ASPECTS OF INFORMATION MANAGEMENT AND RESOURCE ALLOCATION IN
HOSPITALS WITH SPECIAL REFERENCE TO ACCIDENT AND EMERGENCY

GEORGE VASSILACOPOULOS

Royal Holloway and Bedford New College
(University of London)

Submitted for the degree of Ph.D.
ABSTRACT

The management and control process in an Accident and Emergency (A/E) department of a District General Hospital is investigated and the functional relationship between the A/E department and the inpatient hospital service is discussed. Attention is focused on resource allocation and methods are proposed towards reconciling levels of service and resource utilisation.

Within the framework of control problems inside the A/E department, a computerised patient record system has been designed and implemented, on an experimental basis, to allow easy access to patient-related information for performance evaluation. Established statistical techniques are employed to demonstrate how such information can be utilised in medium-term management activities in the A/E department and to provide a sound basis for defining areas where specific problems arise. A method is developed, which uses patient data to the extent that they are routinely available through the patient record system, for allocating physicians to weekly shifts in a way which takes account of the fixed number of physician hours per week; of physician preferences with regard to shifts; and of the patient assessment of the service provided.

With regard to the role of the A/E department as an essential link between the community at large and the hospital service, a simulation model is developed for determining the number of beds in hospital inpatient departments on the basis of expected demand and according to a pre-specified set of measures of hospital efficiency. The measures used are the rapid admission of emergency patients; high occupancy rates; and short lengths of waiting lists. A further study on bed capacity planning concerns the contemplated development of an observation ward in the A/E department. Owing to the increased uncertainty in planning for prospective units, approximation is accepted for the sake of procedural simplicity and an analytic infinite server queueing model is employed to evaluate various numbers of beds for the unit in terms of the average occupancy rates and of hourly and daily service levels.
ACKNOWLEDGEMENTS

I wish to express my gratitude to Professor M.R.C. McDowell for his encouragement, guidance and support.

I am indebted to Professor K.C. Bowen for his keen interest as well as his invaluable comments on the final form of this thesis.

Thanks are due to Mr. P.S.J.B. Rana, FRCS (Ed), Consultant Surgeon, for his assistance in data collection and for the time he devoted to our discussions.

Finally, I should like to acknowledge the debt I owe to the Greek State Scholarship Foundation (I.K.Y.) for their financial support during the course of this study.
"Nothing occurs by chance, but there is a Reason and a Necessity for everything".

Leucippus D 6782

"Reason overrules Necessity by persuading her to guide the greatest part of the things that become towards what is best".

Plato Tim. 47E
Διερεύνεται στους γονείς μου
Σερφώντα και Μαρία
## CONTENTS

Abstract ....................................................... 2  
Acknowledgements ...................................... 3  
Table of contents........................................ 6  
Introduction .................................................. 10  

Chapter 1. A systems approach to the A/E department of St. Peter's hospital .............. 16  
1.1 Introduction .................................. 16  
1.2 Review of the literature.. 17  
1.3 The actual problem situation .............. 20  
1.3.1 Physical and human resources of the A/E department ............................... 22  
1.3.2 Investigation of existing activities ....... 24  
1.4 Systems analysis............................. 30  
1.4.1 Conceptual model building ........ ......... 31  
1.4.2 Definition of problem areas ................. 35  
1.5 Conclusions ................................... 38  

Chapter 2. A/E department patient record system design ........................................ 42  
2.1 Introduction .................................. 42  
2.2 Review of the literature.................... 43  
2.3 Definition of objectives.......... 45  
2.4 Information requirements................. 46  
2.5 Computerised patient record system...... 50
Chapter 3. Aspects of quantitative systems investigation

3.1 Introduction........................................ 59
3.2 Review of the literature........................... 60
3.3 Harmonic analysis..................................... 62
3.3.1 Fourier representation............................ 63
3.3.2 Testing the Poisson hypothesis.................... 66
3.3.3 Data analysis....................................... 68
3.4 Categorical data analysis........................... 71
3.4.1 Methodology....................................... 73
3.4.2 Data analysis....................................... 75
3.5 Conclusions.......................................... 80

Chapter 4. Allocating A/E department physicians to shifts

4.1 Introduction.......................................... 93
4.2 Review of the literature............................ 94
4.3 Discussion of the problem........................... 99
4.4 Problem formulation.................................. 102
4.4.1 Physician allocation per hour of the week.... 102
4.4.2 Physician allocation per shift.................... 105
4.5 The simulation model................................ 106
4.6 Implementation of the method....................... 109
4.7 Conclusions.......................................... 112
Appendix.................................................. 114
Chapter 5. A simulation model for bed allocation to hospital inpatient departments........ 127
5.1 Introduction.................................. 127
5.2 Review of the literature.................... 130
5.3 Problem formulation.......................... 138
5.3.1 Discussion of the problem................... 140
5.3.2 System modelling........................... 144
5.4 Simulation procedure......................... 146
5.5 Data collection and testing.................. 149
5.5.1 Arrival distributions....................... 150
5.5.2 Length of stay distributions................. 155
5.6 Running the model............................ 156
5.6.1 Model validation........................... 158
5.7 Conclusions................................... 162

Chapter 6. Planning a prospective observation ward in the A/E department............... 178
6.1 Introduction.................................. 178
6.2 Review of the literature.................... 180
6.3 Discussion of the problem................... 183
6.4 System modelling............................. 185
6.4.1 Patient arrivals............................ 188
6.4.2 Patient lengths of stay...................... 192
6.5 Queueing modelling........................... 195
6.5.1 The non-stationary case..................... 196
6.6 Conclusions................................... 201
Appendix......................................... 203

Discussion..................................... 212

INTRODUCTION

From the inception of the National Health Service (NHS) in the United Kingdom (UK), there was concern about the provision for the treatment of accidents and the functioning of hospital casualty services as they then were. Two reports which appeared in the early sixties formed the basis for the development of the accident and emergency service at a national level.

A report was published in 1961 from The Accident Services Review Committee of Great Britain and Ireland[4]. Comprehensive provision for all types of injury was envisaged through a three-tier system: a peripheral casualty service for the treatment of minor injuries and the needs of local casual patients, staffed mainly by general practitioners (GPs); an accident unit as part of a District General Hospital (DGH) under consultant supervision, together with the designation of accident beds at a rate of 25 per 100,000 population; and, above that, a number of central accident units, usually at teaching hospitals where there would be additional specialised units so that such a unit, in addition to supplying the needs of the immediate community would be able to treat patients with complicated injuries drawn from a population of perhaps 1-2 million and also serve to co-ordinate the accident arrangements within the 'region'.
A second report appeared in the following year, 1962, from the Standing Medical Advisory Committee of the then Ministry of Health[108]. Broadly speaking the aims were the same, but provision for emergencies other than accidents was also emphasised and the concept of Accident and Emergency (A/E) departments developed. In addition, that report did not support the three-tier concept, making the point that not all the additional specialist services would necessarily be in one hospital within a region. It was thought that the concept of a central accident unit was not sound and a two-tier system of a peripheral service and properly sited A/E departments was preferred. Because of the widening concept of the service, the number of beds suggested for each 100,000 of population rose to 30-35. It was intended that all patients requiring emergency hospital treatment should be taken to such departments and that patients would thereafter be admitted to other hospital departments only if they required specialised treatment.

Since 1962, numerous other reports have been published and formal investigations carried out on A/E services[79,133,82,151,109]. Particular attention has been devoted to the development of the A/E department as a purpose-built modern hospital unit capable of satisfying the need. The proposals which emerged from these studies concern both structural and functional issues including facilities and staffing; management and control in the A/E department; the role and the integration of the A/E department within the total hospital service; and relations and liaison with the
community, peripheral casualty service, general practitioners and ambulance, police and fire services. This thesis is primarily concerned with the management and control process in the A/E department of St. Peter's hospital, Chertsey; attention is also given to the functional relationship of the A/E department with the inpatient hospital service.

Chapter 1 is concerned with the problem identification phase of this study. This phase involved familiarisation with the concept of an A/E department and with the particular system under study (in terms of physical and human resources and of the existing activities); building a conceptual model of the department's activities as derived from a particular viewpoint of the A/E department; and discussions with the consultant on the basis of what was shown in the conceptual model and what was done in the A/E department with respect to resource allocation and management. From these discussions an overall framework of study was set up including the design of a computerised patient record system to support the monitor and control function in the A/E department.

Chapter 2 deals with the design of this computerised patient record system for the A/E department. Emphasis is placed on the provision of pertinent and accurate information to support the daily and longer-term management activities in the A/E department. Standard patient data for clinical purposes and the data-items recommended by the NHS were also collected to be available on request. The system was implemented on an experimental basis for the whole of 1984.
In chapter 3 attention is drawn to some aspects of the patient population of the A/E department. The potential of using established statistical techniques to derive information on the salient features of the A/E department work is demonstrated. Harmonic analysis is applied to describe the hourly pattern of patient demand and categorical data analysis is employed to describe the timeliness of the patient care process with respect to patient waiting and treatment times. The information derived from these analyses provided a sound basis for defining and tackling specific problems within the overall framework discussed in chapter 1.

One of the problems defined was related to patient waiting times. In chapter 4 a method is proposed for a more efficient allocation of physicians to weekly shifts by taking into account that the number of physician hours available per week is fixed and that a feasible set of shifts is provided in advance. Physicians are allocated approximately proportionally to the average patient arrival rate according to the solutions of two dynamic programs. The first, obtains an optimal, with respect to the problem formulated, allocation of physicians per hour-of-week and the second defines the physician allocation policy per shift. The solutions obtained are then assessed by simulation.

A second problem was related to the interaction of the A/E department with the inpatient departments through emergency admissions. Chapter 5 describes a simulation model for bed
capacity planning in inpatient departments. The model is capable of providing information on various criteria of hospital efficiency such as occupancy rates and turnaway rates for emergency patients. The model is validated using data from the General Medical, General Surgical and Orthopaedics departments of St. Peter's hospital. These departments account for more than 60% of the total number of inpatient admissions through the A/E department.

Chapter 6 deals with the problem of estimating the number of beds which is likely to be required in order to meet patient demand in a contemplated observation ward in the A/E department. Issues related to planning for prospective units are thoroughly discussed and consideration is given to the various approaches available in the literature. It is argued that the infinite server queue with Poisson arrivals provides a reasonable approximation for planning purposes; model limitations are fully defined and discussed. The model is used to evaluate the consequences of selecting any number of beds for the unit in terms of average occupancy rates and the percentage of time that patient bed demands are met (service level), for various arrival rates and for various values of the parameters of the patient length of stay distribution considered.

Finally, there is a general discussion on this research in the light of recent proposals towards streamlining the management process within the NHS. Emphasis is placed on resource allocation and management at a facility level and on
associated information requirements.

The term 'A/E department' is used for convenience to refer to similar facilities in other hospitals. In the various contexts, the fact that the full facilities are not necessarily those of an A/E department as formally defined is not important.

The following summarises the main novel points in this thesis.

Innovatively for an A/E department situation, a systems approach was adopted in order to define specific problems explicitly. This resulted in setting up a computerised patient record system designed so as to meet A/E department objectives while serving the information needs of wider systems (e.g. hospital, District). The collected data were analysed statistically to provide a sound basis for the formulation of the problems defined. A new technique was proposed for allocating A/E department physicians to weekly shifts which uses routinely collected data only. A simulation model was developed, using the concept of priority queueing, which determines the number of beds required in hospital inpatient departments so that several criteria of operating efficiency are simultaneously satisfied. Finally, by means of an infinite server queueing model, the number of beds required in the prospective observation ward of the A/E department was determined. A new result was proved which states that when the arrival rate is periodic then the long-run occupancy is also periodic, with period equal to that of the arrival rate, irrespective of the form of the service time distribution.
CHAPTER 1

A SYSTEMS APPROACH TO THE A/E DEPARTMENT OF ST. PETER'S HOSPITAL

1.1 Introduction

Any real situation could in principle be fully represented as a system with far too many variables for us to comprehend its behaviour. Nevertheless, it is common experience that in many human activity systems it is possible to extract a small number of activities that provide a representation of the system that is comprehensible and at the same time seems to 'explain' most of its behaviour as it affects our immediate interest in the situation. The problem situation of our concern was the A/E department of St. Peter's hospital and our immediate interest was in the department's activities related to emergency patient care.

This chapter describes a systems study which was carried out at the initial stage of this research in order to explore the situation in the A/E department of St. Peter's hospital and to identify different ways of viewing its activities so that the resource implications of the type of service provided in the A/E department could be understood better and areas of enquiry and further investigation could be defined.
The overall framework of the study is determined by the particular version of the 'systems approach'\cite{38,39,81} used, emphasising a particular set of concepts, namely human activity systems concepts\cite{38}. As a start, we obtained a detailed picture of the problem situation including physical and human resources, roles of the staff, objectives, activities and relationships to other systems. Then we developed, standing back from the actual situation, a conceptual model of the minimum necessary activities in the A/E department if it was to fulfil its basic purpose. This stage was carried out by first selecting a point of view of the system and then assembling the minimum activities to make it manifest. Having established this model, we initiated a debate with the consultant of the A/E department on the actual activities defined from the detailed picture stage and those in the conceptual model in order to identify areas where improvements might be made in the light of A/E department statutory realities, in the context of the existing health service structure and management, and within the scope of an operational research doctoral thesis.

Before we discuss the stages of our systems study, we present a review of the literature of systems studies in emergency medical services.

1.2 Review of the literature

The nature of systems studies in emergency medical care
varies with the (sub) system of immediate concern and/or the researcher's view of it. The large variety and volume of such studies compels us to limit our review only to a selection.

A number of studies attempt to improve the performance of the emergency medical services (EMS) system which includes transport, communications and medical care sub-systems. Among the most comprehensive studies are those carried out as part of the Yale trauma program in Connecticut[57,60,149,150] and of the Illinois trauma program[30,31,32]. Both studies investigate several aspects of the EMS system with emphasis on the nature of fatal motor vehicle accidents and implications for the hospitals, on the economics of emergency transportation services, on physician staffing of hospital emergency departments and on the effect of the emergency department on the inpatient service. Subsequently, other studies appear in the same area. Hamilton[72] investigates the EMS system in Philadelphia and proposes a simulation model for planning and evaluation of alternatives, Ittig[80] emphasises the need for organisational changes in ambulatory health systems and suggests a linear programming model for service-mix planning, and Sytkowski et al[142] suggest an analytic method for the evaluation of rural emergency medical services using as outcome measure of performance the changes that took place in cardiac morbidity and mortality.

Several studies focus attention on the efficiency and effectiveness of ambulance services. Groom[69] evaluates the
emergency ambulance service in West Glamorgan Area Health Authority by estimating the response time distributions resulting from various arrangements of vehicles and operating conditions. Daskin[47] proposes an expected covering model for ambulance station locations in Austin, Texas, and Taylor and Templeton[143] apply a cut-off priority queueing model to determine the number of emergency ambulances required in an urban fleet.

At the A/E department level, many authors have adopted the point of view of a queueing system after the pioneer work of Bailey[14] in the early 1950's concerning outpatient clinics. It is not surprising, therefore that the systems approach to A/E departments has involved building analytic queueing models or queueing simulations. Pike et al[122] determine the number of beds in a Casualty department in Aberdeen using a discrete version of the M/G/∞ queueing model; Bolling[26] analyses the congestion situation in the emergency room of Richmond Memorial hospital, Virginia by using an M(t)/M/b queueing model; and Ladany and Turban[92] use a simulation model to find the optimal numbers of service facilities in an emergency room of a hospital in Tel-Aviv. Also, Weinerman et al[150] suggest the use of a medical triage (sorting for entry) system in order to reduce congestion in A/E departments.

Finally, other studies consider organisational issues in A/E departments. Mackenzie et al[102] suggest a guideline for the organisation of an A/E department associated with a medical
school. The objectives of the study were to identify operational goals of the medical school pertaining to the A/E department setting; to define the operational objectives of the A/E department pertaining to patient care, education and research; and to develop an organisation of personnel, space, equipment and facilities that will allow the operational objectives to be met. Blair[25] defines the objectives of an A/E department in a teaching hospital and makes recommendations about the department's organisation illustrated by organisational charts and flow diagrams. A detailed account of the organisation of A/E services in the United States is given by Sadler et al[132] and by Luft[101].

1.3 The actual problem situation

The North West Surrey District Health Authority (DHA) is part of the South West Thames Regional Health Authority (RHA) and covers the semi-rural area of the Boroughs of Woking and Runnymede and the Walton and Weybridge part of Elmbridge. Its present population of about 209,000 is projected to rise to 215,000 by 1991[117]. The people live in a mixture of semi-rural and urban areas in largely high quality housing with predominance in social classes I and II although other groups are well represented. The main towns include Walton, Weybridge, Woking, Egham and Chertsey. These are, in large measure, dormitory towns for London but there is some light industry in the area. There are good rail links with main lines to London and a branch line which crosses the District.
Road links include the M3 and A3 trunk routes and the newly constructed M25 outer orbital road.

The District General Hospital (DGH) is St. Peter's hospital, Chertsey with supporting hospitals at Woking, Weybridge, Walton-on-Thames and Egham. St. Peter's was established at Chertsey in 1939 as an emergency war hospital in single storey huts. In the 1960s a comprehensive redevelopment plan was prepared and a start was made on the first phases of upgrading the hospital to a DGH. However, due to capital constraints the phased redevelopment programme has been extended for a longer period of time than was initially anticipated and it is still in progress[119]. In 1984 the hospital had about 600 inpatient beds and its inpatient workload was about 19,000 patients.

Hospital A/E services within the District are provided by the A/E department of St. Peter's hospital and by the casualty units of the peripheral hospitals. The former operates round the clock and can offer the whole spectrum of emergency patient care, from the type of attention expected from the general practitioner (GP) to the initial treatment of severe head injuries; the latter provide patient care for minor casualty cases during working days and hours only. The annual number of patient attendances at the A/E department has shown an almost steady increase, since 1975, as is well illustrated in the series of figures of Table 1.1. This increase is disproportionate to the rise in the A/E department catchment population and appears to be in line
with national trends in A/E department usage[82,109,151].

Table 1.1 New and total (including scheduled revisits) attendances at the A/E department of St. Peter's hospital, Chertsey, Surrey.

<table>
<thead>
<tr>
<th>Year</th>
<th>New attendances</th>
<th>Total attendances</th>
</tr>
</thead>
<tbody>
<tr>
<td>1975</td>
<td>22,411</td>
<td>26,106</td>
</tr>
<tr>
<td>1976</td>
<td>24,083</td>
<td>27,645</td>
</tr>
<tr>
<td>1977</td>
<td>25,332</td>
<td>29,770</td>
</tr>
<tr>
<td>1978</td>
<td>27,028</td>
<td>30,918</td>
</tr>
<tr>
<td>1979</td>
<td>28,086</td>
<td>31,685</td>
</tr>
<tr>
<td>1980</td>
<td>28,472</td>
<td>32,430</td>
</tr>
<tr>
<td>1981</td>
<td>29,374</td>
<td>35,905</td>
</tr>
<tr>
<td>1982</td>
<td>30,748</td>
<td>37,398</td>
</tr>
<tr>
<td>1983</td>
<td>29,740</td>
<td>37,112</td>
</tr>
<tr>
<td>1984</td>
<td>33,110</td>
<td>43,242</td>
</tr>
</tbody>
</table>

1.3.1 Physical and human resources of the A/E department

As part of the redevelopment programme of St. Peter's hospital, the building condition of the A/E department has recently been upgraded to B (B=minor deterioration) from condition C/D (C= major improvements required, D= unusable/unsafe) in which it was before. The A/E department consists of two separate neighbouring buildings.
The one serves as a casualty unit, where minor trauma and minor non-trauma cases are seen, and the other (accident centre) is the area where serious cases are dealt with. This area is used for ambulance-arrived cases and for all cases needing resuscitation. The casualty unit operates from 9am to 9pm during working days only; the accident centre is never closed.

The medical facilities in the casualty unit include six treatment rooms, with one bed in each, and an operating room for minor operations. The accident centre has nine beds: six (cubicles) in the main treatment area, one in the anaesthetic room, one in the plaster of paris room and one in the children's room. The major medical equipment is situated in the accident centre and includes one mobile X-ray machine with day-light processor, three ECG machines and six resuscitation trolleys, one in each cubicle.

The medical staff of the A/E department consists of one consultant, two registrars and four senior house officers (SHOs); the nursing staff consists of three full-time and one part-time sisters, eleven full-time and two part-time staff nurses, six full-time and one part-time state enrolled nurses and four student nurses. The clerical staff consists of one secretary and two receptionists.

The consultant is an A/E specialist and has been appointed by the RHA as director of the department, with both administrative and medical responsibilities. The rest of the
medical staff and all of the clerical staff have been appointed by the hospital, on the consultant's recommendations, and they report to the consultant. The nursing staff have been appointed by the District nursing officer and they report to the department's head nurse.

1.3.2 Investigation of existing activities

The view of the A/E department as a human activity system implies that it has a purpose or mission, that is part of purposeful wholes and that it contains people who are by definition purposeful[6,29,38]. According to the consultant, the primary purpose of the A/E department is 'to provide emergency patient care at the highest level of expertise'.

In addition to listing the physical and human resources, the stage of building up a detailed picture involved investigation and understanding of the existing basic activities of the A/E department with respect to its primary purpose. This was accomplished through discussions with the consultant, and other staff in the department, and through our direct observation of the operation of the department. M'Pherson[114] identifies five observational modes for understanding system structures: hierarchical, structural, nesting, functional and dynamic. For purposes of this part of the study we structured the activities of the A/E department into various activity systems distinguishing between the patient care system (consisting of patient reception, diagnosis, treatment and disposal sub-systems) and
various support systems including the organisation and administration system, the patient record system, the audit system and the communications system.

