$$

There is a series development for γ(a,x) as follows:

$$\gamma(a,x) = e^{-x} x^{a} \sum_{n=0}^{\infty} \frac{\Gamma(a)}{\Gamma(a+1+n)} x^{n}$$

There is a continued fraction development for $\Gamma(a,x)$ as follows:

$$\Gamma(a,x) = e^{-x} x^{a} \left(\frac{1}{x+} \frac{1-a}{1+} \frac{1}{x+} \frac{2-a}{1+} \frac{2}{x+} \right)$$

It is proved that the series development converges rapidly for x less than about a+1, while the continued fraction converges rapidly for x greater than about a+1.

The gamma function is defined by the integral

$$\Gamma(z) = \int_{0}^{\infty} t^{z-1} e^{-t} dt$$

An approximation uses the fact that certain integer choices of ξ and N, and for certain coefficients $c_1, c_2, \ldots, c_N$, the gamma function is given by:

$$\Gamma(z+1) = \left(z + \xi + \frac{1}{2} \right)^{z+\frac{1}{2}} e^{-(z+\xi+1)} \times \sqrt{2\pi} \left[c + \frac{c}{z+1} + \frac{c}{z+2} + \cdots + \frac{c}{z+N} + \varepsilon \right]$$

The error term for $\xi = 5$, $N = 6$ and with a certain set of c's, is smaller than 2×10^{-10}. It is better to implement ln $\Gamma(x)$ than $\Gamma(x)$ since the latter will overflow many computers' floating point representation at quite modest values of x. Often the gamma function is used in calculations where the large values of $\Gamma(x)$ are divided by other large numbers resulting in an ordinary number.

F.5 Weibull Distribution

The Weibull distribution has the great advantage in reliability work that by adjusting the distribution parameters it can be made to fit many life distributions. The Weibull pdf is:

$$f(t) = \frac{\beta}{\alpha} t^{\beta-1} \exp\left[-\left(\frac{t}{\alpha}\right)^{\beta}\right]$$

The hazard rate is $\frac{\beta}{\alpha} t^{\beta-1}$. β is the shape parameter, α is the scale parameter.

When $\beta = 1$, we obtain the exponential reliability function (constant) with mean life equal to α. The exponential distribution is a special case of the Weibull distribution. When $\beta < 1$, this results in a decreasing failure rate reliability function. When $\beta > 1$, we obtain the increasing failure rate reliability function.

The mean and the variance are given by $\alpha\Gamma\left(1+\frac{1}{\beta}\right)$ and $\alpha^2\left[\Gamma\left(1+\frac{2}{\beta}\right) - \Gamma\left(\frac{1}{\beta}\right)^2\right]$.

Appendix G – Programming Structure and the TOCHSIM Shell

It is not appropriate to include a complete listing of the code for the developed programs as described within this thesis. Instead, a description of the main structure, together with some of the units with corresponding procedures and functions, are provided. The PROMPT model was programmed using the TOCHSIM simulation shell. TOCHSIM has benefited from language extensions to give the full power of object-orientated programming: more structure and modularity, more abstraction and reusability built into the simulation shell (see section 6.3.3).

G.1 Application Structure

For each application there exists a project consisting of:

- The project (.DPR) file
- The unit (.pas) files
- The form (.DFM) files
- Source code for units without forms

The project file keeps track of all unit and form files in the application. For example, the project source code for the PROMPT program looks like:

Program PROMPT;	Project name
Uses	Uses clause
Forms;	
Menu in 'MENU.PAS' {MainMenuForm},	Form identifiers
BEVENT1 **in** 'BEVENT1.PAS',	Source code without forms
BEVENT2 **in** 'BEVENT2.PAS',	
.	
.	
{$R *.RES}	Complier directive

```
begin
    Application.CreateForm(TMainMenuForm,MainMenuForm);
    Application.Run;
end.
```

(Note that **bold** fonts are reserved words in Delphi).

In accordance with the distinct elements of the PROMPT model, the following sections describe the coding for patient groups, care units (beds), theatres and workforce.

G.2 Patient Group Code

Defined patient groups are fundamentally either emergency or elective in nature (priority is given to emergencies during the arrival process). Furthermore, if the theatre module is activated, groups are either procedure or non-procedure (require or do not require the use of the theatre respectively). A *Patient_Group* class has been created containing the necessary information on status (emergency or elective), theatre needs (procedure or non-procedure), a care unit priority list (Tlevel_of_care), LoS and operation time (TLoS_distribution), arrival information (Tarrival_vars) and workforce needs (Tpatient_dep_list).

```
TPatient_Group = class
        status: boolean;
        theatre_needs : boolean;
        LoS : TLoS_distribution;
        arrival_var : Tarrival_vars;
        op_time : TLoS_distribution;
        level_of_care : Tlevel_of_care;
        description : Tstringlist;
        patient_dep_list : Tpatient_dep;
        procedure init;
        procedure load(var f : textfile);
        procedure save(var f : textfile);
        end;
```

```
Tarrival_vars = class
    Arrivalhour : array[1..6] of integer;
    ArrivalRateVDay : array[1..7] of integer;
    Arrivalmth : array[1..12] of integer;
    YearlyRate : integer;
    procedure load(var f : textfile);
    procedure save(var f: t extfile);
    end;
```

```
Tlevel_of_care = class
    next_level_of_care:Tstringlist;
    procedure load(var f:textfile);
    procedure save(var f:textfile);
    end;
```

```
Tdistribution = (d_weibull, d_normal, d_lognormal, d_negexp, d_gamma);
```

The LoS and operation times are generated using the LoS_distribution class defined below:

```
TLoS_distribution = class
            distribution : Tdistribution;
            parameter1, parameter2 : real;
            min_stay, max_stay, reduced_los : real;
            mean : real;
            increase_los : integer;
            function random_deviate : real;
            procedure load(var f : textfile);
            procedure save(var f : textfile);
            end;

function TLoS_distribution.random_deviate : real;
var transtime : real;
begin
      repeat
            case distribution of
            d_weibull : transtime := rnd.weibull(parameter1, parameter2);
            d_normal : transtime := rnd.normal(parameter1, parameter2);
            d_lognormal : transtime := rnd.lognormal(parameter1, parameter2);
            d_negexp : transtime := rnd.negexp(parameter1, parameter2);
            d_gamma : transtime := rnd.gamma(parameter1, parameter2);
            end;
      until ((transtime >= min_stay) and (transtime < max_stay));
            if reduced_los >=1 then begin  {reduced_los is a percentage shift to LoS}
                  if increase_los = 0 then  {1 for an increased and 0 for a decreased shift}
                  transtime := transtime - ((reduced_los/100)*transtime)
                  else
                  transtime := transtime + ((reduced_los/100)*transtime);
            end;
      random_deviate := transtime;
end;
```

The Tstringlist object is frequently used throughout the program in order to maintain and manage lists of strings, so that classes may add, delete, insert, move or exchange strings with pre-defined functions in a controlled and safe environment. For example, *next_level_of_care* (care unit priority list attached to each class of patient groups) is defined as a Tstringlist. In the model, a combo-box on the patient group form enables the user to select from a number of currently defined care units and add it/delete it from the list, or move the position (priority) of the care unit on the list (Figure G.1).