Patient care system

This includes the activities to deal with any type of case presenting in the A/E department, from the least to the most severe one. Walking patients report at the reception desk on arrival and are directed to the waiting room from where they are called for consultation with a physician. When a serious case is due to arrive by ambulance, police or fire service, the department is notified via a two-way radio in order to make the necessary preparations for the reception of the patient. In cases of cardiac arrests, the "crash call" team of the hospital is called into the department to attend the patient. Laboratory tests are performed in the Pathology department whose laboratory operates round the clock for emergencies. Special X-rays (e.g. skull or abdomen) are taken in the Radiology department. After the patient has received all appropriate attention from the A/E department, the decision is taken by the attending physician on the patient's disposal. The latter can take one of the following forms. Simple discharge, advice to the patient to see his/her GP or to call back to the A/E department for review, referral to a specialist outpatient clinic or referral to the appropriate inpatient department of the hospital for admission.
In our discussions, the consultant valued the assurance of 'smooth and logical patient flow after arrival' as the most important aspect of his work. In this context, he deemed it essential that patients were examined by physicians as soon as possible and that the physicians had good diagnostic capability, had practical knowledge of the treatment of a wide variety of cases and made sound decisions concerning patients' disposal. However, he mentioned cases where the department's staff have suffered abuse, and even assault, by patients and their relatives due to prolonged delays. He located the problem in the shortage of inpatient beds and outpatient facilities in the hospital and in the slow feedback to the department of the results of laboratory tests and X-rays. In the former case, a patient was detained in the A/E department until a bed became available in the appropriate inpatient department.

**Organisation and administration system**

The A/E department has the same administrative status as other inpatient departments. The organisation and administration of the A/E department is carried out by the consultant and involves defining A/E department policies and standards related to patient care and planning and organising the space, facilities, equipment and staff to carry out these policies and adhere to the standards.

The department does not have a budget and is directly accountable to the hospital acute services management team.
Within this framework, it must seek approval and coordination from this higher management for acquisition of resources. The consultant envisaged the development of a new and purpose built A/E department at St. Peter's but he recognised that it would be unrealistic to expect this within the next ten years or so, under the existing hospital development plans. However, he thought that there was plenty of scope for improvement of the existing facilities. A matter of primary importance to him was the provision of a number of beds in the A/E department for accommodating short stay (observation) cases. At present, such cases are either admitted to the hospital for fear of complications or they are accommodated in the main treatment area of the accident centre occupying the cubicles which are meant to be used for resuscitation and life support.

**Patient record system**

The existing patient record system was completely manual and was based on patient data recorded on various paper slips (according to case) while some data were duplicated in books. Most of the physicians were frustrated by the difficulties of clinical information retrieval from the case notes. In the consultant's view, this system was entirely inadequate for the operation of the department. He envisaged a more flexible system that would reduce paper work, while improving the quality, legibility and consistency of medical records and that would also provide the basis for clinical research.
Audit system

This was concerned with the daily review of the patient case notes but emphasis was placed on auditing the X-ray plates taken during the previous day. It was carried out by the consultant, in conjunction with other medical staff, in order to identify any missed fractures or other missed abnormalities. Corrective action was taken by calling the patient back to the department for review, when necessary. The X-ray audit was also used for the training of junior physicians on X-ray plate reading. The consultant perceived the role of the medical audit in the A/E department as: 'a process which attempts to scrutinise and give credence to the medical care that we offer to our patients and which aims to evaluate our performance according to criteria established by the medical profession'.

Communications system

The A/E department, through the consultant, liaises with the social workers for social cases presenting themselves in the department; with the GPs and the peripheral casualty services to discuss local problems and remedy any deficiencies; and with the local authorities (e.g. health visitors, health councils) to make the general public aware of what the A/E department is intended, how they will benefit and how they can help to develop the service by using it sensibly[82,151]. The consultant acts also as a DHA's advisor on matters related to A/E services either directly or through his participation in the A/E services sub-committee.
Within the hospital, the main communication links of the A/E department are with inhospital departments and outpatient clinics for discussing problems related to patient flow through the hospital and with the hospital administration and other hospital bodies for acquisition of resources. In relation to the acquisition of medical equipment, the consultant mentioned that he would first obtain agreement from other medical specialist groups in the hospital (e.g. division of orthopaedics or medical equipment sub-committee), and then submit a request to the acute services management team rather, than seek direct approval from the latter and have a higher risk of rejection. As an example, he mentioned the recently replaced operating table of the A/E department at a cost of £3,500. It is worth noting at this point that the lack of hierarchy of authority within hospitals has in the past been criticised by Curnow[44] and by Rosenhead[128].

At the same time as this detailed picture was being developed, we studied various official reports on A/E services and on the organisational structure of the NHS and definition of management roles within this structure[4,48,51,52,79,82,108,133,139]. Our primary aim was to identify the environment of the A/E department and the interactions which significantly affect its nature. Also, identified were those parts of the environment with which these interactions occur. Finally, a number of visits were made to other hospitals so that the type of service offered by the A/E department of St. Peter's hospital could be
compared with that elsewhere.

1.4 Systems analysis

From the detailed picture stage several potential problem areas were illuminated. For example, the installation of an adequate patient record system (probably computerised in accordance with the current trend), the patient flow through the department and the resource allocation could all be seen as areas offering the opportunity for further investigation. However, such a piecemeal approach would have excluded consideration of many interacting problems and so could only provide partial and even misleading solutions. The conceptualisation stage of systems analysis attempts to assess the strengths and weaknesses of the existing procedures. Thus, we decided to proceed by first developing conceptual models of the problem situation and then defining areas for further investigation by successive elimination of those parts of the conceptual model to which changes would be either infeasible or undesirable. Also, we believed that such an approach would lead to increased insight into the operation of the A/E department and would provide an elegant and rational way of formulating potential problems.

Conceptual model building constitutes an important element in Churchman's systems approach[39], Mitroff's et al[113] view of the scientific approach to problem-solving and in Checkland's systems approach[38], among others. Also,
Tilquin[145] incorporates conceptualisation into a systems analysis study of both personal and collective health service systems.

1.4.1 Conceptual model building

Having familiarised ourselves with the concept 'A/E department' and with the problem situation being studied, our next step was to select viewpoints of the A/E department and build conceptual models of systems, implicit in these viewpoints, which would hopefully throw light on the problem situation. According to Checkland[38], these viewpoints can be formulated into root definitions which are accounts of 'what the system is'; for each root definition a conceptual model can then be developed which reflects 'what the system must do' in order to be the system named in the corresponding root definition.

After several iterations, during which we examined the implications of various root definitions on the problem situation, the decision was taken to develop a conceptual model of an A/E department based on the view that it is:

A hospital owned system to provide maximum care for all types of accidents and emergencies in the District and to provide continuously effective and efficient emergency patient care at a level of performance acceptable to the community at large but within the constraints imposed by statutory, emergency medicine and hospital limitations.
A number of comments are worth making about this statement:

(1) The statement of ownership in this definition implies that activities must be included in the model which identify, and act on, the policy, constraints and expectations of that owner. The system is owned by the hospital in the sense that the hospital is its immediate superior and responsible for the implementation of DHA, RHA and NHS policies[48,118].

(2) The statement 'most able...District' imply that the system must be properly built, equipped and staffed if it is to fulfil its basic purpose and that it must advise as required outside authorities on matters related to emergency patient care.

(3) The words 'continuously effective and efficient' indicate a need for the system to continuously review and update its knowledge on all aspects of emergency medicine, including building structure where the service is provided and current developments in methods of care, drugs and equipment. The acquisition and use of resources need to be cost effective, in accordance with national health policies[51,56], while efficient use of the system's resources and medical knowledge implies good service to patients.

(4) The performance of the system must be acceptable to the community at large, for which all the medical facilities are deployed[49,52] and whose members are the system's utilisers.
This means that the first priority of the system is how best to serve its patients and determines the world view implicit throughout this study.

(5) The environmental constraints imposed on the system are determined by the state of emergency medicine, by the inhospital facilities with which the system interacts, by its statutory purpose and by the hospital proposed priorities on development needs. Obviously, all of these constraints apply to changing environments.

Thus, at first level of resolution the above definition leads to the four major sub-systems illustrated by Figure 1.1. It must be emphasised that the model of Figure 1.1 is a systems model and is a statement of those sub-systems which must exist in a system described in the root definition. It is not implied that these sub-systems should necessarily exist as organisational entities.

The model of Figure 1.1 is at too broad a level to be of use in the derivation of the resource implications of the type of service provided in the A/E department and to bring about areas for further investigation, but it serves as an intermediate stage in the hierarchical model development process. Following Checkland[38], each of the four sub-systems were redefined as systems and again, through the mechanism of root definitions, models were produced at the next resolution level. These four models were then put together to produce a complete systems model of the A/E
department at this higher level of detail. This is shown schematically in Figure 1.2.

Consider first the resource development sub-system which will be necessary for the system to ensure that resources are available in order to be 'most able to care for all types of accidents and emergencies in the District'. This sub-system will have to recognise the hospital capital and revenue constraints; it will have to appreciate the system's existing resources and their capabilities; it will have to know about modern standards of A/E departments; it will have to appreciate the system's statutory role including how much initial care it can provide; and it will have to assess development needs, establish an order of priority for those needs and feed these priorities into the hospital planning mechanism.

The resource organisation sub-system responds to the resource input and combines this with up-to-date knowledge of emergency medicine to plan the task of providing emergency patient care in conformity with community and hospital expectations. These resources are then organised so that they are capable of carrying out this task efficiently.

Considering the patient care sub-system, this will have available the organised resources and it will have to make efficient and cost effective use of the system's resources and knowledge in executing the task of providing emergency patient care. The main activities of this system will be to
receive, diagnose, treat and dispose patients.

Finally, in order that all these sub-systems operate together in such a way that the system pursues its purpose in a way which meets the expectations of the hospital and which is acceptable by the community, it will be necessary for the system to monitor the total set of its activities and take control action where appropriate. This will require measures of performance for the system to be defined which are appropriate to expectations. There are obviously a number of possibilities, such as cost (total cost, number and mix of services rendered, cost per unit of service, long-term cost trends), quality of care (structure of setting where emergency patient care is provided, process of care, outcome of care), patient satisfaction (patient's evaluation of courtesy, ease of access to care), staff satisfaction (salaries, environment, flexibility in the practice of medicine, communications) and medical research and education. At this level of resolution we only record the activity of setting measures and monitoring with respect to them, rather than including any specific measures. We shall have more to say about the monitor and control function in the next chapter.

1.4.2 Definition of problem areas

The next step in the study plan was to compare the activities contained in the conceptual model with those occurring in the existing A/E department. Since the conceptual model
developed was based on our perception of the situation, however, the strategy adopted in the comparison stage was to use that model as a basis for debate with the consultant on the existing activities of the department. This procedure was facilitated by the fact that the main activities in the A/E department were the same as those shown in the conceptual model.

The debate was about all areas in which it was fruitful to discuss differences (i.e. on areas that could be influenced by the consultant, excluding medical aspects of the A/E department's work). Among these differences two notable ones are worth mentioning here. First, requests for resources were made while the consultant was not informed about the overall hospital development plans and without documented justification for their need. As a result, the consultant experienced sometimes frustrating rejections of his requests. For example, the development of the observation ward (short stay beds) in the department had been discussed in a meeting and considered favourably, but the matter was referred to another meeting at a later date because there was not enough documented justification for its need and size. The consultant's perception of this situation was that the administration wanted to retain the 'status quo'. Our perception was that the hospital preferred to allocate its limited resources to departments where there was a strong case for a particular need and that, in any case, any allocation should be within the existing plans. Second, a monitor and control activity in the A/E department was
virtually non-existent. The consultant perceived the role of monitor and control as identical to medical audit, which deals with the medical aspect of the work only. In our opinion, a monitor and control function in an A/E department should have a much broader scope including both day-to-day (medical audit) and longer term (administrative) activities. Partly due to the cumbersome way of recording and retrieving patient-related data, medical audit was rudimentary and administrative monitoring was non-existent. It must be noted here that the importance of the medical audit function in an A/E department has been emphasised in a number of recent articles[10,99,121]. Also, the role of the consultant in the NHS management has been stressed by Chawner[37].

Overall, these discussions proved fruitful in that they developed our understanding, with the consultant, of areas in the A/E department where particular problems were being encountered. Also, it was realised during the course of these discussions that, for the consultant, his responsibility for running the A/E department (and therefore defining and alleviating problems) was something which he was not prepared to concede to outsiders, but he was willing to consider problems which were convincingly obvious. Further, this realisation had to be taken into account if we were to expect his collaboration during the course of the project.

As a result of these discussions we set out, with the consultant, the following framework for this research. First, to design a patient record system for the A/E
department; second, to provide information on the department's workload and, in particular, with respect to patient flow through the A/E department. In relation to the patient flow, it was agreed that areas of enquiry would include patient waiting and treatment times in the A/E department, the use and size of the prospective observation ward and the interaction between the A/E department and inpatient departments through emergency patient admissions.

1.5 Conclusions

This chapter was concerned with the problem identification phase of our study on the A/E department of St. Peter's hospital. In summary, the use of systems concepts resulted in three main benefits.

First, the structuring of the A/E department activities into mutually dependent, notional sub-systems provided the consultant with a realisation of what he does and, potentially, of what he might do to ensure that his department is able, with respect to the resources available, to provide effective and efficient emergency patient care at the highest level of expertise.

Second, the comparison of the conceptual model with the actual situation revealed that a major shortage was the lack of an adequate patient record system in the A/E department which would not only reduce paper work but would also be
capable of providing information to the consultant for monitor and control purposes in the A/E department. Thus, the need for an adequate patient record system was established and its design will be the subject of Chapter 2.

Third, our systems thinking dictated the route of the work which is described in subsequent chapters in the form of individual case studies. The purpose of the work was to improve those parts of the department's operation where particular difficulties were being experienced. In each case study, emphasis is placed on providing solutions which have practical merit while serving research purposes.
Figure 1.1 First resolution level model of an A/E department.
Figure 1.2  A conceptual model of an A/E department (constraints and key as in Figure 1.1).
CHAPTER 2

A/E DEPARTMENT PATIENT RECORD SYSTEM DESIGN

2.1 Introduction

Design is the process of transforming a system from 'what it is' to 'what it should be'. The design problem can thus be formulated as a difference between an existing system state and a desired system state, subject to constraints imposed by the inner and outer environment of the system. Hence, systems design must be preceded by a clear definition of the objectives and the environment of the system to be designed. The environmental and internal constraints on the system will determine the conditions under which the system must operate while the objectives of the system are expressed in terms of a desired relation between the system and its environment and relations within the system itself.

In the preceding chapter, a basic structure of the A/E department to carry out its activities was described and the need for relevant and timely information to support these activities was established. This chapter is concerned with the problems of designing a patient record system to provide this information and of implementing this system in the A/E department of St. Peter's hospital.
Patient record system design proceeded by the following steps. First, we defined the objectives of the system and its environmental constraints; second, we defined the nature of information requirements to meet these objectives and focused our attention on the patient-related information; third, we designed a suitable patient card for the A/E department to enable collection of standardised patient data; and fourth we set up a computerised data base system to process the data and generate the required information. The designed system was then implemented on an experimental basis in the A/E department.

Before we discuss the process of the patient record system design, we present a review of the literature on computerised patient record systems for A/E departments.

2.2 Review of the literature

To a large extent the health care process consists of gathering, interpretation, selection and use of pertinent information. As information's importance in the health care area continues to grow hospitals have increasingly been using computer systems in support of their management process. Particular attention has been devoted to developing computerised systems for A/E departments.

Computerised patient record systems seem to have been installed in some A/E departments since the early 1960s.
Levy et al[95] report on a computerised patient record system implemented in a diabetes ambulatory clinic of a hospital in order to improve the quality, legibility and consistency of patient records. O' Connell and McFarlane[120] develop a computerised patient record system with the objective of providing information for effective planning of ambulatory services. Grossman et al[68] develop a computerised medical record system capable of meeting, at a reasonable cost, the data needs for patient care, administration, planning and evaluation. McDonald et al[104] design a model for a clinical information system to serve an ambulatory care facility. Several other studies on computer applications used in ambulatory health care delivery, in the United States, are described by Giebink and Hurst[63].

Up to 1981 in the UK, computerised patient record systems had been in use in at least six A/E departments[3]. Among them are the A/E department of the Royal Victoria hospital, Belfast reported by Rutherford and Maynard[130]; and the A/E department of Royal Lancaster infirmary, Lancaster reported by Dalby et al[45] and by Roberts et al[128]. Major advances have been made in the development of computerised patient record systems, since 1981, using microcomputers. In 1981, a microcomputer system was installed at the A/E department of Bangor general hospital, Gwynedd for real time registration and printing of the casualty card of patients attending that department[105]. An improved microcomputer-based patient record system was the product of the CAER (computer-based A/E department records) project[3] conducted under the aegis of
the steering group on health services information (Korner group) and of the computer policy committee of the NHS. The objectives of the system were to allow real-time data entry and retrieval so that paper work is reduced, to facilitate fast access to patient records and to provide standard A/E data for subsequent analysis and interhospital comparison. Interest in such data has been expressed recently by national accident prevention and health education organisations such as the Home Accident Surveillance Scheme and the Child Accident Prevention Trust[3]. The prototypes of the CAER project system have been implemented at the A/E departments of Leeds, Salford and Bangor general hospitals. A third system, which is a combination and expansion of the previous two, has been installed at the A/E department of Derbyshire Royal infirmary[125].

2.3 Definition of objectives

One of the characteristics of any system is that it must be designed in such a way that is capable of achieving its overall objective subject to the constraints imposed by the wider systems of which it forms a part[81]. In our discussion it was agreed with the consultant that the objective of the patient record system was 'to contribute to improved patient care by providing relevant and timely information for increased clinical and administrative control'. Thus, in agreement with Ackoff[5], the design of the patient record system was seen in the framework of
control activities. At this point we note the distinction between a patient record and the traditional medical record. The former provides data which can be transformed into information for both administrative and clinical purposes while the latter is essentially an aide memoire for the physician. A medical record contains much more clinical detail than is necessary for information reporting and at the same time it may be missing certain data items which can give rise to useful information for administrative purposes[115].

The intent of this study was not to restructure the existing medical record, but to encourage the incorporation of certain basic data in the record so that they are available for reporting and analysis as needed by the A/E department or outside authorities.

2.4 Information requirements

In the previous chapter, the monitor and control function was identified as an important component of the organisational structure of the A/E department while the activities of monitoring and control with respect to measures of performance were stated in broad terms only. It is now appropriate to define explicitly these measures of performance and the control actions.

From the consultant's point of view, the primary criterion of performance of his department was the quality of patient care rendered. The environmental constraints on the patient
record system would obviously be the ones external to the total A/E department and those determined by the inner environment, i.e. those imposed by the A/E department itself as the immediate environment of the patient record system. Since the quality of care is not directly measurable, it is common practice to monitor it in terms of proxy measures. At the A/E department level these include[101,132]: structure of setting (qualifications of staff providing the service, keeping and structure of medical records, premises, equipment); ease of access to care (patient reception, waiting times); process of care (dependent on the purpose of patient visit — examinations, diagnoses, treatment procedures, disposal); outcome of care (morbidity, mortality); and patients' acceptance of the services provided (patient's evaluation of courtesy and inconvenience, confidentiality of patient records etc).

Monitoring the performance of the A/E department with respect to these measures will obviously require both formal (patient-related) and what may be called informal (communications among staff, plans and policies of wider systems) information. These information requirements were assembled from the previous analysis. In particular, for the patient-related information we used the idea suggested by Bowen[29] namely 'to regard the patient as an essential body of information flowing through the system'. It is noted, however, that since the conceptual model represents minimum necessary activities, the information categories collected were the patient-related minimally necessary ones to support
these activities. Thus, these are the information categories that should be provided by the patient record system. In other words, the patient record system was seen as the formal information provider in the A/E department, while informal information flows are, of course, retained, as these represent significant features of both the social and the administrative structure of the system.

With respect to the above measures of performance, the consultant could take control actions inside the department (e.g. arrangement of appropriate physician rosters, modification of medical practice, training of junior physicians); recommend action to the hospital (e.g. requests for resources, reporting on prolonged delays of patient flow through the department due to interaction with other in-hospital facilities); and he could take indirect input control action (e.g. encouraging GPs to use the A/E department for real emergencies only, providing community health councils with information about frequent causes of accidents to aid prevention plans).

Up to this stage we had specified the information requirements for monitoring and control in the A/E department. Our next step was to design a patient card which would have the potential for evaluating the quality of patient care rendered, that is, contain the necessary data items whose processing would provide the specified information and which would be logically structured. These data items were the ones which would be standardised and
codified. Here, two additional points were considered. First, that the data should be recorded in a way which ensured compatibility with similar patient records kept in other A/E departments in England[3,125] for inter-hospital comparison. Second, the minimum data-set recommended by the steering group on health services information (Korner group) should also be included since collection of these data will become compulsory in 1987 for all hospitals in England[3,47,53,555]. Therefore, the system adopted for the standardisation and codification of the data items was similar to the one recommended by the Korner group, with minor deviations added by the consultant on the basis of local characteristics.

The codified data are divided into five categories corresponding to the states of the patient flow to and through the A/E department. These are: pre-hospital (place of incident, type of incident, source of referral, mode of transport); registration (patients' identification and other patient characteristics, date and time of arrival); diagnosis (examining physician, time of examination, purpose of visit, types of examination, final diagnoses, anatomical site of illness); treatment (type of operation, other treatment procedures); and disposal (time of departure, mode of disposal, code of inpatient department for patients admitted). In addition to the codified data items other physician notes would be recorded in the traditional narrative way. The designed card was then produced in a form of multiple carbon copies and has since been implemented at
the A/E department of St. Peter's hospital.

Once the information requirements had been specified and once the data items had been defined, consideration was given to the reporting structure and the frequency of reporting. It was realised that the information should be interpreted and displayed effectively before it reached the consultant who would not be expected to have the time or analytic skills to cope with complex statistical reports or the patience to deal with multi-column computer printouts. Therefore, parsimony and simplicity of the reports were judged essential if the system was to play a significant part in improving the quality of patient care rendered in the department. With respect to frequency of reporting, it was realised that certain information (e.g. related to staff scheduling or usage of resources) would be required on a periodic basis while other (e.g. related to audit) should be readily available on a daily basis.

2.5 Computerised patient record system

At the same time as the early stages of the design process were being carried out, we considered with the consultant the question of the means by which the information would be provided, be it manual or computerised. We assumed that computer support could be provided, as discussed below, and we examined those two options in the light of the objective of the patient record system defined earlier; that is, we
ignored for the moment any financial constraint which might rule out any consideration of the second option.

Although the previous system of recording patient data—in various case notes and books—was quite cumbersome to use for any monitoring activity, the agreed one—in which all patient data are recorded on a compact patient card—could be useful for short term activities such as daily audit. The department's workload of about one hundred patients a day would not induce any particular problems. The real problem lay in generating information for support of the longer term activities such as those related to resource acquisition and usage, staff scheduling and to providing information to health authorities and to community interest groups.

These considerations led us towards the computer option on the provisos that the installation of the computerised patient record system would be regarded as an experimental situation for the trial of new ideas and that the existing routines of the department would be disturbed a little as possible. The 'new ideas' referred to techniques of management and control, rather than to the designed system itself. It was then agreed that the Royal Holloway College (RHC) would provide computer support to the A/E department on its VAX 11/780, VMS computer for an experimental period of one year. The user-interface would be carried out from the hospital site through a TeleVideo model 910 terminal, lent by the department of Mathematics for that period. This was linked to the computer via a British Telecom modem. Thus,
the A/E department would be able to operate the system in interactive mode round the clock. Since the data would be stored on the VAX's disk at the RHC site, an encryptor was embedded in the terminal to allow on-line encoding of entered data and decoding of retrieved data. In this way, the confidentiality of patient records was safeguarded.

Our next step was to set up the computerised data base system. This involved setting up the data base, where the data would be stored, and developing the computer programs for input/output processing.

The data base consists of a single file whose records contain all the codified data on the patient card. The structure (organisation) of the file is indexed which allows a combination of sequential and keyed access to the file. This combination is often called Indexed Sequential Access Method (ISAM)[33]. Each patient record has a primary key, which identifies the record uniquely, and one or more secondary keys, which may or may not be unique. Primary Key is the 'A/E number' and secondary keys are 'date of arrival', 'date of birth', 'physician code', and 'group of patient'. All of these keys are embedded in the patient record. If the data base is accessed using a particular key, the records are automatically sorted in ascending order of the key's value by means of the Index table maintained by the Record Management Service (RMS) system of the VAX. Keyed or sequential access to the records can then be performed. When records are added to the data base, the Index table is updated so that the new
Having set up the data base, a computer program was written for data entry, maintenance, retrieval and information reporting. The program was designed in the form of a sequence of modules and was written for portability in standard ANSI FORTRAN 77. The access procedure to each module is menu-driven and there is a set of prompting messages which enable the user to become proficient at operating the system within a few days. The user selects a particular module and, by answering a set of questions displayed on the terminal screen, he/she specifies the required data analysis or item of information. The available modules are:

STORE. This module is accessed for data entry in the format appropriate to the data type. Any typing errors during the data entry can be corrected immediately by typing the character 'E', for error, thereby positioning the terminal cursor in the previous location. The data are automatically checked for logical errors relative to individual data-items (e.g. time of arrival greater than 2400 hrs). After the data have been typed, the user has the option of accessing the CHECK sub-module to detect record-wise logical errors (e.g. males admitted to gynaecology department). If the data are correct they are stored in the data-base. If errors are found, the system responds with a message listing the locations and nature of the errors. These are then corrected by typing 'E' to position the cursor at the
appropriate location and by retyping the relative entry.

RETRIEVE. This module is accessed for data retrieval and for simple aggregations of data. If the user requires aggregated data (e.g. total number of X-rays performed in a particular period of time, average per day, standard deviation, percentages) he/she types /STATS after RETRIEVE in order to enter the relevant sub-module.

REPORT. If aggregations of data are required in the form of two-way cross-tabulations, the REPORT module is accessed. The user indicates the specific cross-tabulation required interactively. There is also the option of accessing the sub-module HIST for histogram plots.

SEARCH. This module allows entering of simple programming instructions for complex retrieval studies to be performed on the data base for research purposes (e.g. hourly counts of patient arrivals).

UPDATE. This module is accessed for updating patient records, in the light of new data not available at the time of the record's initial entry. To guard against updatings of wrong records the system responds by printing on the terminal screen the A/E number of the record and asks for confirmation of the updating request. A 'GO' response from the user results in updating the record.

DELETE. It initiates the deletion of a record. To guard
against mistaken deletions the same procedure as for UPDATE is performed. Here, a 'GO' response from the user results in final deletion of the record and automatic updating of the Index table.

HELP. At any time at which the system is awaiting a reply from the user, the input of the character '?' gives access to the HELP module designed to assist the user on how to use the system.

2.6 Implementation

A continuous theme of this study has been the opportunities for improving the quality of patient care through effective clinical and administrative control supported by the computerised patient record system. There is, of course, another side of this: the management of change. It was anticipated that implementation of the designed system would form the basis for instituting changes in the operation of the department in particular with respect to monitor and control. Therefore, the implementation of the system had to be carefully planned if the system was to fulfil its purpose.

It was agreed with the consultant that the operation of the total (manual and computerised) system would be as follows. The patient cards would be filled in by the receptionist on duty (pre-hospital and registration data items) and by the examining physician (diagnoses, treatment procedures and
disposal data items). Each morning the casualty cards of the previous day would be checked by the clerical assistant, attached to the project, for errors and completeness. Then, the codified data items on the corrected patient cards would be entered in the data base by the clerical assistant who had been adequately trained on both computer interface and on the potential and use of the system. She would also be responsible for the routine information retrieval.

The problems encountered at the early stages of implementation were related either to the lack of previous experience of the department's staff with similar systems (e.g. incomplete cards, mistakes), or to the dependency of the system on the RHC computer (e.g. machine breakdowns, slow response during term-time). It was, therefore, decided to allow a few months for the staff to get used to the new system, while the total system would be in operation, and start the formal data collection and retrieval process on the 1st January 1984.

2.7 Conclusions

There has been a growing movement recently to improve the information flows within the NHS and better patient record systems have been regarded as an integral part of this effort[53,55]. At the A/E department level, a major impetus towards this direction has been given by the CAER project[3]. The objective of that project, as stated by its steering
committee, was 'to develop a transferable on-line microcomputer-based A/E information and administration system'. The system would allow the user to collect and produce various cross-tabulations of the data set recommended by the Korner group\[53,55\], for management at a District level, and of a locally determined optional data set. However, no definition has been given of the potential benefit of such a system to the A/E department itself which, as Black\[24\] states, '...must of course be interested in those facilities for which it is responsible'.

Throughout this study a conscious effort was made to provide a framework within which a patient record system could be created which would be closely related to the achievement of A/E department objectives while serving the information requirements of the wider systems.

In assessing the benefits accrued from the operation of the computerised system, we might conclude that if the system is to be used primarily for monitoring daily activities then the means of a computer might not be cost-effective for a medium size A/E department like the one at St. Peter's hospital (although this conclusion is based on limited experience rather than on a complete feasibility study). On the other hand, the large volume of data required in order to generate information for longer term activities, including research, and for collecting the Korner data set for District purposes would necessitate the use of computer facilities. Such facilities would offer the opportunity for improved control
and management in A/E departments[82,109,151]. In the subsequent chapters we focus our attention on tackling problems with respect to the longer term administrative activities of the A/E department.

1.1 Introduction

The general approach described a qualitative and conceptual overview of the A/E department. By means of a prescribed system of study, a good insight was gained into the nature and the behaviour which led to the design of a system of education which was closely related to the environmental of the A/E department objectives.

The use of data and information in a more quantitative and more scientific manner could be achieved by a more precise description of the characteristics of the work of the A/E department and a description of the dynamic behaviour of central problems in administrative evaluation. Such a characteristic comprises the study of size and nature of patient demand for emergency care and dynamic behaviour concerning the study of the sequence and delay that takes place in the patient care process. This chapter attempts to describe in more detail the nature of patient demand for emergency care and the consequences of the patient care process in the A/E department.

The emphasis in this chapter is the need for a more comprehensive qualitative systems investigation of existing
3.1 Introduction

The preceding chapters described a qualitative and conceptual systems approach to the A/E department. By means of a structured symbolic model, a good insight was gained into the system and its problems, which led to the design of a computerised patient record system in a way which was closely related to the achievement of A/E department objectives. Patient-related data and demonstrable information could be obtained for performance evaluation.

Characterisation of the workload of the A/E department and description of its dynamic behaviour are central problems in performance evaluation. Workload characterisation comprises the study of size and nature of patient demand for emergency care and dynamic behaviour concerns the study of the interactions and delays that take place in the patient care process. This chapter attempts to describe in some detail the pattern of patient demand for emergency care and the timeliness of the patient care process in the A/E department.

The studies described in this chapter formed parts of a more comprehensive quantitative systems investigation of existing
patterns in emergency care delivery in the A/E department. As a result of this investigation further insight was obtained about the workload and operation of the A/E department and a more sound basis for the case studies described in the next chapters was provided. The specific results reported here refer to the complete sample of unscheduled patient visits to the A/E department during a period of 13 weeks, from 2 July to 30 September 1984. The data are analysed using the statistical methods of harmonic analysis and of categorical data analysis while emphasis is placed on providing information for monitor and control purposes at the A/E department level. Therefore, the pattern of emergency demand is described with respect to the hour of the week and the timeliness of the patient care process is investigated with regard to physician shifts and patient waiting and treatment times.

3.2 Review of the literature

Given the important role of the A/E department in community medical care, several studies attempt to describe in quantitative terms the caseload and the working of such departments. The objective of most of these studies is to provide understanding of the present with the prospect to plan for the future.

King and Sox[87] describe in some detail the patient population, types of emergency cases and diagnoses, timing
and locations of demand, transport modes and nature of
treatment as a first step towards a more comprehensive systems analysis study whose objective was to improve the timeliness and adequacy of emergency medical services in San Francisco. Reed and Reader[126] present data pertaining to the types of patients, their utilisation of hospital facilities and the nature of their complaints from a sample of patients who used the New York hospital emergency room during the year 1965; Vaughan and Gamester[147] examine retrospectively the reasons that make patients use the A/E departments of the hospitals in Michigan. Goss et al[66] explore the range of patients' experience on different shifts and days of the week with respect to total visit time and waiting time to see a physician in the A/E department of the New York Hospital and Torrens and Yevvab[146] demonstrate the variation among emergency room populations by carrying out a comparative study in four hospitals in New York City.

Several studies of similar nature have also been reported for UK situations. Brownie[35] analyses the workload of the Casualty Department of Glasgow Royal Infirmary; Rutherford and Maynard[130] investigate the load of referrals from work at the Casualty department of the Royal Victoria hospital; and Roberts et al[128] demonstrate the use of a computer system in the analysis of the attendance profile of the Casualty department of Royal Lancaster Infirmary. Also, one unpublished study is worth mentioning. It was carried out by the Wessex Regional Hospital Board in 1973 and resulted in a comprehensive report on the A/E services in that Region[151].
Special emphasis was placed on the analysis of major and minor case load and the development and utilisation of resources. A/E department patient record systems were also investigated and recommendations were made on their improvement with respect to compatibility and methods of storage and retrieval.

All of the aforementioned studies are purely descriptive in the sense that the reported data analyses are restricted to counts and percentages. Such analyses certainly develop understanding. However, if something more than a macroscopic view is required the analyses are best carried out using mathematical tools such as those employed in this chapter.

3.3 Harmonic analysis

The first type of analysis described in this chapter concerns the investigation of the pattern of patient demand for A/E department care with respect to the hour of the week. This was carried out by applying the deterministic harmonic model to the time series of hourly counts of unscheduled patient attendances to the A/E department considering as null hypothesis that a purely random process of counts is homogeneous Poisson. A detailed theoretical background for the use of harmonic analysis is provided by Anderson[9]. The specific problem of describing the pattern of counts of events has been dealt with by Pocock[123] in a study on
industrial absenteeism.

Two comments are worth making about the choice of statistical methodology. First, the assumption of the Poisson process of counts is justifiable since the arrival process consists of a superposition of arrivals from a large number of sources (e.g. patients with different complaints)[43]. Second, although the exact times of arrivals were available, we used counts of these arrivals in successive hourly intervals since it was the hourly variation in the pattern of patient demand that was of interest. The grouping of arrivals in successive equispaced time intervals to form the corresponding counts of arrivals is a satisfactory basis for testing since if the data were from a Poisson process these counts would be independent variates with possibly different means[43]. Later on in this thesis we use statistical methods based on the actual times of patient arrivals[97], in order to investigate time of day and time of week effects.

3.3.1 Fourier representation

Consider the number of unscheduled patient attendances at the A/E department over a period of N weeks and divide these N weeks into k hourly intervals. Then, \( k = N \times 168 \). Define \( X_t \) as the number of patient arrivals during the \( t \)-th hour for \( t = 1, 2, 3, \ldots, k \). One basic result from the theory of Fourier analysis is that there exist constants \( \{A_p\} \) and \( \{B_p\} \) such that
The model (3.1) is called the Fourier series representation of \( \{X_t\} \) and the constants \( \{A_p\} \) and \( \{B_p\} \) are called Fourier coefficients. These are given by the expressions:

\[
A_o = \frac{1}{k} \sum_{t=1}^{k} X_t = \bar{X}
\]

\[
A_p = \frac{2}{k} \sum_{t=1}^{k} X_t \cos \frac{2\pi pt}{k}
\]

\[
B_p = \frac{2}{k} \sum_{t=1}^{k} X_t \sin \frac{2\pi pt}{k}
\]  

In model (3.1) it is seen that the Fourier series has \( k \) parameters to describe \( k \) observations and so can be made to fit the data exactly. This explains why there is no error term in (3.1).

The overall effect of the Fourier analysis of the data is to partition the variability of the time series of counts into components at frequencies \( 2\pi/k, 4\pi/k, \ldots, \pi \). The component at frequency \( \omega_p = 2\pi p/k \) is commonly referred to as the \( p \)-th harmonic. Its period is \( N/p \) weeks and its frequency per week is \( p/N \) cycles.

Those harmonics with \( p \) equal to an exact multiple of \( N \), the number of weeks, have their frequencies per week equal to an integer. This means that these harmonics have a cycle which
is repeated an exact number of times per week and therefore follow the same pattern in all weeks. For example, for \( N=13 \) weeks, harmonics with \( p=13, 26, 39, \) etc have periods 1, 1/2, 1/3, etc weeks and frequencies per week 1, 2, 3, etc cycles. Furthermore, for any linear combination of such harmonics the same pattern is followed in all weeks. Thus, any such linear combination has a cycle of exactly one week.

Consider the full linear combination of all harmonics with frequency per week equal to an integer. That is, define

\[
W_t = \sum_{p=1}^{k} \left( A_p \cos \frac{2\pi pt}{k} + B_p \sin \frac{2\pi pt}{k} \right)
\]

Then, \( W = \{ w_t \}_{t=1}^{k} \) is defined as the hour-of-week effect on \( \{ X_t \}_{t=1}^{k} \). This effect has a cycle of one week. The remaining harmonics can be combined to form what might be called 'irregular' effect on the \( \{ X_t \}_{t=1}^{k} \). This effect includes evolutionary trends and other week-to-week sources of variation.

The total corrected sum of squared deviations is defined as

\[
\sum_{t=1}^{k} (X_t - \bar{X})^2 = \sum_{t=1}^{k} \left( \sum_{p=1}^{k/2} \left( A_p \cos \frac{2\pi pt}{k} + B_p \sin \frac{2\pi pt}{k} \right) \right)^2
\]

which, because of the orthogonal properties of the harmonic components, reduces to

\[
\sum_{t=1}^{k} (X_t - \bar{X})^2 = \sum_{p=1}^{k/2} \frac{k(\lambda_p^2 + B_p^2)}{2}
\]
Dividing through by $k-1$ we have that the sample variance of the observations is given by

$$\sum_{t=1}^{k} (X_t - \bar{X})^2 / (k-1) = \sum_{p=1}^{k/2} \frac{k(A_p^2 + B_p^2)}{2(k-1)}.$$  

(3.6)

Thus, $k(A_p^2 + B_p^2)/2(k-1)$ is the contribution of the $p$-th harmonic to the total variance of the observations and (3.6) shows how the total variance is partitioned.

3.3.2 Testing the Poisson hypothesis

Consider the null hypothesis, $H_0$, that patient arrivals occur randomly in time according to a Poisson process with hourly mean $\lambda$, say. The alternative hypothesis, $H_1$, is that $\{X_t\}$ is a non-homogeneous Poisson process. The index of dispersion of $\{X_t\}$ is defined as (43)

$$I = \sum_{t=1}^{k} (X_t - \bar{X})^2 / \bar{X}(k-1).$$  

(3.7)

Under $H_0$, $I(k-1)$ has approximately a $\chi^2$ distribution on $k-1$ degrees of freedom. Therefore, it is possible to test $H_0$ using $I$. Furthermore, under $H_0$, it can be shown (9) that

$$E(A_p) = E(B_p) = Cov(A_p, B_p) = 0$$

$$Var(A_p) = Var(B_p) = 2\lambda / k.$$  

(3.8)

Thus, $E\left[\text{sum of squares for } p\text{-th harmonic}\right] = 2\lambda/\text{for } p=1, 2, \ldots, k/2-1$ while for $p=k/2$, $E\left[\text{sum of squares}\right] = \lambda$. Also, $\{A_p\}$ and $\{B_p\}$ can be assumed to have a normal distribution.
provided that $k$ is reasonably large and that the overall sample mean is greater than one[43], so that

$$\frac{k(A_p^2 + B_p^2)}{2\lambda} \sim \chi^2_k. \quad (3.9)$$

Replacing $\lambda$ by the sample mean $\bar{X}$ we have

$$\frac{k(A_p^2 + B_p^2)}{2\bar{X}} \sim \chi^2_k \quad (3.10)$$

except for $p = k/2$ when $kA_p^2 / 2\bar{X} \sim \chi^2_1$. Thus, we can test the significance of each harmonic component by using the $\chi^2$ distribution.

Since the sums of squares are additive, the sum of squares due to hour-of-week can be tested for deviations from $H_0$ using the fact that under $H_0$

$$\sum_{p=1}^{k/2} k(A_p^2 + B_p^2) \sim \chi^2_M \quad (3.11)$$

where $M = 167$, the number of parameters to be estimated. Similarly, to test whether the irregular sum of squares is significantly different from zero one can use the fact that, under $H_0$

$$\sum_{p=1}^{k/2} k(A_p^2 + B_p^2) - \sum_{p=1}^{k/2} kA_p^2 / 2\bar{X} \sim \chi^2_{k-M-1} \quad (3.12)$$

which amounts to comparing the irregular sum of squares with its distribution under $H_0$.

Now, the deviations from the null hypothesis can be examined by dividing the sample variance of $\{X_t\}$ into random and
individual harmonic contributions as described in Pocock[123].

Since a Poisson variate has variance equal to the mean \( \lambda \) (estimated by \( \bar{X} \)), the random component of sample variance is defined as \( \bar{X} \). The amount of sample variance in excess of \( \bar{X} \) can be divided into harmonic contributions.

The contribution of the p-th harmonic to the (excess) sample variance is defined as

\[
\frac{k(A_p^2 + B_p^2)}{2} - 2\bar{X} \quad \frac{1}{k-1}
\]  

(3.13)

Thus, the hour-of-week component of sample variance is defined as

\[
\sum_{p=1}^{N-1} \frac{k(A_p^2 + B_p^2)}{2} - \bar{M}X
\]  

(3.14)

while the remaining component of variance is attributed to irregular variations (where the term 'irregular' was defined earlier).

3.3.3 Data analysis

For the data mentioned in the introduction to this chapter we have that \( N=13 \) weeks (\( k=2184 \) hours). Thus, using the notation of sub-section 3.3.1, \( X_1 \)=the number of arrivals between 0000 hours and 0100 hours on the 2nd of July and \( X_{2184} \)=the number of arrivals between 2300 hours and 2400
hours on the 30th of September of the year 1984. Figure 3.1 shows a collapsed plot of the data by hour-of-week. No statistical test is required to see that there is a significant hour-of-week effect on the pattern of patient arrivals, especially with respect to hour-of-day. A detailed description of this pattern, in statistical terms, is presented below.

The total number of unscheduled patient arrivals during that period of 13 weeks is 8,711. Thus,
The sample mean of arrivals per hour = 4.0
The sample variance of arrivals per hour = 11.1

Using the definition of equation (3.7), the index of dispersion \( I = \frac{11.1}{4.0} = 2.8 \) which, as expected, is highly significant under \( H_0 \).

We now apply harmonic analysis to the above mentioned data. Since there are 2184 hours, a total of 1092 harmonics were considered (the last one consisted of the cosine term only). The hour-of-week effect on each \( X_t \), defined by (3.3), is formed by combining 84 out of the 1092 harmonics while the irregular effect is formed by combining \( 1092 - 84 = 1008 \) harmonics.

The total corrected sum of squares has a value of 24176.4 while the sums of squares due to the hour-of-week effect and due to irregular effect are equal to 14564.0 and 9612.5 respectively. By formulae (3.11) and (3.12) both of these
effects are highly significant.

The sample variance 11.1 has a random component equal to 4.0 (36.0%). From equation (3.14), the component of sample variance attributable to the hour-of-week effect has a value of 6.4 (57.5%) and the remaining value of 0.7 (6.5%) is attributed to irregular variations.

Now consider the importance of individual harmonic components. Figure 3.2 shows a plot of the contributions of individual harmonics to the total sum of squares. Twenty nine harmonics are significant at the 0.01 level and a further 76 at the 0.05 level. Eight of the former harmonics and 5 of the latter are components of the hour-of-week effect defined by (3.3). However, as expected, by far the most significant harmonic is the one with period 24 hours. From formula (3.13), the percentage contribution of this harmonic to the total sample variance is 50.3% and to the hour-of-week variation 87.4%. In addition, the harmonic with period 12 hours contributes 2.57% to the sample variance and 4.46% to the hour-of-week effect; the harmonic with period 8 hours accounts for 2.67% of the sample variance and for 4.65% of the hour-of-week variation; and the harmonic with period of one week accounts for 0.61% of the sample variance and for 1.06% of the hour-of-week variation. The rest of the significant harmonics have much smaller percentage contributions to the sample variance.

Overall, the results indicate a substantially anomalous
pattern of the unscheduled patient arrivals per hour to the A/E department. However, since 93.5% of the total variation is attributable to the hour-of-week and random components, we may assume that, for all practical purposes, the hourly arrival rate is constant for each hour of the week which, in turn, means that the weekly arrival rate is time-independent over the period of observation. Unbiased estimates for those 168 hourly arrival rates may then be obtained by averaging the corresponding counts of patient arrivals.

The pattern of hourly patient arrivals described above is quite similar to the ones prevailing during the rest of the 13 week periods of the year 1984: 2 January - 1 April, 2 April - 1 July and 1 October - 30 December. Table 3.1 gives summary results for all of the four 13-week time periods of the year 1984.