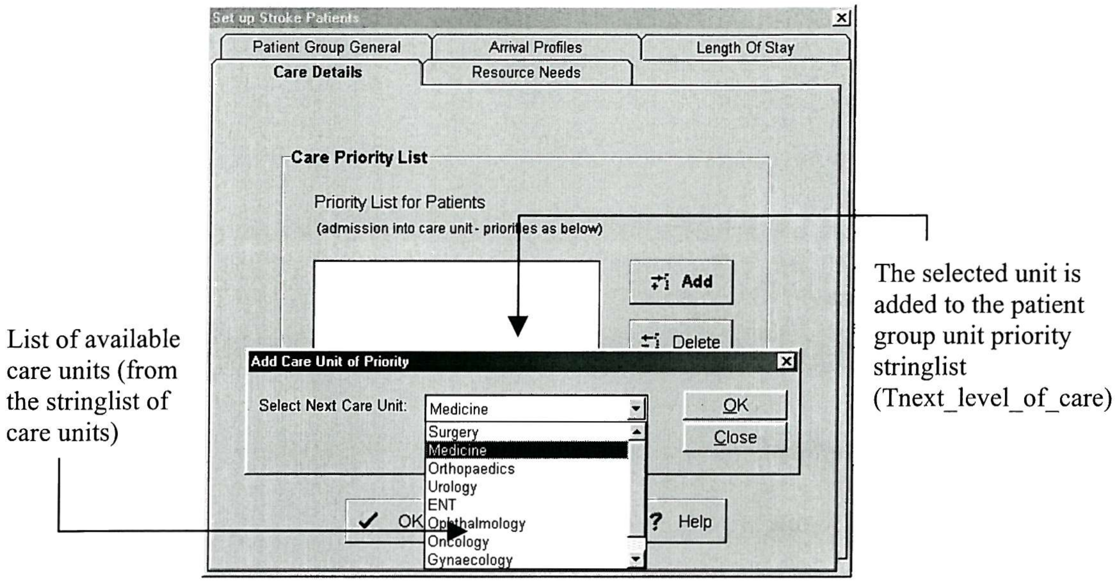


Figure G.1: Editing care unit priority lists: the TStringList object in action

G.3 Care Unit (Bed) Code

A patient attempts to find a suitable bed in the hospital by examining each care unit on their priority list in turn and examining the current care unit bed availability whilst adhering to relevant admission rules. The class *Care_Unit* has been created containing all of the necessary statistics (Pcounter and Pwaitstatistic), the unit name and the number of beds in each unit (Tresource).

```
TCare_Unit = class
        patient_count : integer;          Patients in the unit
        care_unit_beds : Tresource;       Number of beds
        number_of_beds : Tnum_beds;
        waitingtime : real;
        deferral : Tdeferment;            Admission rule variables
        patient_freq : Pcounter;
        arrivals : Pcounter;
        transfers : Pcounter;
        no_stayed : Pcounter;             Statistics
        no_moved : Pcounter;
        patientwaitingtime : Pwaitstatistic;
        patientlostime : Pwaitstatistic;
        monthly_bed_time_used : array [1..12] of real;
```

```
procedure init(wardname : string; def_time : real; no_def_to_emerg,
              no_free, fixed_def_time : integer; time_waiting : real);
procedure changename(wardname : string);
procedure remove;
constructor create;
end;
```

G.4 Operating Theatre Code

An operating theatre is described by the TOpt class that contains all of the necessary statistics and rules governing its use. The Ttheatre_times class controls the opening times and duration of each theatre session (by day).

```
TOpt =class
      number_of_sessions : Tnum_sessions;
      theatre_name : string;                         ⎫
      times : Ttheatre_times;                        ⎪
      overrun_percentage,                            ⎬ Theatre scheduling rules
      schedule_method : integer;                     ⎪
      cut_point : integer;                           ⎭
      currently_being_used : integer;                                          ⎫
      time_in_use, time_over, time_under : real;                               ⎪
      no_operations, sessions_under, sessions_over : Pcounter;                 ⎬ Statistics
      operation_waiting_time, operation_server_time,                           ⎪
      slack_time : Pwaitstatistic;                                             ⎭
      constructor create;
      procedure load(var f : textfile);
      procedure save(var f : textfile);
      procedure remove;
      destructor destroy;
      end;

Ttheatre_times = class
   duration:integer;
   starttime:string;
   currentlybeingused:boolean;
   starttimeindays,endtimeindays:real;
   lasttimeused:real;
   constructor create;
   procedure save(var f:textfile);
   procedure load(var f:textfile);
   procedure copyandpaste(var D:Ttheatretimes);
   function acquire(timerequired,group_mean:real):boolean;
   function overrun(timerequired,transtime:real):boolean;
   function return:boolean;
   procedure produceactualtimes;
   destructor destroy;
   end;
```

G.5 Workforce Code

A number of workforce resources may be defined. For each patient group, the necessary workforce-to-patient ratios must be defined. This is further broken down by patient dependency state and shift of the day.