3.4 Categorical data analysis

The primary question we seek to answer in this section is: Do patients arriving during certain physician shift periods experience longer waiting and treatment times than others or is the pattern random throughout the week? Of interest are not only the average waiting and treatment times per shift but also how many patients experience what delays and whether these numbers of patients are (statistically) significantly different from expectations. Some secondary questions, arising from the primary one, are also considered.
The term 'waiting time' is defined here as the time spell from a patient's arrival to the time of the patient's first encounter with a physician ready to deal with the patient's complaint; the term 'treatment time' is defined as the difference between the time of that encounter and the time of the patient's departure from the A/E department. Thus, waiting times are underestimates and treatment times overestimates of the actual ones. For example, waiting times due to service interruptions and due to waiting for the results of laboratory tests are included in the treatment times, which also include delays to admission of some patients due to full occupancy of the inpatient departments concerned. On the other hand, more accurate measurements of waiting and treatment times were impossible with the resources available. The problem of accurate measurements of patient waiting and treatment times has also been reported in the study on the analysis of the workload of the A/E department of Leeds general infirmary[2].

The above question was examined using the statistical method of categorical data analysis. Patient visits were cross-classified according to physician shifts during which the arrivals occurred and according to five waiting and treatment time intervals, respectively. Before we discuss the results of the data analysis we present a brief account of the methodology used.
3.4.1 Methodology

Consider an r x c contingency table consisting of counts $n_{ij}$ of patients who are classified according to the i-th category of the row variable and the j-th category of the column variable ($i=1,2,...,r; j=1,2,...,c$). Denote by $m_{ij}$ the expected frequency in the (i,j) cell of the table and let

$$y_{ij} = \ln m_{ij}.$$  \hspace{1cm} (3.15)

Then, the $y_{ij}$s may be decomposed in the form

$$y_{ij} = \mu + u_1(i) + u_2(j) + u_{12}(ij),$$\hspace{1cm} (3.16)

where $\mu$ denotes the overall mean and $u_1$, $u_2$ and $u_{12}$ denote the row, column and interaction effects, respectively, under the assumption that

$$\sum_i u_1(i) = \sum_j u_2(j) = \sum_i u_{12}(ij) = \sum_j u_{12}(ij) = 0.$$ \hspace{1cm} (3.17)

One hypothesis of interest is whether the row and column variables of the contingency table are independent of each other. This hypothesis is equivalent to

$$H_0: u_{12}(ij) = 0 \text{ for all } (i,j).$$ \hspace{1cm} (3.18)

The hypothesis of independence, $H_0$, may be tested by using the likelihood ratio test statistic defined as \cite{22}

$$G = 2 \sum_i \sum_j n_{ij} \ln \left(\frac{n_{ij}}{\hat{m}_{ij}}\right).$$ \hspace{1cm} (3.19)

Under $H_0$, $G$ has the $\chi^2$ distribution on $(r-1)(c-1)$ degrees of freedom. The $\hat{m}_{ij}$s in (3.19) are the maximum likelihood
estimates of the expected cell frequencies which are given by
\[ \hat{m}_{ij} = n_{i+} n_{+j} / n_{++}, \] (3.20)
provided that there are no zero entries in the contingency table. In the presence of zero entries, the \( \hat{m}_{ij} \)s are usually obtained iteratively while the degrees of freedom of the \( \chi^2 \) distribution are reduced by the number of zeroes[22]. In equation (3.20) the '++' subscript denotes summation over the corresponding index.

A significant overall chi-squared for an \( r \times c \) contingency table indicates non-independence of the two variables, but provides no information as to whether non-independence occurs throughout or in a specific part of the table. Therefore, one would like to make additional comparisons of cells within the table as in the analysis of variance where, having found that a set of means differ, one wants to identify just which means differ from which others. The procedure used here in order to identify the categories responsible for a significant chi-squared value involves examination of the adjusted residuals defined by
\[ d_{ij} = e_{ij} / \sqrt{v_{ij}}, \] (3.21)
where \( e_{ij} \) are the standardised residuals of the form
\[ e_{ij} = (n_{ij} - \hat{m}_{ij}) / \sqrt{\hat{m}_{ij}}, \] (3.22)
and \( v_{ij} \) are their variances estimated by...
Both the standardised and the adjusted residuals are approximately distributed according to the standard normal distribution\( [22,71] \). It is therefore possible to test their statistical significance using the standard normal distribution. However, Haberman\([71] \) shows that the adjusted residuals are more sensitive than the standardised residuals in detecting significantly larger or smaller observations (outliers) than expected under the model of independence. Smaller (larger) observations are indicated by negative (positive) values of the adjusted residuals.

3.4.2 Data analysis

The data analysed in this section refer to the complete sample of patient visits to the A/E department from the 21st hour of 1 July to the 9th hour of 30 September. This time period contains exactly 13 of each weekly physician shift. Sample means and sample standard deviations of patient waiting and treatment times per shift are shown in Table 3.2. These aggregate statistics as well as the two-way contingency tables analysed in this section were produced routinely by the computerised patient record system described in the last chapter.

We first test the hypotheses that patient waiting and treatment times are independent of the physician shift during

\[
v_{ij} = \left(1 - \frac{n_{i+}}{n_{++}}\right)\left(1 - \frac{n_{+j}}{n_{++}}\right). \quad (3.23)
\]
which the arrival occurred. In the absence of any guidelines for waiting time limits in an A/E department we considered five time intervals of waiting and treatment times selected subjectively following discussion with the consultant: four of length 15 and 30 mins, respectively, and one which includes waiting and treatment times longer than 1 and 2 hours respectively.

In relation to the theory outlined in the last section, the above hypotheses are equivalent to the hypothesis of independence, as stated in (3.18), in the corresponding contingency tables. However, it is noted that the test of this hypothesis is conditioned by the specific time intervals considered. Also, at this level of analysis, the possible presence of autocorrelations in the data is ignored.

Tables 3.3(a) and 3.3(b) show the contingency tables shift by waiting times and shift by treatment times, respectively. The values of the G statistic of equation (3.19) were found equal to 409.1 and 208.5 for the waiting and treatment time tables, respectively. Although both of these values are highly significant, under $H_0$, they provide evidence that the pattern in the treatment times table is less systematic (or more random) than the pattern in the waiting time table. Individual interactions between the categories of the row and column variables can be examined by use of the adjusted residuals given in parentheses in the corresponding tables.

In Table 3.3(a) the values of the adjusted residuals indicate
a significantly disproportionate number of patients waiting for longer times to be seen by physicians during the weekends when the A/E department tends to be very busy. Overall, the significant interactions between certain physician shifts and waiting times give reason to suspect that the number of medical staff allocated per shift does not reflect the pattern of patient demand. This point is pursued in the next chapter where we describe a procedure for allocating physicians to shifts proportionally to the patient arrival rate.

On the other hand the significant interactions in the treatment times table are more difficult to interpret. At first sight, it might be expected that the mix of patients per shift, the duration of time required by a physician to deal with patients and delays of the nature described in section 3.4 are all randomly distributed. However, some or all of these possible expectations are proved false by the significant value of the G statistic for the shift by treatment times table. A study to determine which of these factors affect significantly the pattern of the patient treatment times would be very complex, if at all possible. It would involve defining a severity index of the type of case as well as a detailed recording of both the time spells of face-to-face encounters of patients with physicians and the nature and duration of any delays. Therefore, a study of such detail was not undertaken here.

Instead, we examined whether there is a significant relationship between the duration of treatment times and
patients' disposal. Patients were classified as discharged, referred to an outpatient clinic and admitted. There are three reasons for this classification. First, the patient flow, upon completion of the appropriate treatment given in the A/E department, can only be disturbed for admitted patients due to a full occupancy of the inpatient departments concerned. For all other patients the time of departure coincides with the time of treatment termination. Second, it was pointed out by the consultant that patients requiring hospitalisation require, in general, longer spells of 'actual treatment' in the A/E department for stabilisation of their condition. However, actual treatment times of more than two hours would rarely be necessary. For example, cardiac arrest cases would be admitted to the coronary (or intensive) care unit immediately after resuscitation had been given in the A/E department and, similarly, seriously injured patients would be admitted to the appropriate inpatient department. Therefore, it was of interest to examine the existing relative tendency. Excessive treatment times, longer than two hours, for admitted patients would, it was thought, indicate that these patients experience prolonged stays in the A/E department due to the inability of inpatient departments to accommodate them at the time when a request for a bed was issued. Again, it was difficult to investigate the validity of this argument in quantitative terms. Third, patients who were referred to outpatient clinics were grouped together to form a separate category because such patients are regarded as intermediate cases, in terms of seriousness of their condition, between discharges and admissions. Also,
the time of departure of these patients coincides, in
general, with the time of completion of their service in the
A/E department because no requests for beds, or other
'restrictive' facilities, are made by the A/E department to
the outpatient clinics.

Tables 3.3(c) and 3.3(d) show the contingency tables shift by
disposal and disposal by treatment times, respectively. The
_corresponding values of the G statistic were found equal to
134.4 and 2939.9 which, under Ho, are highly significant when
compared with the values of the χ² distribution on 46 and 8
degrees of freedom. In Table 3.3(c) the values of the
adjusted residuals indicate that during some shifts (e.g. Fri
9pm - Sat 9am and Sat 9pm - Sun 9am) there are
disproportionally many patients needing inpatient admission;
the values in Table 3.3(d) provide evidence that patients who
are to be admitted to the hospital through the A/E
department experience very long stays in that department
compared to other patients. This is particularly so for
treatment times longer than two hours. Overall, by comparing
the values of the adjusted residuals shown in Tables 3.3(b–d)
it may be concluded (within the degree of accuracy imposed by
the data limitations) that one of the reasons for longer
treatment times during certain shifts is that during these
shifts there are many arrivals of admitted patients. In
general, these patients stay longer in the A/E department.
One of the reasons for these long stays is that these
patients experience the results of some sort of instability
in the hospital operation produced by urgent requests for
service to others. In order to reach these conclusions we used here both quantitative analysis of available data and the consultant's judgement and experience. However, similar conclusions with respect to treatment times in the A/E department are reported in a study covering the A/E departments of the whole Wessex health region [151].

Tables 3.3(b-d) may be seen as consisting of the marginal totals of the underlying three-dimensional table: shift by treatment times by disposal. Although the analysis described above was adequate for our purposes it excludes consideration of conditional independence and three-factor interaction. However, as a matter of curiosity, we analysed the above three-dimensional table in which the 'outpatients' category of disposal was collapsed with the 'discharged' category. The analysis showed that the only adequate fit to the data was given by the complete three-dimensional model [22].

3.5 Conclusions

The analysis presented in this chapter served the dual purpose of providing an analytical framework, based on established statistical methods, to obtain information about the timeliness of the patient care process in the A/E department and of providing background information for the case studies described in the next chapters.

Although the investigation described in this chapter was
specifically about the timeliness of the patient care process in the A/E department, the statistical methods are readily applicable to many other investigations. For example, it might be of interest to study the way patients use A/E services or the incidence of occurrence of specific types of emergencies. The requirements for such analyses are that the data consist of counts of patients having one or more characteristics of interest. Since such data are now readily available in many A/E departments in the UK, the information produced from such data analyses can also be used for interhospital comparison at regional and/or national level.

Owing to their relative simplicity, the statistical methods used in this chapter could be incorporated in computerised patient record systems such as the ones which are currently being installed in many A/E departments in the UK. Changes in population attitudes towards utilisation of A/E department resources and the department's performance would then be able to be monitored on the basis of more detailed analysis and, therefore, more sound information. For, as mentioned in the Wessex Region report[151], 'no solution to the increasing problems of the A/E departments can usefully be proposed without adequate and detailed information of what is actually occurring in the hospitals at the present time'.
Figure 3.1  Collapsed plot of patient arrivals to the A/E department per hour of the week during the period 2 July - 30 September 1984.
Table 3A. Random, base, day of week and angular coefficients of the sample variance of weekly counts of scheduled patient visits to the AIDS Outpatient Clinic. Certain periods of the year are.

(d) Harmonics 820-1092
Table 3.1  Random, hour-of-week and irregular components of the sample variances of hourly counts of unscheduled patient visits to the A/E department, during four consecutive 13-week periods of the year 1984.

<table>
<thead>
<tr>
<th>Component of variance</th>
<th>Period of year</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2 Jan.-1 April</td>
</tr>
<tr>
<td>Random</td>
<td>3.3 (40.1%)</td>
</tr>
<tr>
<td>Hour-of-week</td>
<td>4.7 (56.2%)</td>
</tr>
<tr>
<td>Irregular</td>
<td>0.3 ( 3.7%)</td>
</tr>
<tr>
<td>Total</td>
<td>8.2 (100%)</td>
</tr>
</tbody>
</table>
Table 3.2  Means and standard deviations of patient waiting and treatment times per physician shift of patient arrival during the period 2 July - 30 September 1984.

<table>
<thead>
<tr>
<th>Shift no.</th>
<th>Duration</th>
<th>No. of patient arrivals</th>
<th>Waiting time Mean</th>
<th>Std. dev.</th>
<th>Treatment time Mean</th>
<th>Std. dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Mon 0900-Mon 1300</td>
<td>346</td>
<td>16.9</td>
<td>17.2</td>
<td>50.3</td>
<td>48.7</td>
</tr>
<tr>
<td>2</td>
<td>Mon 1300-Mon 1700</td>
<td>354</td>
<td>19.5</td>
<td>19.3</td>
<td>61.8</td>
<td>88.8</td>
</tr>
<tr>
<td>3</td>
<td>Mon 1700-Mon 2100</td>
<td>370</td>
<td>22.6</td>
<td>20.0</td>
<td>54.3</td>
<td>49.5</td>
</tr>
<tr>
<td>4</td>
<td>Mon 2100-Tues 0900</td>
<td>254</td>
<td>20.3</td>
<td>20.5</td>
<td>57.9</td>
<td>48.8</td>
</tr>
<tr>
<td>5</td>
<td>Tues 0900-Tues 1300</td>
<td>306</td>
<td>17.4</td>
<td>16.9</td>
<td>55.5</td>
<td>92.4</td>
</tr>
<tr>
<td>6</td>
<td>Tues 1300-Tues 1700</td>
<td>307</td>
<td>15.3</td>
<td>14.0</td>
<td>59.1</td>
<td>55.0</td>
</tr>
<tr>
<td>7</td>
<td>Tues 1700-Tues 2100</td>
<td>308</td>
<td>18.3</td>
<td>15.2</td>
<td>53.5</td>
<td>49.8</td>
</tr>
<tr>
<td>8</td>
<td>Tues 2100-Wed 0900</td>
<td>249</td>
<td>17.8</td>
<td>18.8</td>
<td>56.1</td>
<td>50.8</td>
</tr>
<tr>
<td>9</td>
<td>Wed 0900-Wed 1300</td>
<td>321</td>
<td>15.5</td>
<td>15.8</td>
<td>58.1</td>
<td>118.8</td>
</tr>
<tr>
<td>10</td>
<td>Wed 1300-Wed 1700</td>
<td>299</td>
<td>14.4</td>
<td>15.6</td>
<td>56.6</td>
<td>47.8</td>
</tr>
<tr>
<td>11</td>
<td>Wed 1700-Wed 2100</td>
<td>276</td>
<td>19.1</td>
<td>15.1</td>
<td>55.2</td>
<td>92.1</td>
</tr>
<tr>
<td>12</td>
<td>Wed 2100-Thurs 0900</td>
<td>241</td>
<td>16.3</td>
<td>14.7</td>
<td>64.1</td>
<td>99.3</td>
</tr>
<tr>
<td>13</td>
<td>Thurs 0900-Thurs 1300</td>
<td>314</td>
<td>16.1</td>
<td>13.5</td>
<td>50.3</td>
<td>49.8</td>
</tr>
<tr>
<td>14</td>
<td>Thurs 1300-Thurs 1700</td>
<td>277</td>
<td>19.0</td>
<td>17.9</td>
<td>65.0</td>
<td>97.7</td>
</tr>
<tr>
<td>15</td>
<td>Thurs 1700-Thurs 2100</td>
<td>319</td>
<td>16.7</td>
<td>17.5</td>
<td>66.2</td>
<td>93.3</td>
</tr>
<tr>
<td>16</td>
<td>Thurs 2100-Fri 0900</td>
<td>246</td>
<td>22.2</td>
<td>21.5</td>
<td>62.1</td>
<td>47.1</td>
</tr>
<tr>
<td>17</td>
<td>Fri 0900-Fri 1300</td>
<td>312</td>
<td>16.6</td>
<td>16.1</td>
<td>56.1</td>
<td>46.9</td>
</tr>
<tr>
<td>18</td>
<td>Fri 1300-Fri 1700</td>
<td>327</td>
<td>18.9</td>
<td>18.4</td>
<td>71.7</td>
<td>95.5</td>
</tr>
<tr>
<td>19</td>
<td>Fri 1700-Fri 2100</td>
<td>340</td>
<td>18.2</td>
<td>16.2</td>
<td>65.8</td>
<td>138.7</td>
</tr>
<tr>
<td>20</td>
<td>Fri 2100-Sat 0900</td>
<td>263</td>
<td>18.2</td>
<td>18.5</td>
<td>58.1</td>
<td>51.7</td>
</tr>
<tr>
<td>21</td>
<td>Sat 0900-Sat 2100</td>
<td>1016</td>
<td>21.3</td>
<td>18.0</td>
<td>59.9</td>
<td>122.1</td>
</tr>
<tr>
<td>22</td>
<td>Sat 2100-Sun 0900</td>
<td>312</td>
<td>24.1</td>
<td>22.8</td>
<td>69.4</td>
<td>58.8</td>
</tr>
<tr>
<td>23</td>
<td>Sun 0900-Sun 2100</td>
<td>1053</td>
<td>24.7</td>
<td>20.1</td>
<td>52.7</td>
<td>96.0</td>
</tr>
<tr>
<td>24</td>
<td>Sun 2100-Mon 0900</td>
<td>276</td>
<td>14.0</td>
<td>22.0</td>
<td>30.1</td>
<td>47.7</td>
</tr>
</tbody>
</table>
Table 3.3  Pairwise classifications of patient visits to the A/E department with regard to physician shifts (shown in Table 3.2), waiting times, treatment times and disposal. The values of the adjusted residuals are given in parentheses (significance at 5% level is indicated by an *).

(a) Shift vs. Waiting time

<table>
<thead>
<tr>
<th>Shift no.</th>
<th>0000-0015</th>
<th>0016-0030</th>
<th>0031-0045</th>
<th>0046-0100</th>
<th>0101-</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>170 (1.87)</td>
<td>126 (1.64)</td>
<td>31 (-2.71)*</td>
<td>10 (-1.99)</td>
<td>9 (-1.56)</td>
<td>346</td>
</tr>
<tr>
<td>2</td>
<td>158 (0.15)</td>
<td>107 (-0.88)</td>
<td>60 (1.69)</td>
<td>14 (-1.10)</td>
<td>15 (-0.02)</td>
<td>354</td>
</tr>
<tr>
<td>3</td>
<td>137 (-2.85)*</td>
<td>125 (0.59)</td>
<td>71 (3.00)*</td>
<td>17 (-0.56)</td>
<td>20 (1.12)</td>
<td>370</td>
</tr>
<tr>
<td>4</td>
<td>117 (0.60)</td>
<td>69 (-1.60)</td>
<td>42 (1.23)</td>
<td>8 (-1.51)</td>
<td>18 (2.26)*</td>
<td>254</td>
</tr>
<tr>
<td>5</td>
<td>154 (2.19)*</td>
<td>93 (-0.75)</td>
<td>39 (-0.60)</td>
<td>12 (-1.04)</td>
<td>8 (-1.45)</td>
<td>306</td>
</tr>
<tr>
<td>6</td>
<td>159 (2.72)*</td>
<td>109 (1.19)</td>
<td>26 (-2.80)*</td>
<td>7 (-2.36)*</td>
<td>6 (-2.04)*</td>
<td>307</td>
</tr>
<tr>
<td>7</td>
<td>134 (-0.26)</td>
<td>120 (2.52)*</td>
<td>34 (-1.48)</td>
<td>12 (-1.07)</td>
<td>8 (-1.47)</td>
<td>308</td>
</tr>
<tr>
<td>8</td>
<td>126 (2.05)*</td>
<td>70 (-1.46)</td>
<td>28 (-1.23)</td>
<td>14 (0.28)</td>
<td>11 (0.13)</td>
<td>249</td>
</tr>
<tr>
<td>9</td>
<td>175 (3.70)*</td>
<td>103 (-0.11)</td>
<td>25 (-3.23)*</td>
<td>8 (-2.24)*</td>
<td>10 (-1.03)</td>
<td>321</td>
</tr>
<tr>
<td>10</td>
<td>179 (5.54)*</td>
<td>80 (-2.11)*</td>
<td>27 (-2.48)*</td>
<td>9 (-1.75)</td>
<td>4 (-2.55)*</td>
<td>299</td>
</tr>
<tr>
<td>11</td>
<td>118 (-0.50)</td>
<td>102 (1.65)</td>
<td>37 (-0.24)</td>
<td>12 (-0.67)</td>
<td>7 (-1.44)</td>
<td>276</td>
</tr>
<tr>
<td>12</td>
<td>127 (2.68)*</td>
<td>67 (-1.64)</td>
<td>31 (-0.48)</td>
<td>12 (-0.18)</td>
<td>4 (-2.03)*</td>
<td>241</td>
</tr>
<tr>
<td>13</td>
<td>160 (2.44)*</td>
<td>101 (-3.09)</td>
<td>39 (-0.78)</td>
<td>9 (-1.91)</td>
<td>5 (-2.38)*</td>
<td>314</td>
</tr>
<tr>
<td>14</td>
<td>126 (0.43)</td>
<td>86 (-0.48)</td>
<td>40 (0.26)</td>
<td>13 (-0.41)</td>
<td>12 (0.06)</td>
<td>277</td>
</tr>
<tr>
<td>15</td>
<td>142 (0.10)</td>
<td>98 (-0.64)</td>
<td>51 (1.09)</td>
<td>15 (-0.43)</td>
<td>13 (-0.17)</td>
<td>319</td>
</tr>
<tr>
<td>16</td>
<td>105 (-0.50)</td>
<td>65 (-2.21)*</td>
<td>33 (-0.23)</td>
<td>23 (-2.95)*</td>
<td>20 (3.05)*</td>
<td>246</td>
</tr>
<tr>
<td>17</td>
<td>157 (2.21)*</td>
<td>108 (0.86)</td>
<td>30 (-2.23)*</td>
<td>7 (-2.41)*</td>
<td>10 (-0.94)</td>
<td>312</td>
</tr>
<tr>
<td>18</td>
<td>156 (1.29)</td>
<td>95 (-1.61)</td>
<td>43 (-0.40)</td>
<td>18 (0.23)</td>
<td>15 (0.30)</td>
<td>327</td>
</tr>
<tr>
<td>19</td>
<td>153 (0.29)</td>
<td>120 (-1.17)</td>
<td>45 (-0.37)</td>
<td>7 (-1.43)</td>
<td>10 (-1.23)</td>
<td>340</td>
</tr>
<tr>
<td>20</td>
<td>129 (1.60)</td>
<td>77 (-1.09)</td>
<td>25 (-2.09)*</td>
<td>20 (1.76)</td>
<td>12 (0.25)</td>
<td>263</td>
</tr>
<tr>
<td>21</td>
<td>378 (-4.80)*</td>
<td>345 (1.15)</td>
<td>196 (5.28)*</td>
<td>53 (1.60)</td>
<td>31 (3.81)*</td>
<td>374</td>
</tr>
<tr>
<td>22</td>
<td>129 (-1.05)</td>
<td>72 (-3.57)*</td>
<td>53 (1.60)</td>
<td>31 (3.81)*</td>
<td>27 (3.91)*</td>
<td>312</td>
</tr>
<tr>
<td>23</td>
<td>318 (-9.78)*</td>
<td>400 (4.15)*</td>
<td>173 (2.52)*</td>
<td>90 (5.16)*</td>
<td>72 (4.42)*</td>
<td>1053</td>
</tr>
<tr>
<td>24</td>
<td>135 (1.59)</td>
<td>74 (-2.01)*</td>
<td>29 (-1.66)</td>
<td>17 (0.71)</td>
<td>21 (2.80)*</td>
<td>276</td>
</tr>
<tr>
<td>Total</td>
<td>3842</td>
<td>2812</td>
<td>1208</td>
<td>454</td>
<td>370</td>
<td>8686</td>
</tr>
</tbody>
</table>
### Shift vs. Disposal

<table>
<thead>
<tr>
<th>Shift no.</th>
<th>Discharged</th>
<th>Outpatients</th>
<th>Admitted</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>269 (2.16)*</td>
<td>25 (0.64)</td>
<td>52 (-2.75)*</td>
<td>346</td>
</tr>
<tr>
<td>2</td>
<td>258 (0.09)</td>
<td>23 (0.08)</td>
<td>73 (-0.14)</td>
<td>354</td>
</tr>
<tr>
<td>3</td>
<td>268 (-0.11)</td>
<td>27 (0.72)</td>
<td>75 (-0.13)</td>
<td>370</td>
</tr>
<tr>
<td>4</td>
<td>157 (-3.95)*</td>
<td>20 (0.97)</td>
<td>77 (3.74)*</td>
<td>254</td>
</tr>
<tr>
<td>5</td>
<td>233 (1.38)</td>
<td>20 (0.10)</td>
<td>53 (-1.58)</td>
<td>306</td>
</tr>
<tr>
<td>6</td>
<td>216 (-0.93)</td>
<td>27 (1.74)</td>
<td>64 (-0.03)</td>
<td>307</td>
</tr>
<tr>
<td>7</td>
<td>207 (-2.19)*</td>
<td>26 (1.49)</td>
<td>75 (1.51)</td>
<td>308</td>
</tr>
<tr>
<td>8</td>
<td>170 (-1.58)</td>
<td>12 (-1.03)</td>
<td>67 (2.36)*</td>
<td>249</td>
</tr>
<tr>
<td>9</td>
<td>234 (0.09)</td>
<td>21 (0.11)</td>
<td>66 (-0.16)</td>
<td>321</td>
</tr>
<tr>
<td>10</td>
<td>213 (-0.57)</td>
<td>24 (1.17)</td>
<td>62 (-0.08)</td>
<td>299</td>
</tr>
<tr>
<td>11</td>
<td>200 (-0.08)</td>
<td>18 (0.08)</td>
<td>58 (0.04)</td>
<td>276</td>
</tr>
<tr>
<td>12</td>
<td>160 (-2.22)*</td>
<td>13 (-0.65)</td>
<td>68 (2.82)*</td>
<td>241</td>
</tr>
<tr>
<td>13</td>
<td>240 (1.52)</td>
<td>19 (-0.26)</td>
<td>55 (-1.51)</td>
<td>314</td>
</tr>
<tr>
<td>14</td>
<td>190 (-1.55)</td>
<td>22 (1.07)</td>
<td>65 (1.06)</td>
<td>277</td>
</tr>
<tr>
<td>15</td>
<td>232 (0.02)</td>
<td>13 (-1.73)</td>
<td>74 (1.02)</td>
<td>319</td>
</tr>
<tr>
<td>16</td>
<td>165 (-2.00)*</td>
<td>9 (-1.78)</td>
<td>72 (3.27)*</td>
<td>246</td>
</tr>
<tr>
<td>17</td>
<td>227 (0.03)</td>
<td>17 (-0.70)</td>
<td>68 (0.39)</td>
<td>312</td>
</tr>
<tr>
<td>18</td>
<td>234 (-0.46)</td>
<td>17 (-0.91)</td>
<td>76 (1.05)</td>
<td>327</td>
</tr>
<tr>
<td>19</td>
<td>245 (-0.26)</td>
<td>22 (0.05)</td>
<td>73 (0.26)</td>
<td>340</td>
</tr>
<tr>
<td>20</td>
<td>175 (-2.27)*</td>
<td>17 (0.04)</td>
<td>71 (2.46)*</td>
<td>263</td>
</tr>
<tr>
<td>21</td>
<td>763 (3.34)*</td>
<td>52 (-1.78)</td>
<td>181 (-2.59)*</td>
<td>1016</td>
</tr>
<tr>
<td>22</td>
<td>212 (-1.91)</td>
<td>16 (-0.94)</td>
<td>84 (2.66)*</td>
<td>312</td>
</tr>
<tr>
<td>23</td>
<td>827 (4.55)*</td>
<td>83 (2.09)*</td>
<td>143 (-6.25)*</td>
<td>1053</td>
</tr>
<tr>
<td>24</td>
<td>198 (-0.36)</td>
<td>13 (-1.17)</td>
<td>65 (1.09)</td>
<td>276</td>
</tr>
</tbody>
</table>

**Total** 6313 556 1817 8686
(d) Treatment time vs. Disposal

<table>
<thead>
<tr>
<th>Treatment time</th>
<th>Discharged</th>
<th>Outpatients</th>
<th>Admitted</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>0000-0030</td>
<td>2956 (33.46)*</td>
<td>150 (-4.70)*</td>
<td>42 (-33.83)*</td>
<td>3148</td>
</tr>
<tr>
<td>0031-0100</td>
<td>1988 (19.53)*</td>
<td>220 (6.06)*</td>
<td>256 (-15.18)*</td>
<td>2464</td>
</tr>
<tr>
<td>0101-0130</td>
<td>786 (-11.85)*</td>
<td>99 (1.73)</td>
<td>440 (11.95)*</td>
<td>1325</td>
</tr>
<tr>
<td>0131-0200</td>
<td>307 (-19.56)*</td>
<td>37 (-1.56)</td>
<td>389 (22.37)*</td>
<td>733</td>
</tr>
<tr>
<td>0201-</td>
<td>276 (-34.65)*</td>
<td>50 (-2.05)*</td>
<td>690 (39.19)*</td>
<td>1016</td>
</tr>
<tr>
<td>Total</td>
<td>6313</td>
<td>556</td>
<td>1817</td>
<td>8686</td>
</tr>
</tbody>
</table>
CHAPTER 4

ALLOCATING A/E DEPARTMENT PHYSICIANS TO SHIFTS

4.1 Introduction

The problem of cyclical allocation of manpower to weekly shifts arises in situations where the demand for service fluctuates in the course of a week. The simplest type of allocation involves non-overlapping shifts where demand fluctuations are reflected by requirements that vary at time intervals coinciding with the shift lengths. Then each shift can be treated independently in determining appropriate allocations, giving a fluctuating staff size which follows the same pattern as the shift-to-shift requirement. However, in many situations the time intervals over which demand fluctuates do not coincide with the shift length. This type of demand fluctuation gives rise to the so-called overlapping shift allocation problem.

The problem we are concerned with in this chapter may be viewed within the framework of the overlapping shift allocation problem. We describe a method for allocating physicians to weekly shifts in an A/E department so that the patient waiting times and queue sizes in the department's waiting room are reduced. The method takes into account that the hourly patient arrival rate varies with time; that there
must be at least one physician present in the department at any time; that the number of physicians employed in the department is fixed; and that a feasible set of physician shifts is provided a priori (by the consultant). The method uses two models solved by dynamic programming. The first finds the optimal solution to the problem of allocating physicians per hour of the week proportionally to the corresponding patient arrival rate; and the second uses this solution to smooth out the hourly discrepancies in the number of physicians within a shift. The solution is then assessed by computer simulation.

4.2 Review of the literature

The problem of allocating manpower to shifts falls within the broader class of problems concerned with the optimal distribution of effort[148]. In general, such problems may be posed as: given a demand for staff services in each shift within a planning horizon, how does one allocate staff to meet an appropriate objective? For example, it might be desirable to determine staff needs per shift to meet demand requirements at a minimum cost or, given the total staff size, to allocate staff to shifts in an efficient way according to a selected criterion. The shifts may be fixed or variable while the demand may be deterministic or probabilistic. A useful and much studied special case of the general shift allocation problem is the cyclic staff allocation problem, in which both shifts and demand for staff
services are cyclic. The techniques proposed in the literature to tackle such problems include mathematical (mainly integer) programming, queueing theory and simulation.

Baker[18] surveys the basic mathematical models used, up to 1975, for manpower allocation to cyclic shifts with cyclic demand for staff and reviews several problem areas in which these models have been applied. Two other papers by the same author deal with the problems of scheduling full-time staff only[17] and full-time and part-time staff[16], to meet a predetermined cyclic demand. Also, Baker and Magazine[19] generalise the work of Brownell and Lowerre[34] on the problem of scheduling days off in continuous (seven-day-a-week) operations under a variety of days-off policies, when demand for manpower differs between weekdays and weekend days. For each policy, Baker and Magazine[19] give a formula for the minimum workforce size and a schedule construction algorithm.

Luce[100] tackles the problem of scheduling telephone operators. He shows that integer programs are impractical to solve, due to the large number of shift patterns, and he suggests the use of a heuristic round-off procedure. He also notes that even a linear programming approach will ultimately result in the need to resort to heuristic tactics in solving practical versions of the allocation problem. Luce proposes an objective of matching requirements with allocations as closely as possible by minimising a quadratic function heuristically. Bartholdi[20] formulates the cyclic staff
allocation problem as a set covering problem which he solves using a round-off heuristic with a prespecified bound on the absolute error. Following Luce, he further establishes the appropriateness of the heuristic approach by showing that cyclic staff scheduling is NP-complete.

Zanakis and Lawrence[156] propose a mixed integer programming model for allocating scarce resources (including manpower) in a military environment and Kolesar et al[90] determine the number of police patrol cars required to meet demand for service by using queueing theory and integer programming.

In the health operational research literature a good deal of effort has been devoted to nursing staff scheduling. Abernathy et al[1] provide a comprehensive discussion of the nurse staffing problem within the overall hospital objective to hold costs at a reasonable level while maintaining or improving the quality of care. Stimson and Stimson[141] describe the role of operational research in the solution of nurse staffing problems.

Wolfe and Young[152,153] devise a method for predicting nurse staffing requirements in the face of the day-to-day variability of the daily census and of the seriousness of conditions of the patients. Then, they propose a controlled variable staffing policy in which fewer nurses are permanently assigned to particular wards while 'float nurses' are moved between wards on flexible assignments. The authors argue that this policy would enable the hospital cope
more efficiently with the stochastic nature of patients' demands by varying the daily staffing patterns to match the fluctuating requirements. The nurse scheduling model is solved by linear programming and the solution provides staffing needs that reflect both the qualitative and the quantitative aspects of specific assignments of personnel. Maier-Rothe and Wolfe[103] describe a mathematical procedure for nurse allocation to meet average staffing requirements in a manner consistent with hospital personnel policies and employee's preferences. The procedure allows adjustment of staffing twice daily in accordance with an index of the latest patient care requirements. Hershey et al[77] evaluate the fixed and variable nurse staffing policies by monte carlo simulation using work sampling estimates of the workload generated by patients in different 'need categories'. Their illustrative example shows that a reduction in nurse hours of the order of 10% could be achieved by variable staffing.

The problem of physician allocation to shifts has mostly been viewed as a management problem of the queueing type. Fetter and Thompson[61] use a simulation approach to allocate physicians to shifts in an outpatient setting (where patients are mainly seen on an appointment basis), so that some balance is achieved between the patients' waiting time and the physicians' idle time. Bolling[26] projects the emergency room demands of the population by month of the year, day of the week and hour of the day and suggests the use of queueing modelling to determine the number of treatment rooms, waiting room facilities and staffing
patterns for various planning horizons. Glenn and Roberts[64] devise a simulation model to investigate the relationship between the amount of resource allocation (including physician allocation) to an outpatient clinic and selected measures of clinic performance such as the time spent by patients in the clinic and the utilisation of resources. The objective of the model is to provide insight into the sensitivity of the clinic's performance to changes in resource allocation and to provide a method for making resource allocation decisions. Rising et al[127] develop a simulation model of a university health service outpatient clinic with the objective to schedule physicians and patient appointments more efficiently. The implementation of the model resulted in appreciable improvement of the system's performance. Keller and Laughunn[84] describe an application of queueing theory as an analytical framework for evaluating the adequacy of physician capacity in an outpatient clinic for medically indigent patients. The authors focus their attention on balancing physician cost and patient waiting cost. Ladany and Turban[92] model an emergency room of a hospital as a priority queueing system to provide the framework for the planning and staffing process. Carlson et al[36] describe a study carried out for an outpatient health care setting in the United States in order to determine the best combination of facilities and personnel available, so that the profit to the clinic is maximised while its performance remains at acceptable levels. The authors obtain a solution to the problem using a recursive method which combines linear programming and simulation models.
4.3 Discussion of the problem

In chapter 1 we identified the ease of access to medical care as one of the most important attributes in the patient's assessment of the quality of service provided. In particular, for emergency patients, any time lost before high level medical care is rendered may be critical to the patient's chances of recovery. This critical time is defined as the period between the time a person first seeks emergency medical help and the time when emergency care is rendered in the A/E department. The A/E department's response time is defined as the part of the critical time which begins when a patient arrives at the A/E department and ends when the patient meets a physician ready to provide the particular care needed. This response time is clearly a function of the available facilities (including staff) and the number and mix of patients in the department at the time of patient's arrival.

Most people assume that there is a trade off between the A/E department's response time and the number of available resources. This assumes that pure waste of resources is rare and that improvements require increase in the quantity of the existing resources. However, reallocation of the existing resources may improve the situation.

Owing to the lack of any documented information about the
pattern of the patient arrival rate, the existing physician shift allocation policy at the A/E department of St. Peter's hospital was based on pure intuition and on physician preferences. As a result, the existing physician allocation pattern did not correspond to the pattern of the patient arrival rate. This imbalance was observed when the collapsed plot of hourly counts of patient arrivals, shown in Figure 3.1, was compared with the number of physicians allocated per shift and was reflected by the significant differences in patient waiting times between shifts, discussed in the last chapter. Therefore, an alternative way of allocating physicians to shifts was sought that would result in a reduction of the overall patient waiting times and queue sizes. In view of the management realities in the A/E department, discussed in chapter 1, three constraints should be taken into account in formulating the problem:

(1) that the physician shifts are prespecified (i.e. physician preferences should be taken into account),
(2) that at least one physician is present in the A/E department at any time, and
(3) that the number of physician-hours available per week is limited to the existing one.

At first sight, the problem just described could easily be identified as being amenable to analysis by queueing modelling. However, observation of the operation of the A/E department of St. Peter's hospital brought about the realisation that queueing system modelling would only marginally represent reality. In support of this argument we
cite the following two reasons.

(1) There is no clear-cut queueing discipline in the system. Patients are, in principle, seen on a first-come first-served (FCFS) basis. The one official class of priority cases (such as those needed resuscitation) may be ignored at this point since they represent a very small proportion of the total workload. However, other patients may be given priorities to service unofficially (e.g. young children, old people and pregnant women irrespective of the severity of their complaint; and other patients according to the apparent nature and severity of their complaint). Further, these priorities are not clearly ordered.

(2) The beginning of service of one patient does not necessarily coincide with another patient's departure. This is a very usual situation since patients may be temporarily accommodated in beds and seen periodically by a physician. This fact complicates the calculation of the actual treatment (service) times of patients. It was also noted that some patients are attended to by more than one physician at a time (e.g. severe cases).

Up to this point, the problem of allocating physicians to shifts has been seen within the framework of the existing operating rules of the A/E department. Alternatively, one might find it desirable to introduce new operating rules (e.g. a 'triage system' in order to assign priorities for service to patients) so as to make the system conform approximately to the assumptions of existing queueing models. However it was infeasible to introduce such innovations in
the A/E department of St. Peter's hospital. Neither was it possible to time in detail the physician-to-patient interactions, for service time data collection, since such a process would require additional staff to carry out this task.

4.4 Problem formulation

From the above considerations we concluded that we were faced with a problem of a 'queueing nature' that, in view of the system's realities, was not amenable to queueing theory or queueing simulation analysis. In order to tackle the problem we decided to use a more general (or flexible) model in the sense that its application would be independent of the specific arrival and service time distributions and of the queueing discipline. As such a model we selected the utilisation factor which, as is well known, has direct effect on all of the performance characteristics of queueing systems. In fact, we decided to allocate physicians to shifts proportionally (or approximately so) to the patient arrival rate, on the assumption that the -unknown- service rate is constant over time. The physician allocation methodology is described below.

4.4.1 Physician allocation per hour of the week

The harmonic analysis results of the last chapter showed that more than 90% of the sample variance of the time series of
hourly counts of patient arrivals are explained by the random and hour-of-week effects. Therefore, it is reasonable to assume that the patient arrival rate \((\lambda_t)\) is constant for each hour \(t\) of the week over the period considered. This assumption enables us to tackle the problem of allocating physicians to shifts as a discrete one.

We first define the following notation:

\(N\) : no. of available physician-hours per week.
\(M = N - 168\) : no. of remaining physician-hours after allocating one physician per hour of week.
\(n_t\) : no. of physicians allocated to the \(t\)-th hour of the week \((t = 1, 2, \ldots, 168)\); and \(r_t = n_t - 1\).
\(b\) : no. of physicians whose working hours per week are considered in the allocation.

In order to allocate physicians per hour of the week in an approximately proportional manner to the corresponding patient arrival rate, we first determine the (real) numbers, \(q_t\), by

\[
q_t = \frac{M}{\sum_{t=1}^{168} \lambda_t} \lambda_{t=1, 2, \ldots, 168}, \tag{4.1}
\]

where \(\lambda\) is assumed to be constant following the above discussion on the patient arrival process. Then, the integer numbers \(r_t\) are obtained by solving the following problem by dynamic programming
minimise \( f_1 = \max_{1 \leq t \leq 168} \{|q_t - r_t|\} \)

Pl: subject to 
\[
\sum_{t=1}^{168} r_t = M \\
0 < r_t < b-1 \\
r_t \text{ integers.}
\]

Selection of this criterion for allocating physicians per hour of the week has certain advantages which hold, within the constraints of the problem, under the fairly 'loose' condition that the weekly patient arrival rate is constant over the period considered.

(1) It is suggested that the expected number of arriving patients corresponding to each physician on duty is approximately constant throughout a week. This might be judged desirable from the physician's point of view.

(2) It is ensured that the number of physicians present in the A/E department at any time of the week is closely related (in phase) to the patient arrival rate.

(3) For a queueing situation, the utilisation factor per hour of the week is kept approximately constant throughout a week, provided that the service rate is constant. This is expected to reduce unduly long queues during those hours of the week when the department tends to be very busy[13].
4.4.2 Physician allocation per shift

Let \( S = \{ s_0, s_1, \ldots, s_k \} \) denote a feasible set of the hours of the week when the number of physicians present in the A/E department may change. This set is provided by the consultant in charge of the A/E department.

We introduce the following additional notation:

\[ l_i = s_i - s_{i-1}; \quad i = 1, 2, \ldots, k \] shift duration in hours.

\[ d_i; \quad i = 1, 2, \ldots, k \] number of physician-hours, in excess of \( l_i \), allocated to time period \( l_i \) (\( d_i \) divisible by \( l_i \)).

Using the solution to problem P1, we smooth out the hourly discrepancies in the number of physicians within each time period \( l_i \). The solution vector \( (d_i)_{1 \leq i \leq k} \) is obtained by solving the following dynamic programming problem

\[
\text{minimise } f_2 = \max \left\{ \left| d_i - \sum_{j=s_i+1}^{s_i} r_j \right| \right\}
\]

\[ \text{P2: subject to } \sum_{i=1}^{k} d_i = M \]

\[ 0 \leq d_i \leq (b-1)l_i \]

\[ d_i \text{ integers divisible by } l_i \]

Computationally, end-effects are avoided by letting \( s_k = 168 \) depending on the choice of \( s_0 \). At this point it is noted that problem P2 does not always have a feasible solution. Necessary and sufficient conditions for the existence of a
non-empty set of feasible solutions to problem P2 are given in the following lemma.

**Lemma**

Let \( l \) denote the minimum of the \( l_i \)s (\( i=1,2,\ldots,k \)). Suppose that

\[
l_i = a_i l, \quad i=1,2,\ldots,k
\]

where the \( a_i \)s are integers such that

\[
1 \leq a_i \leq b-1, \quad i=1,2,\ldots,k.
\]

Then, a feasible solution to problem P2 exists if and only if \( M=h l \) for some integer \( h, \ h>0 \).

The proof is given in the Appendix of this chapter.