A Tstringlist of human resources is constructed (THuman_Resource).

```
THuman_Resource = class(Tstringlist)
        constructor create;
        procedure load(var f : textfile);
        procedure save(var f : textfile);
        destructor destroy;
        end;
```

Patient dependency lists are created within the TPatient_Group class (section E.2) which define the number and duration of each dependency state, together with the numbers of each resource required for each of the three daily shift (time_period_1, 2, and 3).

```
Tpatient_dep = class
        name : string;
        percent : integer;
        time_period_1 : Ttime_resource_list;
        time_period_2 : Ttime_resource_list;
        time_period_3 : Ttime_resource_list;
        constructor create;
        procedure load(var f : textfile);
        procedure save(var f : textfile);
        procedure copy_info(old_info : Tpatient_dep);
        destructor destroy;
        end;
```

G.6 Simulation Object

The inherited simulation object is used by queues, resources, all statistic objects, the simulation timer and the simulation model. The common method for these objects is described by:

```
PsimObj = TsimObj
TsimObj = Object(Tobject)
        Constructor init;
        Procedure Start_runs;
        Procedure Start_new_run;
        Procedure End_run;
        Procedure End_runs;
        Function return_value : real;
        Destructor done;
        end;
```

The object orientated programming approach enables the programmer to hide detail in units that have become part of the simulation shell. For example, when a queue is created in the initialisation at the start of the application, the queue is added to the simulation object list using the following code:

```
Object_list.add_object(self)
```

The programmer no longer needs to empty the queue at the start of the run since it automatically follows the Start_runs procedure in the application.

G.7 B-Events

The B-events for the PROMPT model are:

- Generate initial arrival times and cause next arrivals within each patient group.
- Check, and if necessary adjust, bed numbers to reflect step-up and step-down of beds over time (daily and monthly event).
- Open theatre sessions as necessary (daily event).
- Cause post-operation LoS to commence when patient arrives back from theatre to ward.
- Cause patient to leave hospital on completion of their LoS or cause re-admission arrival time for deferred elective patients.

All B-events are objects of Bevent, defined by:

```
TBevent = Object(TObject)
        Constructor init;
        Procedure Start_new_run;
        Procedure do_event(ent : Pentity);
        Destructor done;
        end;
```

Pseudo code for the B-events are provided below:

b_arrive : TB_event1.init;
initialise entities {for each patient group}
generate initial arrivals {based on arrival profiles for each group}

b_arrive : TB_event1.do_event;
add to queue {for care unit}
cause next arrival

b_arrive : TB_event2.do_event;
release bed {patient leaves}
dispose of entity
check queue {for waiting times of queueing patients}

b_arrive : TB_event3.init;
calculate any changes in bed numbers during year

b_arrive : TB_event3.do_event;
change bed numbers {for each care unit as appropriate over time}

b_arrive : TB_event4.do_event;
cause departure {as patient arrives back from theatre}

b_arrive : TB_event5.do_event;
Open theatre {for each session based on opening times}

G.8 C-Events

The C-events for the PROMPT model are:

- Start patient stay (having already found an available and suitable hospital bed).
- Start operation in theatre (having queued and been admitted in to a suitably available and open theatre session).

C-events are defined by:

```
PCevent = TCevent
    TCevent = Object(TObject)
        {Variables}
        event_on : boolean;
        {Procedures}
        Constructor init;
        Procedure do_event(ent : Pentity);
        Destructor done;
end;
```

Pseudo code for the C-events are:

```
TC_start_service.do_event;
    check bed availability
    if a suitable bed is free:
        acquire bed
        generate LoS
        cause departure
        log arrival time
        update statistics
    else:
        add to queue

TC_theatres.do_event;
    if at least one patient waiting for theatre:
        acquire theatre
        generate operation time
        cause departure from theatre
        update statistics
```

TOCHSIM has been developed over a number of years with the programming skills in particular of Dr Daryl Gove and Dr Simon Jones. My specific contribution to TOCHSIM was to help move the shell from the Pascal to Delphi environment, thus developing a 32-bit object-orientated version.

Appendix H – Hospital Feedback and Research Evaluation

Acceptance and ease of use of the models by the hospital staff are necessary conditions for obtaining the benefits from the developed models. It was the intention that the variety of operational tools and frameworks as developed during the research and as discussed in this thesis meet these necessary conditions. A key aspect of the evaluation process was entailed in the use the models by the participating NHS Trusts. In this context, it was with great pleasure that letters of thanks and appreciation were received from the two major NHS Trusts at Reading and Portsmouth. Messages of thanks were also received from managers and consultants from the various critical care units.

An excellent working relationship has been forged between the Institute of Modelling from Healthcare (IMH) at the University of Southampton and the participating hospitals. As a consequence, a number of future projects have already been proposed (some ideas of which are described within the further research section of Chapter 10). Furthermore, due to the network of senior hospital staff within the UK and in particular between regional NHS Trusts, a number of other hospitals have approached IMH with the intention of conducting similar capacity planning exercises. The letters from the two NHS Trusts are shown on the following pages.

H.1 Correspondence from The Royal Berkshire and Battle Hospitals NHS Trust

ROYAL BERKSHIRE & BATTLE HOSPITALS NHS TRUST

ADULT MEDICINE

London Road, Reading, Berkshire RG1 5AN Tel (0118 87 7911)

Paul Harper
Institute of Modelling for
Healthcare Faculty of Mathematics Studies
University of Southampton
SO17 1BJ

HB/1736/gjc
23 April 2001

Dear Paul

<u>Re: Activity Modelling for the Royal Berkshire and Battle Hospitals NHS Trust</u>

We have now completed our fourth year of activity modelling for the Trust. I wanted to thank you for all the help and support you have given us in developing this tool. It has proved to be very accurate in projecting our activity.

We use this information on an annual basis to modell our activity both for our surgical profiling and our Acute medical capacity. To date entirely based on the modelling we have opened 60 additional medical beds, some being newly staffed beds others re-allocated from other specialities.

The whole activity modelling forms part of our contract negotiations with our local Primary Care Trusts.

It has formed the basis for our winter pressure planning. Our Regional Office visited us in March to evaluate the effectiveness of our plan and we were given a glowing report. There is no doubt with the information of projected peaks and troughs in our activity we can plan to meet the demand with as much efficiency as possible. Thus emergency patients are admitted speedily and elective surgery is not cancelled.

Kind regards

Yours sincerely

Heather Bunce
CSUM
Deputy Director of Operations

H.2 Correspondence from Portsmouth Hospitals NHS Trust

Portsmouth Hospitals NHS
NHS Trust

Dr Peter Howlett MBA
Executive Director

Ground Floor
Room 0.23
De La Court House
Queen Alexandra Hospital
Southwick Hill Road
Cosham
Portsmouth PO6 3LY
Hants
Tel: (023) 92 286000 Ext. 6342
Fax: (023) 92 286073

Mr P Harper
Institute for Modelling for Healthcare
Faculty of Mathematical Studies
University of Southampton
Southampton
Hants
SO17 1BJ

Our Ref: PH/SMHW/012

18th June 2001

Dear Paul

Re: Modelling of Patient Activity for Portsmouth Hospitals NHS Trust

I am writing to thank you for the crucial work that you are doing to support the plans for the redevelopment of Queen Alexandra Hospital. As you know the redevelopment plans involve a £120m building programme combining existing hospital buildings with significant new ones to provide one of the largest acute hospital complexes on the south coast.

In an area as complex as health care identifying the capacity of such facilities against future patient workload is not straightforward. We do have some tools which are widely used in the health service for identifying numbers of beds and numbers of operating theatres for specific workloads in specific specialties, but these are static modelling tools, which I describe as measuring the 'size of the pipe'. The crucial element of your own work and why it is so valuable to us is that your modelling tools provide a dynamic model which enables us to effectively test 'what if' questions. We can take our own current workload figures and ask the question what if we were to attempt to manage this workload through this set of new facilities. We can then anticipate growth in that workload and again test against the facilities. Effectively, your modelling work provides us with a dynamic tool to look at the flow through the pipe. Given that much of our workload is dominated by emergency work which is subject to significant fluctuations, peaks and troughs in workload, it is essential for us to be able to anticipate when we might fail to have sufficient facilities for a particular workload. The failure of hospitals to anticipate fluctuations in workload and to have sufficient facilities to meet workload peaks is of course the subject of a considerable amount of adverse publicity.

Given too the significant levels of investment in new hospital facilities it is essential that we are able to fully test our specification proposals for the building of new hospitals.

I think I can therefore safely say that your own personal input and the development of those modelling tools has been enormously valuable to us. We recognise the degree to which those tools and the modelling is evolving and is the subject of both research and development. We are very happy to be working with you in looking at how that can be used in a very timely way in one of the largest acute hospitals in the country undergoing a massive change programme.

We look forward to making further progress with you in meeting our planning needs and hopefully stimulating your research interests.

Yours sincerely

Dr. Peter Howlett
Director of Planning