### 4.5 The simulation model

Once an optimal solution to the problem P2 has been obtained, the solution must be tested to see if it leads to improvement of the system's performance-characteristics of interest. In the absence of any suitable test-procedure, a rough assessment could be obtained by simulating the A/E department as a queueing system. Following the discussion of section 4.3, it is emphasised that the mere objective of the simulation was to provide some measure for comparison of the performance characteristics of the A/E department, obtained with the existing and with the proposed shift allocation patterns, under the assumptions underlying existing queueing models. Within this context only, the simulation exercise was judged
The simulation model was coded into a FORTRAN 77 computer program, using the event advancing (discrete event) programming technique, and it was run on the VAX 11/780, VMS computer of the RHC. Random numbers were generated from the appropriate NAG library routines. The logic of the program was a detailed flow account for each patient from the time of arrival until the time of his/her departure from the A/E department. The events were: patient's arrival, patient's initiation of service, patient's departure, change of hour of the week, change of shift and end of simulation. The queueing discipline was assumed FCFS while the department was assumed to have a finite capacity.

The time-dependent patient arrival distribution was incorporated into the simulation model as interarrival times using the following procedure. Let \( \{ X_{tn} : t=1,2,\ldots,168 ; n=1,2,\ldots,N \} \) be the sample record of the numbers of hourly arrivals over \( N \) weeks. Then,

\[
S_t(k) = \frac{1}{N} \sum_{n=1}^{N} \delta_{tn}(k)
\]

where

\[
\delta_{tn}(k) = \begin{cases} 
1, & X_{tn} \leq k \\
0, & \text{otherwise} 
\end{cases}
\]

denotes the sample cumulative distribution function for the number of arrivals on hour \( t \) of each week.
Suppose that we want to generate an arrival in hour \( r \). Let \( U \) denote a uniform deviate on \((0,1)\). Then, the number of arrivals on hour \( r \) is \( I_r = \inf \{ k: U < S_t(k); t = r \mod 168 \} \) and the \( j \)-th arrival on that hour occurs at \( r + V_j \) where \( \{ V_j: j = 1, 2, \ldots, I_r \} \) is a sequence of independent uniform variates. By this procedure we were reasonably assured, and the results confirmed it, that the arrival pattern of patients in the simulation model would replicate the pattern of patient arrivals which would actually occur. The appeal of the prescribed procedure for generating interarrival times lies in the simplicity and complete reliance on untransformed uniform variates. However, it requires that the arrivals generated per hour of the week are ordered chronologically, which may be time consuming.

As indicated earlier the actual service times were not easy to measure. However, since the simulation was carried out for comparison purposes only, it was considered meaningful to generate random service times from some continuous distributions of the exponential family. Three such distributions were tried namely, negative exponential, lognormal and gamma with various values of the parameters.

The basis of the simulation model is a two-dimensional matrix \( A = (a_{ij})_{1 \leq i \leq n}^{1 \leq j \leq 4} \), where \( n \) is the maximum capacity of the system. For each patient \( i \) in the system, the values of the elements \( a_{i1}, a_{i2}, a_{i3}, \) and \( a_{i4} \) represent the time of arrival, time of initiation of service, time of departure and number of shift during which the arrival occurred.
Initially, each element of the matrix is assigned the value -1. The same value is assigned to the elements of the i-th row of the matrix on the i-th patient's departure. The matrix is then sorted so that the maximum value k such that $a_{kl} > 0$ represents the number of patients in the system. Therefore, the state of each patient in the system is completely identified by the values of the elements of the corresponding row of the matrix. For example, if $a_{i1} > 0$ and $a_{i2} < 0$ then the patient is in the waiting room queue.

The simulation output quantities were computed using these patient vectors. These quantities include the average waiting time, average queue-size and average busy time of the system per physician shift as well as total average quantities (e.g. total average waiting time and total average busy time).

4.6 Implementation of the method

As mentioned earlier, the medical staff of the A/E department of St. Peter's hospital consists of one consultant, two registrars and four SHOs. After subtracting the working hours of the consultant and those of the SHO who is assigned to the subsidiary unit of the department, at Woking Victoria hospital, there are $N=292$ ($M=124$) physician hours which remain to be allocated to the A/E department of St. Peter's hospital.
Table 4.1 shows a feasible set of shifts together with the existing and proposed physician allocation patterns. The latter was obtained by solving problems P1 and P2 using patient arrival data collected over the period from 2 July to 30 September 1984. It is noted that with the above physician-hours to be allocated and with the shift pattern shown in Table 4.1, the conditions of the Lemma are satisfied. By comparing the proposed physician allocation pattern with the average arrival rate per shift, shown in Figure 4.1, it is seen that the advantages (1) and (2) mentioned earlier hold approximately.

The existing and the proposed physician allocation patterns were then compared in terms of the above operating characteristics given as simulation outputs. The simulation model was run with the three service time distributions, mentioned earlier, using various values of their parameters. Three independent runs were made for each distribution using, respectively, three different random number seeds in the random number generator. The same three seeds were used for both the existing and the proposed physician allocation patterns so that the results were comparable. The simulation period for each run was 104 weeks but we collected data for the last 52 weeks only in order to allow the system to 'stabilise'. The internal validity of the model was tested against the analytic (numerical) results for the $M/M/d(t)$ queueing system with time-dependent arrival rate (for simplicity only the harmonic with period one day was considered).
Figures 4.2(a-c), 4.3(a-c) and 4.4(a-c) show plots of the average waiting times per shift, the average queue sizes per shift and the percentages of physician busy times per shift, respectively, for the three service time distributions: negative exponential ($\mu=18.0$), lognormal ($a=e, b=0.59$) and gamma ($a=6.0, b=3.0$). The parameters of these distributions were selected so that the mean service time is 18 mins. Thin (thick) lines in the plots denote results obtained with the existing (proposed) pattern. It is seen that under the proposed pattern the performance of the simulated system is improved considerably. Also, with the above negative exponential, lognormal and gamma distributions the average waiting times were 47.2 and 38.3 mins, 40.8 and 29.2 mins and 38.8 and 28.4 mins under the current and proposed pattern, respectively. The proposed physician allocation pattern gave better results for all other parameter values of the service time distributions used. The improvement was more evident for larger mean values or for the same mean values and larger variances (for the lognormal and gamma distributions). Overall, there is some evidence from the simulation results that implementation of the proposed pattern would lead to considerable improvement of the performance characteristics of the actual system.

When the optimal solution, shown in Table 4.1, was presented to the consultant he thought that, at least initially, no changes should be made during the period from Friday 9am to Tuesday 9am. These additional constraints were then
incorporated in the formulation of problems P1 and P2. The new optimal pattern obtained, suggested that the only changes to be made were to move one physician from shift 6 to shift 9 and from shift 14 to shift 15 (see Table 4.1). The simulation results obtained with the new pattern were still better, than those obtained with the existing pattern, but the difference was small for all service time distributions. For example, the average waiting times obtained with the above neg. exponential, lognormal and gamma distributions were 46.5, 37.6 and 36.3 mins, respectively. The new pattern was implemented in the A/E department of St. Peter's hospital due to its additional advantage, over the existing one, that it is more closely related to the average arrival rate during the corresponding shifts.

4.7 Conclusions

A method is provided for determining the number of physicians to be allocated per shift in an A/E department based on the proportional allocation of the numbers of physicians per hour to the corresponding patient hourly arrival rate. The results obtained suggest that this is a suitable criterion for an A/E department situation.

Implementation of the method requires only routinely collected data and it may be used on a suitable micro computer system. However, it suffers the disadvantage that it relies heavily on the estimates of the patient arrival
Therefore care must be taken to recognise and eliminate or modify those observations (hourly counts of patient arrivals) which are due to non-systematic effects (e.g. disasters or epidemics). In the presence of such effects it might be more appropriate to use statistically robust estimation procedures, such as Winsorisation, in order to estimate the hourly arrival rate than to use the simple arithmetic average.

The method, in conjunction with the simulation model, may also be used in order to examine the effect of changing the number of available physician-hours, provided that the A/E department under study can be approximated by a queueing system model.
Lemma

Let \( l \) denote the minimum of the \( l_i \)s (\( i=1,2,\ldots,k \)). Suppose that

\[
 l_i = \frac{a_i l}{i=1,2,\ldots,k} \tag{1}
\]

where the \( a_i \)s are integers such that

\[
 1 \leq a_i \leq b-1, \quad i=1,2,\ldots,k \tag{2}
\]

Then, a feasible solution to problem P2 exists if and only if \( M = h l \) for some integer \( h \), \( h \geq 0 \).

Proof

Let \( (d_i)_{i=1}^{k} \) denote a feasible solution. Then, from the definition of problem P2, there exist integers \( n_i \) (\( i=1,2,\ldots,k \)) such that \( d_i = l n_i \). From (1) it is easily verified that \( M = h l \), where \( h = \sum_{i=1}^{k} a_i n_i \).

Let \( (n_i^*)_{i=1}^{k} \) be integers, such that \( 0 \leq n_i^* \leq b-1 \), which give the maximum allocation \( M^* \) of the available physician hours (units) to the \( k \) shifts. Then,

\[
 M^* = \sum_{i=1}^{k} \frac{1}{i} n_i^* = \sum_{i=1}^{k} a_i n_i^* \tag{3}
\]

In what follows we shall prove that \( M = M^* \) by contradiction.

Suppose that

\[
 M = M^* + r, \quad r > 0. \tag{4}
\]

From (4) and from the assumption that \( M = h l \) it follows that
\( r = \theta \ell, \theta \text{ integer} > 0 \) \hspace{1cm} (5)

and from (2) we have that

\[ l_i \leq \ell(b-1) \quad i = 1, 2, \ldots, k. \] \hspace{1cm} (6)

Note that \( l_i \) are the units allocated to the shortest shift since, otherwise, \( M^* \) would not give the maximum allocation.

The fact that there are \( r \) unallocated units means that

\[ r < \min \left\{ l_i : n_i^* < b-1 \right\}. \] \hspace{1cm} (7)

Since \( M \leq (b-1) \sum_{i=1}^{k} l_i \), such \( l_i \)s always exist. Consider one of these, \( l_J \) say. Then, from (6), \( l_J \leq \ell(b-1) \). Therefore, by taking out \( l_J \) units from the shortest shift and adding them to the \( J \)-th shift would give

\[ n_J' = n_J^* + l \leq b-1. \] \hspace{1cm} (8)

Now adding \( \ell \), say, units to the shortest shift would result in a remainder of

\[ r' = r - \ell \] \hspace{1cm} (9)

units.

The last result implies that the initial \( n_i^* \)s do not give the maximum allocation as assumed. Therefore, there exist \( (n_i)_{1 \leq i \leq k} \) which give zero remainder in (4). A feasible solution \( (d_i)_{1 \leq i \leq k} \) to problem P2 may then be obtained by taking \( d_i = l_i n_i \) \( (i = 1, 2, \ldots, k) \).
Table 4.1  Current and proposed physician allocation per shift  
(Differences are indicated by an *).

<table>
<thead>
<tr>
<th>Shift no.</th>
<th>Duration</th>
<th>Current allocation</th>
<th>Proposed allocation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Mon 0900-Mon 1300</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>Mon 1300-Mon 1700</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>3</td>
<td>Mon 1700-Mon 2100</td>
<td>2</td>
<td>3*</td>
</tr>
<tr>
<td>4</td>
<td>Mon 2100-Tues 0900</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>Tues 0900-Tues 1300</td>
<td>3</td>
<td>2*</td>
</tr>
<tr>
<td>6</td>
<td>Tues 1300-Tues 1700</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>7</td>
<td>Tues 1700-Tues 2100</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>8</td>
<td>Tues 2100-Wed 0900</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>9</td>
<td>Wed 0900-Wed 1300</td>
<td>3</td>
<td>2*</td>
</tr>
<tr>
<td>10</td>
<td>Wed 1300-Wed 1700</td>
<td>3</td>
<td>2*</td>
</tr>
<tr>
<td>11</td>
<td>Wed 1700-Wed 2100</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>12</td>
<td>Wed 2100-Thurs 0900</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>13</td>
<td>Thurs 0900-Thurs 1300</td>
<td>3</td>
<td>2*</td>
</tr>
<tr>
<td>14</td>
<td>Thurs 1300-Thurs 1700</td>
<td>3</td>
<td>2*</td>
</tr>
<tr>
<td>15</td>
<td>Thurs 1700-Thurs 2100</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>16</td>
<td>Thurs 2100-Fri 0900</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>17</td>
<td>Fri 0900-Fri 1300</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>18</td>
<td>Fri 1300-Fri 1700</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>19</td>
<td>Fri 1700-Fri 2100</td>
<td>2</td>
<td>3*</td>
</tr>
<tr>
<td>20</td>
<td>Fri 2100-Sat 0900</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>21</td>
<td>Sat 0900-Sat 2100</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>22</td>
<td>Sat 2100-Sun 0900</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>23</td>
<td>Sun 0900-Sun 2100</td>
<td>2</td>
<td>3*</td>
</tr>
<tr>
<td>24</td>
<td>Sun 2100-Mon 0900</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>
Figure 4.1  Average patient arrival rates during the periods of physician shifts shown in Table 4.1.
Note:  the patient arrival rate per shift is not constant during a shift period.
Figure 4.2  Average waiting time per shift obtained by simulating the A/E department as a queueing system with time-dependent arrival rate and with neg. exponential, lognormal and gamma service time distributions.

- Under current physician allocation pattern
- Under proposed physician allocation pattern

(a) Neg. exponential (\(\mu=18.0\))
(b) Lognormal (a=e, b=0.59)
(c) Gamma (a=6.0, b=3.0)
Figure 4.3  Average queue-size per shift obtained by simulating the A/E department as a queueing system with time-dependent arrival rate and with neg. exponential, lognormal and gamma service time distributions.

- Under current physician allocation pattern
- Under proposed physician allocation pattern

(a) Neg. exponential ($\mu=18.0$)
Average queue-size

(c) Gamma (a=6.0, b=3.0)

Figure 4.4  Average busy time per shift obtained by simulating the A/E department as a queuing system with time-dependent arrival rate and with two exponential, log-normal and gamma service-time distributions.

Legend:
- Under current physician allocation pattern
- Under proposed physician allocation pattern
(b) Lognormal (a=e, b=0.59)
Percentage busy time

Mon 9am  Tues 9am  Wed 9am  Thurs 9am  Fri 9am  Sat 9am  Sun 9am  Mon 9am

(c) Gamma (\(a=6.0, b=3.0\))
5.1 Introduction

Planning is a matter of deciding how the future pattern of activities should differ from the present; identifying the changes necessary to accomplish this; and specifying how these changes should be brought about[50,54]. Broadly speaking, the steps that comprise systematic hospital planning are as follows: define a reference community and the roles of its hospitals, forecast the future volume of hospital service in that community, transform future service volumes into facility requirements, compare present and required facilities and determine facilities required to accommodate future service volumes; and identify alternative ways to provide, maintain and administer the required facilities (e.g. evaluate total costs of these and select the least costly).

Following the reorganisation of the NHS in 1974 and its restructuring in 1982, the District Health Authority (DHA) is regarded as the basic planning unit[54,118]. The planning process at the DHA level has two strands: the strategic plans
(which look 5-10 years ahead) and the operational plans (which cover a one year Annual Programme and a second year Forward Look). A major responsibility of any hospital in a District is to influence the DHA plans by making recommendations on hospital development needs. A hospital is also responsible for ensuring that its constituent parts (e.g. patient care units) work within the framework of the District's policy and that the management throughout the hospital produces 'good results'.

A dominant theme in the DHA planning process is the determination of the least costly, in terms of both capital and operational costs, arrangement of hospital facilities and services consonant with community acceptability of the service provided. Probably the most expensive hospital facility unit is the (staffed) inpatient bed. Therefore, the determination of the correct number of beds to meet demand placed on each inpatient department of a hospital is of primary importance. An excess of beds results in inflated operating and construction costs, while a bed shortage is unacceptable for a variety of reasons, such as the lack of quality care to the community.

The 'required' number of beds in a hospital inpatient department depends on such factors as the volume of the specific service (i.e. the number of patient days per planning period) to be provided, the pattern of patient demand for each service, the service producing capacity of
the facility (i.e. the output that the facility is capable of providing during the planning period) and the desired bounds of service-standard acceptability.

The simulation model presented in this chapter addresses the facility size stage of the planning process; it is assumed that the volume of a specific service to be provided, the service producing capacity of each inpatient department and the service standard-acceptability have somehow been determined. These factors are determined by the DHA in accordance with Regional Health Authority (RHA) plans and, therefore, involve both quantitative (e.g. demand forecasts) and qualitative (e.g. national policies) components.

More explicitly, the problem we are concerned with in this chapter is: 'given that the DHA wishes to provide a certain number of patient days of a particular inpatient service (e.g. medical, surgical) per annum for its catchment population, how many beds are required to accommodate this volume of service in a cost-effective and community-acceptable way?' In the model developed, cost-effectiveness is represented by bed occupancy rates and community-acceptability is expressed in terms of delays to admission for emergency patients and waiting list lengths. However, the model can easily be modified to incorporate additional or alternative quantitative measures of operating efficiency. We shall take note that measures of 'goodness' are invariably surrogate measures based on value judgements.
The model is (partly) validated using data from the General Medical, General Surgical and Orthopaedics departments of St. Peter's hospital. It is intended to be used for evaluating the adequacy of any bed allocation policy, with various sizes and patterns of patient demand for inpatient care, in terms of an administratively selected set of criteria.

5.2 Review of the literature

As pointed out by Goldman et al[65], 'the bed complement of a hospital is of prime importance since it provides for the presence of the patient for whom all facilities and services are organised'. This importance is reflected by the large number of studies covering all steps of the bed capacity planning process.

In allocating beds to inpatient departments the overall objective is to balance high occupancy rates with the number of beds subject to the constraint that emergency patients are admitted without delay. The key factor for achieving this objective is to reduce the coefficient of variation (ratio of mean over standard deviation) of the daily occupancy. This fact constitutes the background information for most studies on bed capacity planning.

Bailey[15] suggests that the daily occupancy (C) can be
approximated by the product of the average daily admission rate ($R$) times the average length of stay in days ($S$). That is, $C = RxS$. This approximate formula is applicable both for single inpatient departments and for the entire hospital; it becomes exact when the daily bed occupancy follows the Poisson distribution in which case $C$ is the mean of that distribution.

Bailey's formula implies that the variation of the daily occupancy is made up of essentially two components: the variation in daily admissions and the variation in the length of stay. Therefore, by exercising control over admission or discharge procedures, a more stable occupancy can be achieved with all its associated benefits.

On these lines, Young[154,155] presents a thorough discussion of various aspects of inpatient admissions and discharges and conducts an analysis of admissions procedures and inpatient occupancy by employing concepts from the theory of queues. His work extends a measure of admissions control into models published previously by specifying both the bed complement and a critical number $B$ of beds in a given inpatient department, such that scheduled patients could be admitted only when occupancy falls below $B$. When $B$ or more beds are already filled, any empty beds are reserved for emergency admissions only. Young's work comprises one of the earliest attempts to develop a stochastic model allowing administrative control on inpatient admissions other than the
total number of beds. However, when Young evaluated the properties of his B-model using simulation, he found that the simulation results were significantly different from the theoretical ones and, in fact, proved closer to observation than those obtained from the mathematical solution. It turned out that the discrepancy was due to the use of discrete time intervals in the simulation. That is, in the simulation, as in the real world, the scheduled patients all arrive at (about) the same time of day whereas, in the queueing theory model, arrivals are assumed to occur at all times of day or night.

Bithell[23] presents a class of discrete models for the study of the control of inpatient admissions. The models he proposes are based on first order non-homogeneous Markov chains, in order to take account of differences between weekdays, with occupancy rates corresponding to states. These models overcome the difficulties of continuous mathematical models, as described above, by assuming discrete distributions of patient lengths of stay, such as geometric and Pascal. The use of discrete distributions limits the utility of these models since, as has been found in many studies, patient lengths of stay are best fitted by continuous distributions, such as negative exponential, lognormal and gamma.

Kolesar[89] develops a Markovian decision model for hospital admissions scheduling. Describing the state of the system as
the number of beds occupied at the start of a given period \( t \), he introduces a linear programming model that exploits the Markovian structure in providing a basis for determining an optimal control policy for scheduling admissions. The utility of Kolesar's model to the hospital administrator has been severely criticised\[138\]. The objection centres around the oversimplification of the system dynamics, the attempt to apply a Markov decision model to a problem that is essentially non-Markovian and the inherent difficulties involved in gathering realistic data.

Connors[41] proposes an admissions scheduling procedure that employs both probabilistic and deterministic constraints constructed from the hospital's current status and the patients' requirements and characteristics. A major innovation in Connors' work is that an attempt is made to incorporate patients' inconvenience in the model. It is noted in passing that the dehumanised nature of the application of OR to hospitals has been criticised by Stimson and Stimson[140] who comment that 'the patient is rarely mentioned in the studies except as a thing that requires the expenditure of hospital resources'.

Esogbue and Singh[59] develop a stochastic model for determining an optimal distribution of beds in an inpatient department on the assumptions that patients can be classified into emergency and scheduled, that admissions follow Poisson distributions and that patient lengths of stay in the
inpatient department follow the negative exponential distribution. For a fixed number \( N \) of beds in an inpatient department, the authors seek to define an optimal value of occupancy, \( C \), such that if the number of occupied beds is greater than or equal to \( C \), then scheduled patients are not admitted. This model extends the one developed by Young in that it attempts to determine, rather than take as given, the cut-off occupancy level \( C \) and in that it can be easily adapted to treat the situation where emergency patients are temporarily housed in a buffer accommodation.

Milsum et al.\[112\] outline control strategies which facilitate a system's effectiveness, for various organisational attributes, and identify three interacting control variables: bed occupancy, stability of patient flow and patient mix selection. Thus, in contrast to the models reported hitherto, Milsum et al. suggest a holistic approach to the admissions scheduling problem in order to achieve high occupancy levels. Also, the authors survey the major mathematical models and evaluate their use versus the use of simulation. This evaluation led them to conclude that 'the deficiencies of the available theoretical models are so large that simulation studies seem necessary in realistic cases.'

The aforementioned studies are mainly concerned with the day-to-day management of inpatient departments in that they focus on reducing the occupancy variation by exerting some sort of control on non-emergency patient admissions. The
implicit assumption in these studies is that the inpatient care process begins at the time of a patient's admission; other measures of hospital performance such as the length of the waiting list and the time spent by patients on the list are sidestepped. In fact, most of the admissions scheduling models are based on the assumption that there are infinitely long waiting lists so that the desired number of scheduled admissions would occur. In relation to this, Esogbue and Singh[59] argue that '...delays to admission do not seriously inhibit the outcome of inpatient treatment for non-emergency patients'. Although this argument may be valid, to some extent, it assumes that the performance of a health care system is judged upon the outcome of care only, in contrast to the widely accepted[50,62,101] additional measures, such as ease of access to care and the process of care itself (also mentioned in chapters 1 and 2).

The primary concern of Shonick[135,136] is to develop an accurate tool capable of assisting the hospital planner in deciding how many beds are required in each inpatient department to meet pre-determined demands for inpatient care. Shonick assumes Poisson arrivals, defined as requests for admission, and exponential lengths of stay and derives the probability distributions for a number of countervailing criteria-variables (e.g. occupancy, waiting list length and turnaways of emergency patients) for given choices of the number of beds. These distributions are derived using a queueing theory model and it is shown that this model
explains most of the published results of empirical field studies of the behaviour of the daily occupancy. In contrast to Young's B-model, no managerial control of setting a B level has been assumed by Shonick but, instead, the queueing discipline calls for admitting any arrival if an empty bed is available.

Shonick and Jackson[137] publish an extended version of Shonick's earlier model in which a B level could be set. It turns out that Shonick's previous model is a special case of the Shonick and Jackson's model, with B equal to the total bed complement, as is Young's B-model when the waiting list is assumed infinite.

McLain[106] undertakes the task of carrying out a sensitivity analysis of Shonick and Jackson's general model. The author uses simulation to investigate the importance of discrete time and reserved beds on this model and arrives at the conclusion that the errors incurred by using analytic queueing theory approaches 'are serious enough that one should be cautious in applying the results in bed capacity planning whenever any significant probability exists that the facility will be full'.

Hancock et al[73] investigate the effect on hospital occupancy of the number of beds, the percentage of patients who are emergencies, the percentage of non-emergency patients who are scheduled and the average lengths of stay of
emergency and non-emergency patients and they suggest a method for estimating the optimal size of each inpatient department on the basis of expected demand. The method incorporates a simulation-based admissions scheduling procedure, derived in the Bureau of Hospital Administration of the University of Michigan, which allows some patients (call-ins) to be admitted on a very short notice 'in order to restore occupancy'. On the assumption that there is a large backlog of patients on the waiting lists, the authors argue that such patients could always be found and admitted. Hancock et al[74] apply this method to an existing 200-bed adult Medical/Surgical department and show that the occupancy can be improved from 88% to 95% by closing 12 beds.

In addition to the studies reported hitherto, many others are reviewed by Fries[62], by Griffith et al[67], by Shuman et al[138] and by Stimson and Stimson[140].

A major assumption made in the studies on bed capacity planning is that patient arrivals and lengths of stay are independent of time. Shonick[136] avoids this problem by arguing that 'the planner would not wish to adopt as desirable bed complements those which are unnecessarily large because of avoidable day-of-week occupancy fluctuations resulting from administratively faulty admissions and discharge policies'. Hancock et al[73] say that 'there may be very few hospitals which cannot reach the simulation predictions (of the model mentioned above) but only rarely
and with a good reason such as extremely high (greater than 50% of all admissions) emergency arrival rate or an extreme time-dependence in arrivals'. However, it has been shown in many studies that such time effects exist in the patient arrival rates of both scheduled\cite{83,96} and emergency \cite{26,97} patients. Further, in the theoretical studies the assumption that the lengths of stay are exponentially distributed limits the usefulness of the proposed models.

McLain\cite{106} says that if one does not assume that the time-dependence in patient arrivals and lengths of stay should be ignored, then the use of stationary results is called into question and points out that 'it would be of great interest to analyse the impact of time-dependencies on average occupancy'. The simulation model presented in this chapter takes account of the fact that patient arrival rates may be time-dependent.

5.3 Problem formulation

The NHS of the UK may be seen as a hierarchy of systems such as the RHAs, the DHAs and the hospitals. Hospital facilities are then sub-systems of each one of these wider systems, lower down in this hierarchy. Jenkins\cite{81} defines as 'a property of systems' that they form parts of a hierarchy. He also points out that it is the fact that the outputs from one sub-system provide the inputs for other sub-systems that is
responsible for the interaction of the performance of a given sub-system with the performance of other sub-systems at the same level of the hierarchy; hence, Jenkins states, 'the performance of a given sub-system cannot be studied in isolation'.

The A/E department is the hospital sub-system which constitutes the 'free-access' point of the hospital to the general public. It consumes hospital resources both within and outside its boundary, in order to provide patient care, and refers a large proportion of patients to other hospital units (e.g. outpatient clinics, inpatient departments) for follow up treatment. Moreover, since all of these requests for service (including referrals) are urgent, considerable disturbance is caused in the smooth operation of the hospital. Thompson and Webb[144] investigate the effect of the A/E department on inpatient service and conclude that emergency admissions have important consequences for the hospital since they may result in a shortage of beds at a time when there already exists a controversy over hospital waiting lists; they may also result in hindering the hospital in controlling its own resources and their use which, in turn, results in an inability of the hospital to reconcile costs and levels of service. Also, any disturbance in the smooth hospital operation has an obvious and direct impact on the performance of the A/E department itself.

In agreement with Jenkins[81], it was clear that any
improvement in the performance of the A/E department of St. Peter's hospital, with respect to patient flow, could only be achieved as a result of the improved performance of the total hospital system, with respect to the same criterion. These considerations suggested tackling the problem of bed allocation to inpatient departments.

5.3.1 Discussion of the problem

In chapter 3 it was shown that emergency patients who needed admission to inpatient departments of St. Peter's hospital experienced prolonged stays in the A/E department; the argument was put forward that the main reason for this was the inability of inpatient departments to provide a bed at a time a request was issued. Also, the hospital annual statistics indicated that the hospital suffered from low bed-occupancy rates and unduly long waiting lists. More specifically, in 1984 the hospital's inpatient workload was about 19,000 patients of which 7,817 had been admitted as emergencies. During that year, the total average daily number of available beds in the hospital was 592.2 while the total average daily bed occupancy was 400.2. On the 31st of December there was a total of 2241 patients on the waiting lists of the seven inpatient departments which keep such lists.

In view of this situation it was evident that an increase in the number of beds would reduce the waiting list and times to
admission for emergency patients but would worsen the already low occupancy rate in the hospital. Further, as mentioned earlier, any changes in the bed complement is not a matter that concerns the day-to-day management of inpatient departments. Therefore, we decided to treat the problem at the planning (operational or strategic) level. We developed a general 'priority-queueing' simulation model that can be used to assist the hospital planner in determining the appropriate bed complement for each inpatient department on the basis of the expected demand for inpatient care and according to certain measures of hospital performance.

The model is concerned with the community daily demand for inpatient care and takes into account that some patients (scheduled) can wait for admission, whereas others (emergencies) cannot. Therefore, the queueing discipline (admissions and waiting list policy) entails admitting patients according to their priorities, if there are empty beds during a day. The model is general in that it is not tailored to a particular hospital's practices concerning the day-to-day admission and discharge control (e.g. summoning patients at short notice to fill up space capacity in the wards; turning away patients who had been asked to come when the wards are too full; and deciding whether to discharge some patients early if there is, or is expected to be, a large demand for beds). The rationale of the model lies in the view that, for a fixed number of beds, the implicit purpose of exercising these practices is essentially to
approximate a priority queueing discipline on a daily basis: the ideal is that each day no beds of an inpatient department should be empty, while there are patients on the waiting list, and emergency patients should be admitted 'first'. (It is noted that a common day-to-day hospital practice is to set aside a number of beds for emergencies[59,137,154]).

Within this framework, the objective of the simulation model presented in this paper can be clearly stated as follows. 'For each inpatient department determine the number of beds required by the community to meet its hospitalisation demand. The determined numbers should simultaneously satisfy several measures of hospital performance such as high occupancy rates, immediate admission of emergency patients and low waiting lists.' For a fixed number of beds, the model uses the patient day as a unit of measurement of hospital service and evaluates these measures of performance as well as the waiting times for admission of scheduled patients and the 'turnaway' rate for emergency patients. The latter is defined, here, as the proportion of time that emergency patients cannot be admitted immediately due to full occupancy of the inpatient departments concerned. The model may easily be modified to include more than two priority classes so that examination and monitoring of the structure of the waiting list with regard to length of time on the list and the relative seriousness of the cases are possible.

A number of comments are worth making about the problem
stated above.

1. By 'community hospitalisation demand' we mean the demand for each inpatient service created by the hospital's catchment population. The question of estimating need and demand and the relation between the two have been treated by many authors and a number of reviews of the different approaches have appeared in the literature[21,28]. A generally acceptable and reliable methodology for projecting demand in terms of the change of socio-economic and demographic factors has not been developed as yet, although partial attempts to do so have been made[21,46]. However, DHAs produce forecasts of the patient demand for inpatient care for planning purposes.

2. The question of setting levels for the measures of performance, considered in the objective stated, introduces additional difficulty in the planning process although the broad criteria of high occupancy rates (in order to reduce average costs per patient day), rapid admissions of emergency patients, and short waiting time of scheduled patients (and waiting times on the list) are highly desirable. As far as the waiting lists are concerned, it is recognised that they play an important role in applying an admissions scheduling procedure to reduce the occupancy variation, as discussed in section 5.2, and the model should provide information on just how far one can go using the waiting list as a means for reducing the occupancy variation. Excessively long waiting lists are, naturally, not readily acceptable to the community as mentioned by Aldridge[8] and by Duncan and Curnow[58].
In selecting measures of operating efficiency for the inpatient departments, we were aware of the fact that there might not exist a number of beds which satisfies simultaneously all of these measures to the preset levels. This point has been discussed and illustrated by Shonick[136] who also states that 'choosing the bed complement which results in a given occupancy rate, automatically fixes the waiting lists and waiting times as well as the frequency with which the facility is full in the face of arriving emergency patients'. In the light of this fact, we first ran a pilot simulation model to define a suitable range of numbers of beds in terms of the above set of measures (and corresponding desirable levels) of operating efficiency; then, we ran the priority queueing simulation model with each number of beds in the estimated range in order to determine the most appropriate one. This procedure is discussed below.

5.3.2 System modelling

The policy of the NHS allows patients to be admitted to hospitals through either the A/E department or the outpatient clinic of the speciality concerned. While patients may refer themselves or be referred by GPs to the A/E department, patients may only be referred by GPs to the outpatient clinic. At this point a patient's condition is assessed by a specialist physician and the patient is either discharged or put on the waiting list for admission. Since outpatient clinics operate only five days a week no patients are added
to waiting lists at weekends. Figure 5.1 illustrates schematically the patient flow to an inpatient bed.

The patient flow diagram shown in Figure 5.1 can be translated into a number of priority-queueing simulation models. The selection of the most suitable one for modelling a particular system can be determined by studying the characteristics of the system in the light of the symbolic model shown in Figure 5.1. For example, if it is decided that the possibility of scheduled patients (i.e. patients admitted through the outpatient clinics) being admitted as emergencies should be included in the model, the system might be better described by a dynamic-priority queueing model; and if it is decided to include dependencies of the patient lengths of stay on other factors[12,70,96,126] (e.g. clinical and demographic) conditional (i.e. suitably adjusted) length of stay distributions would be required. However, the general methodology would be identical to the one presented in this chapter in which an inpatient department is modelled as a priority queueing system with the following characteristics.

(1) There are two input streams generated by two sources, namely the A/E department (emergency patients) and the outpatient clinic (scheduled patients).

(2) The arrival distributions of these two types of patients are arbitrary (possibly time-dependent). However, the assumption is made that patients of different type arrive independently of each other: scheduled patients are not
allowed by the model to be admitted as emergencies. Also, it is assumed that once a patient has been put on the waiting list he/she cannot renege but would wait to be admitted.

(3) The length of stay distributions are general independent. That is, it is assumed that patient lengths of stay depends only on patient type (i.e. emergency or scheduled).

(4) The priorities assigned to patients are the obvious ones. Emergency patients are assigned the highest. As noted in (2) these priorities are not allowed to change.

(5) There are sufficient ancillary facilities and staff so as not to restrict the flow of patients: the bed complement has direct implications for manpower and medical equipment requirements and all these would be considered together in policy decision making.

The basic unit of measurement of hospital service is the patient day. The sum of patient days over a period of n days divided by n is the average daily occupancy for that period. Same day admissions and discharges of patients are not allowed by the model since such patients are admitted to the day-ward of the hospital.

5.4 Simulation procedure

As indicated earlier, the simulation model seeks to determine the bed complement in inpatient departments on the basis of a sensible balance among the following measures of operating
efficiency.

(1) Emergency patients are admitted without delay.
(2) The occupancy does not fall below a prespecified level at any time.
(3) The length of the waiting list does not exceed a prespecified level at any time.

One way of determining the bed complement is to run the simulation model with various numbers of beds and select the one that gives the best results in terms of the above measures. In order to avoid extensive experimentation with a wide range of numbers of beds, a pilot simulation model could be run first to provide estimates of a shorter range. In this model, the above measures of operating efficiency are included as constraints and the simulation output should provide information as to the most likely range of numbers of beds with regard to constraint satisfaction. The general structure of this model is shown in Figure 5.2.

A typical simulated day is as follows. At the start of the day the beds of those patients who are scheduled to be discharged during the day are freed. Then, the emergency and scheduled patient arrivals are generated from different (in general) distributions. Since constraint (1) must be satisfied, the number of beds is increased, if necessary, so that all emergency patients are admitted. The scheduled patients are put on the waiting list and they are admitted on a FCFS basis. While scheduled patients are admitted the
number of beds is varied so that constraints (2) and (3) are satisfied to the prespecified levels. The scheduled patient admission process continues as long as there are empty beds and patients on the waiting list. Note that the creation or elimination of beds is purely a model concept for estimation purposes. The resultant number of beds, from the daily transactions in the model, is recorded at the end of each simulated day.

Output statistics of the simulation include the frequency distributions of the various values of number of beds, the length of intervals (in days) between changes in the number of beds, the cause of change and the empirical probability of a change when the number of beds has a certain value. These 'empirical' statistics would provide a guide as to the range in which the most appropriate number of beds would lie so that constraints (1), (2) and (3) are satisfied. Each of the values in that range is then assessed by the priority queueing simulation model. The general structure of this model is shown in Figure 5.3.

The arrival mechanism is the same as in the pilot model whereas, now, both types of patients are assumed as joining the queue and are selected for admission according to their priorities. Patients of the same priority are admitted on a FCFS basis. The constraints (1), (2) and (3) above are not included in this model but they are computed as measures of operating efficiency of the simulated system and they are
produced as simulation outputs. The minimum number of beds that results in a desirable balance among those measures can then be selected.

5.5 Data collection and testing

For the simulation model presented in this chapter, the input information required consists of the arrival and length of stay distributions of the two types of patients (emergencies and scheduled). Owing to the fact that the patient record systems in the inpatient departments were completely manual, we resorted to the following data collection procedure. Emergency patient arrival data were collected from the A/E department of the hospital. The data consist of the actual times of patient discharges from the A/E department with direction to the inpatient department concerned. These times were recorded on the casualty cards of the emergency patients and stored in the A/E department's computerised data base from which they were easily retrievable. Scheduled patient arrival data were collected from each of the inpatient departments in the form of weekly counts. Daily counts of scheduled patient arrivals were not available. Finally, data on the lengths of stay of both emergency and scheduled patients were collected, separately for each of the two types of patients, from the patient records of the inpatient departments according to a 50% sampling scheme.
The data refer to the General Medical (MED), the General Surgical (SUR) and the Orthopaedics (ORTH) departments of St. Peter's hospital and were collected over a period of 26 weeks starting from 1st January 1984. In running the simulation model, we assumed that scheduled patient arrivals follow Poisson distributions with daily means 1.4, 4.6 and 2.1 for the MED, SUR and ORTH departments, respectively (the weekly counts supported the Poisson assumption). These numbers were obtained by averaging the weekly counts. At this point it must be noted that the MED department does not keep a waiting list; for convenience we retain the name 'scheduled patients' for patients admitted through the outpatient clinic of this department.

5.5.1 Arrival distributions

In sub-section 5.3.2, it was mentioned that the A/E department is one of the two sources which 'generate' demand for inpatient care. It is current practice that emergency patients may be admitted to inpatient departments of the hospital at any time of the week (i.e. the hospital is always 'on take'). If there are no beds available in the inpatient department concerned, at the time when a demand for a bed is generated, the patient will inevitably have had to be temporarily accommodated in the A/E department. This introduces some bias in the arrival data mentioned above, in the sense that the times of emergency patient arrivals do not necessarily coincide with the times at which requests for
admissions are issued. In the absence of any recorded data on such delays, as discussed in chapter 3, we sidestepped this complication and assumed that emergency patients arrive in an inpatient department after a random length of stay in the A/E department for stabilisation of their condition.

The homogeneous Poisson distribution has been found to be a good fit for emergency patient arrivals in many studies. Therefore, we have considered this distribution as the null hypothesis in testing the pattern of emergency patient arrivals to the three inpatient departments considered.

As a preliminary analysis, each one of the three sets of data was collapsed according to day of week, and the hypothesis of time-homogeneity in the patient arrivals was tested by the dispersion test[43]. Plots of the collapsed data are shown in Figures 5.4(a-c). For the MED, SUR and ORTH arrivals, the values of the dispersion statistic were found to be 30.084, 15.584 and 4.667 which, when compared with the value 12.6 of the $\chi^2$ distribution with 6 d.f. at 0.05 level, led to rejection of the homogeneous Poisson hypothesis for MED and SUR arrivals.

More detailed information about the patient arrival processes can be obtained by basing our analyses on the actual times of arrival. Due to the fact that in the simulation models described in section 5.4, the numbers of patient arrivals are generated at the beginning of each day, we focused our
analysis on the investigation of evolutionary or cyclical trends in the data with respect to time of lunar month or time of week.

In testing for linear trend we used the statistic

\[ U = \sum_{i=1}^{n} \frac{t_i}{n} \sum_{i=1}^{n} \frac{t_i - t_0}{2} \]

suggested by Cox and Lewis[43]. Here, \( t_0 \) is the total period of observation, \( n \) is the number of arrivals in the period \( t_0 \) and the \( t_i \)s denote the actual times of arrival. For MED, SUR and ORTH the absolute values of the U statistic were 0.401, 1.065 and 0.223 which when compared with the value 1.96 of the standard normal distribution suggest no linear trend in the data at the 0.05 level. Potential presence of polynomial trend of higher order was tested using McLean's[107] estimation and testing procedure. Again the data were found to be trend-free.

Next we searched for any cyclical effects in the data. Statistical tests based on the periodogram of the actual times of arrival showed the presence of three harmonic components for MED arrivals and one harmonic component for SUR arrivals (with periods greater than one day). Our previous acceptance of a homogeneous Poisson hypothesis for the ORTH arrivals was supported by this test.

At this point, it must be noted that the periodogram test for
a homogeneous Poisson process is subject to the undesirable effect that individual periodogram components have correlations of the order of \((n+1)^{-1}\), where \(n\) is the number of arrivals in the time period considered. However, since the values of \(n\) for MED and for SUR arrivals were 1328 and 625, respectively, we were not overly concerned about the validity of the test (i.e. the significance of each harmonic could, in effect, be tested independently). A more serious problem arises in the estimation of the parameters of the arrival rate function when the test suggests the presence of more than one harmonic, as was the case for MED arrivals. Although the estimation problem is straightforward mathematically, the parameter estimates are difficult to compute even numerically. This issue has previously been discussed by Cox[42].

The latter difficulty was overcome by estimating the rate function for emergency MED arrivals by a harmonic model based on daily counts. The sufficiently large sample average of the original 182 daily counts, equal to \(m=7.295\), suggested the transformation

\[ y_t = (n_t + c)^{\frac{1}{2}} , \quad t=1, 2, \ldots, 182, \]  

(5.2)

where \(n_t\) denotes the number of arrivals on day \(t\), and \(c\) is a parameter to be chosen. Should the distribution of the original observations \(n_t\) be homogeneous Poisson, the distribution of the \(y_t\)'s would tend rapidly to normal with mean equal to \(m^{\frac{1}{2}}\) and (stabilised) variance equal to 0.25.
For this purpose, the choice of the parameter \( c \) to be 0.386 has been found optimal by Kihlberg et al.\[86\] for \( m \geq 5 \).

For emergency MED arrivals, three harmonic components at frequencies 1, 2, and 2.7 cycles per week were found significant at the 0.05 level, and the daily rate of the non-homogeneous Poisson distribution was estimated to be

\[
\lambda_t = \left[ 2.714 + 1.872 \sin \left( \frac{2\pi t}{7} + 0.526 \right) \\
+ 1.886 \sin \left( \frac{4\pi t}{7} + 0.583 \right) \\
+ 2.042 \sin \left( \frac{5.4\pi t}{7} + 0.773 \right) \right]^2 - 0.386. 
\]

(5.3)

For emergency SUR arrivals, the corresponding rate was estimated by the sample of the actual times of arrival as described by Lewis\[97\]. This rate is given by

\[
\lambda_t = 3.403 \exp \left\{ 0.19 \sin \left( \frac{4\pi t}{7} + 1.136 \right) \right\}. 
\]

(5.4)

Finally, for emergency ORTH arrivals the estimated rate of the homogeneous Poisson distribution is

\[
\lambda_t = 2.885. 
\]

(5.5)

It must be noted at this point that the above harmonics were also found significant when the whole year's (52 weeks) data were analysed. However, as expected, the patient demand for all of the three specialties increased considerably during the summer months.
5.5.2 Length of stay distributions

First we tested the hypothesis that emergency and scheduled patient lengths of stay come from the same distribution. The values of the Kolmogorov-Smirnov two-sample test statistic were 0.078, 0.099 and 0.07 for MED, SUR and ORTH, respectively, corresponding to critical values 0.104, 0.092 and 0.1 at the 0.05 level. Thus, the homogeneity between the length of stay distributions for emergency and scheduled patients was accepted for MED and ORTH and rejected for SUR.

The Kolmogorov-Smirnov one-sample test was then used in order to fit theoretical distributions to the data. For MED and for ORTH, we used the samples formed by combining the individual ones for emergency and scheduled patients. For SUR, the lengths of stay of the emergency and of the scheduled patients were tested separately. Four theoretical distributions were tried, namely lognormal, negative exponential, gamma and Weibull.

At the 0.05 level, the Kolmogorov-Smirnov test accepted the fit of the exponential distribution with parameter 0.156 to the MED lengths of stay. Satisfactory fits were found with the gamma distribution with parameters 1.734 and 4.844 for the emergency SUR lengths of stay, and with the exponential distribution with parameter 0.169 for the scheduled SUR lengths of stay. For the lengths of stay of the ORTH patients, none of the distributions tested gave a
satisfactory fit.

Figures 5.5(a-c) give a comparison of the empirical cumulative distribution functions with the fitted theoretical ones.

5.6 Running the model

The simulation model developed was coded into a FORTRAN 77 computer program using the time-slicing programming technique. The model was run on the VAX 11/780, VMS computer of the RHC. In order to generate random numbers we used the parametric distributions estimated in the previous section, except for ORTH lengths of stay where we had to resort to the empirical distribution. For scheduled patient arrivals, the random numbers were generated from the homogeneous Poisson distributions mentioned earlier. In what follows we discuss our simulation strategy and the results obtained by running the simulation model.

Along with the construction of the simulation model, consideration was given to the problem of selecting a simulation strategy (i.e. method and size of sampling) which would be capable of providing reliable estimates of the required measures of operating efficiency of the simulated system[93,94]. In relation to the simulation strategy, the prime question to answer is whether the simulated system is
stationary in the sense that all of the simulation output processes of interest are stationary (at least in theory). For example, if the bed-occupancy is the only such quantity, stationarity can be reached in configurations with ever-increasing waiting lists irrespectively of the pattern of patient demand (requests for inpatient care). In the simulation model described in this chapter, however, both the length of the waiting list and the emergency-turnaway output processes are also of interest. It follows that the MED and SUR departments are non-stationary; the ORTH department can reach stationarity when the utilisation factor is less than one.

On the other hand, theoretical studies on queues with non-homogeneous and cyclical Poisson arrivals and negative exponential service times have shown that the state probabilities of the system become periodic in the long-run\cite{13,40,91,131}. In turn, this long-run periodicity can play the role of the stationary state in the time independent case. A further long-run queueing theory result, related to bed occupancy, is the approximate relationship

\[
\text{mean no. of beds occupied per day} \approx \text{mean no. of admissions per day} \times \text{mean length of stay in days.}
\]

On the basis of these long-run results it may be argued that in simulations of inpatient departments for bed capacity planning, estimates of the quantities of interest should be
based on long run-length, deletion of some amount of initial output and multiple replications. The latter, are particularly suitable in simulations of non-stationary systems[91,98]. In running the priority queueing simulation model, we used relationship (5.6) as a guide for defining the amount of initial simulation output that would be deleted. Although a simulation strategy which incorporates relationship (5.6) as a criterion is more empirical than formal, it should be noted that such empirical methods have been advocated by many authors on simulation methodology[93,94]. It should also be noted that the method of replication and deletion has been found to perform well in a number of recent studies[85,93,94,98] and is used widely in practical applications[7,83,93]. Its simplicity may also be considered as an advantage.

5.6.1 Model validation

Any model is a simplified representation of reality. Therefore, the validity of the representation should be tested before the results produced by the model are relied upon. The validation process for a simulation model involves running the model with the actual system parameters (for the same time period) and testing whether model results bear a measurable (absolute or relative) relationship to the corresponding results in the real world[88,110,111].

Testing the validity of models for bed capacity planning is a
rather difficult task. Actual hospital figures, corresponding to the quantities measured by the model, are often unavailable\cite{106,138}; those available are mostly 'biased' by the day-to-day admission and discharge procedures\cite{83,138}. It is not surprising, therefore, that the bed capacity planning models are usually 'tested' for reasonableness rather, than validated against actual hospital figures\cite{65,73,74,83,106,136}.

At St. Peter's hospital, the available aggregate figures for each inpatient department, relevant to the simulation model presented in this chapter, were the mean bed occupancy per day; the mean no. of admissions per day; and the mean length of stay in days. The lengths of the waiting lists were only recorded for the 31st December of each year; the turnaway rates for emergency patients were not recorded. Therefore, any attempt to validate the simulation model developed was bound to be limited.

In the light of this fact, we decided to sidestep a validation process based on the existing numbers of beds in the inpatient departments considered. Instead, we ran both the pilot and the priority queueing simulation models to see whether the application of the whole simulation procedure leads to results which show reasonable agreement with the corresponding hospital figures available. However, it should be noted that an extensive validation process, which would hopefully reveal the weaknesses and strengths of the model
developed, is intended to be carried out as part of our further work in this area.

Ten runs were made of each of the two simulation models (pilot and priority queueing), and for each of the inpatient departments considered, using different random number seeds in the random number generator. The initial conditions incorporated in the models were representative of the conditions in the corresponding inpatient departments on the 31st December 1983: in MED, 53 beds were occupied; in SUR, 45 beds were occupied and 179 patients were on the waiting list; and in ORTH, 34 beds were occupied and 214 patients were on the waiting list. The simulation period was 1092 days (three years) but we collected data for the last 364 days only (on the basis of the results obtained from preliminary runs).

In running the pilot model, we incorporated the constraints that emergency patients should be admitted immediately; that no more than 10% of the beds should be empty at any time; and that no more than 150 patients should be on the waiting list at any time. For MED, no waiting list was allowed to form. It is noted that the results obtained from the pilot model can be a valuable guide in determining the appropriate numbers of beds, when the simultaneous satisfaction of several countervailing measures of performance is sought.

Figures 5.6(a-c) show the plots of the cumulative frequency distribution for the numbers of beds in the three inpatient
departments. This distribution is defined as

\[ F_k = \sum_{n=1}^{k} f_n \]  

(5.7)

where \( f_n \) denotes the relative frequency of occurrence of the number \( n \) of beds. Obviously, \( F_k \) gives the proportion of time for which the levels of the measures of operating efficiency incorporated in the model, are satisfied with up to \( k \) beds. The numbers of beds to be subsequently assessed by the priority queueing model were determined as the minimum numbers \( b \) such that

\[ P, \]

where the value of the parameter \( \beta \) was chosen empirically. For SUR and ORTH we selected \( \beta = 0.6 \) giving \( b = 51 \) and \( b = 35 \), respectively. For MED, since no waiting list was kept, we selected \( \beta = 0.8 \) giving \( b = 68 \).

Table 5.1 compares the actual bed complements, the mean number of admissions per day, the mean length of stay in days and the mean number of beds occupied per day for the actual system with the corresponding average results over ten simulation runs. It is clear that the simulation results show close relative relationship with the actual hospital figures. In agreement with 'common-sense' expectations, the results indicate that the occupancy rates are affected considerably by whether or not the simulated inpatient department keeps a waiting list (e.g. the MED department has the lowest occupancy). The occupancy seems to be affected also by the 'smoothness' of the emergency patient arrival process. This result would be anticipated from intuitive grounds in that the number of beds set aside for emergencies
each day could be estimated more accurately for random arrivals. The fact that there is not a large difference between the occupancies in the actual SUR and ORTH departments, indicates that the admission procedure applied in the former department results in 'smooth' patient admission rate. This is an implication of the existence of a long waiting list. These characteristics of the actual systems are reflected by the simulation results.

The constraint on the waiting lists (i.e. less than 150 patients) was satisfied for about 91% and 87% of the time for SUR and ORTH, respectively. A plausible explanation for this can be attributed to the overall higher value of the utilisation factor for the ORTH department. Unfortunately, comparable hospital figures were not available. The delays to admission for emergency patients were always less than 5 hours for SUR and ORTH and less than about 9 hours for MED. Clearly, the latter department will have to operate under even lower occupancy in order to allow that all of its patients are admitted without delay. It may be noted at this point that with the provision of an observation ward in the A/E department (for short-stay cases) it is possible to raise occupancy because the probability of turning an emergency away is greatly reduced. The provision of an observation ward in the A/E department of St. Peter's hospital is the subject of the next chapter.
5.7 Conclusions

On the basis of the results discussed above it may be argued that the simulation model presented in this chapter performs in a manner consistent with the 'real world' operation at the overall level. Should the model be further validated against aggregate hospital figures, with respect to all of the measures of operating efficiency considered, it could be used repeatedly in order to experiment with and gain insight into various bed allocation policies. For a given mix of scheduled and emergency patients various measures of performance, along with their desirable levels, could be administratively determined and their impact on the bed complements in the hospital inpatient departments be evaluated.

Alternative versions of the model presented in this chapter might consider the inclusion of more than one well defined priority class among scheduled patients (e.g. according to severity of case or age); also, upper limits might be set on the length of time spent by patients on the waiting list. Operationally, it might be more useful to place emphasis on the time spent by patients on the waiting list than on the actual length of the list. Future research on these lines could provide fertile areas for substantiating claims for extra hospital resources.
Clearly, selection of the most appropriate model will primarily depend on the purpose of the bed-allocation simulation, be it examination of alternative system configurations and operational rules or study of existing systems under the same operational rules. One point should be made clear however. There are going to be emergency patients needing inpatient care without delay and there are going to be patients put on waiting lists. Hence, it seems reasonable to model inpatient departments as priority queueing systems in order to provide a basis for estimating bed requirements.
Figure 5.1 Possible patient flow paths to hospital inpatient departments.
Figure 5.2  Flow chart of the pilot simulation model.

NBED = no. of beds
NEM = no. of emergency patient arrivals
NSCH = no. of scheduled patient arrivals
NEB = no. of empty beds
Figure 5.3 Flow chart of priority queueing simulation model.

Note: This flow chart is identical to the one shown in Figure 5.2 up to node B.

NWL = no. of patients on waiting list (queue)
Figure 5.4 Collapsed daily plots of emergency patient arrivals to inpatient departments.

(a) General medical
(b) General surgical
Figure 5.5
Theoretical and empirical cumulative distribution functions of patient lengths of stay.

(a) All medical patients (neg. exponential)
(c) Scheduled surgical patients (neg. exponential)
Figure 5.6  Cumulative plots of the relative frequency distributions of daily requirements in beds predicted by the pilot simulation model.

(a) General medical
Cumulative relative frequency

Number of beds

(c) Orthopaedic
<table>
<thead>
<tr>
<th>Actual system</th>
<th>Simulation model</th>
<th>No. of beds</th>
<th>Mean no. of admissions per day</th>
<th>Mean length of stay in days</th>
<th>Mean no. of beds occupied per day</th>
</tr>
</thead>
<tbody>
<tr>
<td>MED</td>
<td>MED</td>
<td>67</td>
<td>8.0</td>
<td>6.4</td>
<td>49.3 (51.0)*</td>
</tr>
<tr>
<td>SUR</td>
<td>SUR</td>
<td>54</td>
<td>6.7</td>
<td>7.3</td>
<td>47.8 (48.6)*</td>
</tr>
<tr>
<td>ORTH</td>
<td>ORTH</td>
<td>40</td>
<td>4.8</td>
<td>7.7</td>
<td>37.7 (37.7)*</td>
</tr>
</tbody>
</table>

+ Averages over ten runs

* Results obtained from relationship (5.6) of text
CHAPTER 6

PLANNING A PROSPECTIVE OBSERVATION WARD IN THE A/E DEPARTMENT

6.1 Introduction

Health care delivery within a hospital consists of the provision of a sequence of services to patients according to their 'care requirements'. These care requirements are determined by physicians' judgements derived from the command of a sufficient body of facts (e.g. symptoms, results of investigations) and from the skill to combine such facts appropriately. It is obvious that unsound decisions of physicians can, at best, result in waste of expensive resources and, at worst, in poor care of patients.

The nature of the work of an A/E department is characterised by three factors which are probably unique in the hospital service: there are no barriers in terms of direct access by the general public; the diversity of symptoms A/E department physicians are called upon to analyse cover, virtually, the whole spectrum of health problems; and medical decisions concerning diagnosis and treatment have, not unfrequently, to be made in an atmosphere of tension and haste. In addition, the rising numbers of patients who attend A/E departments, combined with the lack of purpose-built modern facilities to satisfy the need[82,109,151], make the task of control and
management in these departments even more difficult. This may have important implications for the hospital service: as it has been pointed out [109], 'the character of an acute hospital can be greatly influenced by the presence or absence of an A/E department and its operational policies'.

Some A/E department patients have problems that are probably not urgent but need observation until a diagnosis is clear and safe disposal can be made. From this need arose the concept of observation wards, also known as short-stay beds, observation beds, holding units and short-stay wards. In the absence of such beds, patients are either housed in the main treatment area of the A/E department or admitted to the hospital for observation. In the former case, overcrowding situations are likely to be caused in an area which is designated for the treatment of the seriously ill; in the latter, beds intended for inpatients are occupied with the consequent implications for hospital efficiency.

An observation ward (OW) need not be a separate room in the A/E department; it may well consist of a number of beds which are set aside in the existing premises. In the UK, it has been recommended that the size and function of OWs should be agreed through the A/E sub-committee by taking local conditions into account [82]. However, it is widely believed that patients should not be detained in an OW for more than 24 hours [82,102,109] and if a safe disposal cannot be made after this period, the patient should be admitted to an inpatient department.
The provision of an OW in the A/E department had been included in phase IV of the redevelopment programme of St. Peter's hospital and initial discussions on the site, size and function of the prospective unit had already been held. However the matter had been referred to another meeting at a later date while the consultant of the A/E department had been asked to assess the need for an OW as well as the bed requirements. The objective of the study described in this chapter was to assist the consultant to make documented and appropriate recommendations to the hospital administrator on the prospective unit. More specifically, we sought to provide the consultant with information which could be useful for planning decisions on the OW, within the constraints of time and of the unavailability of historical data. Attention was focused on the nature, size and pattern of the patient demand for the OW, with reference to the total patient population of the A/E department; on the patient lengths of stay; and on occupancy-related quantities.

6.2 Review of the literature

The concept of an OW in an A/E department is not new but very little documentation on their use and operational application is available in the literature.

Pike et al [122] draw attention to the use of short-stay beds in Casualty departments (predecessors of A/E departments) in
order to reduce the number of emergency admissions to inpatient departments. The authors begin their study with an analysis of historical data, obtained from the patient records of the Casualty department of Aberdeen Royal Infirmary, and find that patient arrivals are homogeneous Poisson; that patient lengths of stay (measured in days) are geometric; and that patients suitable for the short-stay ward are selected when first seen by physicians with about 95% accuracy. Pike et al estimate the number of beds required in the short-stay ward of that Casualty department based on the result that the bed occupancy follows the Poisson distribution if the input is Poisson with constant mean, whatever the length of stay distribution provided that it is independent of other factors. This result, proved by the authors for discrete length of stay distributions, corresponds to the well known continuous case one for the M/G/∞ queue in steady state.

Boose[27] carried out a comprehensive investigation of the uses of observation beds in A/E departments in the USA, based on factual information from hospital management publications. Boose lists the most frequently encountered uses of observation beds, such as to avoid unnecessary admissions of A/E department patients, to assist physicians of A/E departments in making sound decisions for or against hospitalisation, and to facilitate the care of patients in need of and awaiting placement in nursing homes or transfer to other health facilities. Boose concludes that observation beds have been used basically to support the care of patients
having non-acute conditions.

Hannan[75] develops a simulation model of the queueing process in the holding unit of the A/E department of a hospital in Springfield, Massachusetts. The model provides a convenient tool for evaluating the effects of changes in demand and utilisation policies upon the congestion level in the unit. The utilisation policies evaluated by Hannan are in essence the ones identified by Boose[27].

The studies by Pike et al[122] and by Hannan[75] were concerned with existing units. Therefore, the authors were able to collect and analyse historical data in order to provide inputs to their models and to test model validity. In contrast, the study described in this chapter deals with the planning of a non-existing unit. Therefore, the operational policy and parameters of the prospective unit were inevitably derived from physicians' judgements, past experience and generally accepted functioning of similar units.

The provision of short-stay beds in A/E departments has been one of the principal recommendations of many reports and investigations in the UK[82,108,109,133,151]. In particular, in the Wessex Region report[151] it is mentioned that 61% of all self-poisoning cases seen in the A/E departments of that Region are admitted to the hospital for observation; it is also pointed out that 'the provision of beds (in A/E departments) will mean that overdose (self-poisoning) cases
requiring observation only, after their initial assessment and treatment, can be managed in the A/E department thereby saving admission to an acute medical bed'.

6.3 Discussion of the problem

All of the alternative options for the first stage of phase IV of the redevelopment programme for St. Peter's hospital included the upgrading of the A/E department[119]. Part of this upgrading was the development of an OW as an integral component of that department.

At the time when the study described in this chapter was initiated, initial discussions on the prospective OW had been held. These covered such issues as the site, size and use of the unit, the construction and operational costs involved and the priority that should be given to the development of the unit within the overall framework of the phase redevelopment programme of the hospital. However, all of these issues had yet to be finally agreed.

The advice of the technical officers was that the unit should be installed as an extension of the existing accident centre (see section 1.3) both because of the fact that OWs are essentially supportive units to the patient care activities of the A/E department and for staffing reasons. It was thought that a unit of about eight beds would be sufficient to cover the needs and it would also be relatively
inexpensive to operate, with respect to fixed costs involved, due to the fact that it would not require expensive equipment or dedicated medical staff (since OWs are not used for acute cases).

The consultant of the A/E department advocated the immediate development of the unit on the grounds that it would serve the multiple purpose of reducing the number of admissions through the A/E department (emergency admissions); of temporarily accommodating patients needing admission when inpatient departments were full; of aiding A/E department physicians to make sound decisions on patients' disposal; of accommodating patients because of social (including psychiatric) or geographic considerations; of preparing patients for operative procedures; and of caring for them afterwards.

An important issue to be settled concerned the number and mix of staff required to ensure proper operation of the unit. The opinion of the nursing officer was that three more nurses would be required in the A/E department in order to cover also the duty-rotas for the OW. The consultant thought that this number could be reduced if the development of the unit was accompanied with the location of the existing nursing station, currently situated in the accident centre, inside the OW so that a nurse would 'keep an eye' on the accommodated patients most of the time.

Although staffing and cost of the unit were not touched upon
directly in this study it was assumed -not unrealistically- that these are closely related to the size of the unit. Therefore, the primary concern of this study was to provide a suitable set of indices for decisions relating to the size (number of beds) of the prospective unit.

6.4 System modelling

Four general approaches can be used in planning the bed capacity of hospital units[67]: the use of normative occupancy levels, the use of empirical occupancy data or similar data on the experience of comparable hospitals, the use of mathematical (mainly queueing) models and the use of simulation.

The major advantage of the first two approaches is that they are simple to use. However, the premium for this simplicity is that no probabilistic statements can be made about the behaviour of the various system parameters of interest. This results in depriving the planner of the flexibility to evaluate the consequences of his choices, with respect to a suitable set of criteria, except at the most superficial level. A thorough discussion on the weaknesses and deficiencies of these approaches is given by Shonick[135,136], by authors in Griffith et al[67] and by Hancock et al[73]. Moreover, these approaches are hardly applicable to units like OWS where the 'patient demand' depends mainly on the operational policy of the A/E
department and on hospital realities. A similar hospital unit to the OW, in this sense, is the pre-discharge ward which is used for accommodating patients during the last few days of their treatment in order to facilitate inpatient departments to meet emergency patient demand. The provision of such units has been suggested by Newell[116].

Since the remaining two approaches are based on queueing concepts, our next step was to identify how queueing would actually occur in the OW. To this end, the consultant's previous experience was of valuable assistance.

The usual practice is that patients are accommodated in the OW after they have been attended to by physicians in the main treatment area of the A/E department. Therefore, strictly speaking, the time of a patient's arrival at the unit coincides with the time the attending physician decides that the patient needs observation before safe disposal can be made. This time coincides also with the beginning of the patient's waiting time until a bed (server) is available in the OW. Once a bed becomes available, the patient is transferred to the OW from the main treatment area of the A/E department and, as a matter of standard policy, is detained in the unit for no more than 24 hours.

Two comments are worth making on this description. First, it seems unlikely that objective data on patient arrivals, as defined above, can be collected even for an existing system. More realistically, the patient arrival process can be assumed
to be triggered at the time a patient is first seen by a physician or at the time of admission to the OW. For planning purposes, the former times could be relied upon since they are routinely recorded on the patient record cards and, therefore, they are readily available. The finding of Pike et al [122] that 'patients suitable for the short-stay ward were selected when first seen by physicians with about 95% accuracy' could also be invoked to justify considering these times. Second, since patients once seen by physicians will somehow receive what might be called 'observation service' if required, it is immaterial to these patients whether this service is going to be provided in the OW or in the main treatment area of the A/E department or in an inpatient bed. Here, it is assumed that physicians will not normally decide to discharge patients whose condition is ambiguous.

The corollary to these comments is that the system (i.e. the observation service) can be adequately represented by an infinite server queueing model. Therefore, by focusing attention on the number of patients in the system or, equivalently, on the number of beds occupied by OW patients (either in the OW or elsewhere) at any time, the selection of any number of beds for the OW can be evaluated in terms of the average occupancy achieved and of the percentage of time patient bed demands are met (or, equivalently, the percentage of patients that can be accommodated in the unit at any time).

Both of these quantities can obviously be calculated by
either an analytic queueing model or a simulation. For the sake of procedural simplicity, which was both desirable and necessary to accept in this study, it was decided to reject the simulation approach in favour of the easier to use analytic infinite server queueing model with homogeneous Poisson arrivals and general-independent lengths of stay, M/G/\infty. The main reasoning behind the selection of this model, as well as the main limitations resulting from the application of the model to the system under study, are discussed below.

6.4.1 Patient arrivals

Patient arrivals to an OW can be viewed as derived demands generated by the number and types of patients attending the A/E department. In other words, the patient population of the A/E department rather, than the population of the health district (catchment area) is the population 'at risk' for the OW.

Information on the patient population of the A/E department (e.g. in terms of age, sex, diagnosis, treatment and pattern of demand) was readily available, since the implementation of the computerised patient record system, from the quantitative systems investigation mentioned in chapter 3. It was now required to obtain information on the possible size, nature, and pattern of the demand for the OW. This could be used as a yardstick for OW planning decisions. To obtain this information, we advocated the method of collecting OW arrival
data prospectively (i.e. through A/E department judgements of the patients' prospective needs for observation). There are three reasons for this. First, there is a great variability in the patient populations and characteristics of A/E departments even in the same health Region\[109,151\]. Therefore, it was thought that there would be a greater risk of error if we based our conclusions on the experience of other A/E departments. Second, there is the concept of clinical freedom. By this we mean that it is recognised throughout the NHS that, within the constraints of the resources available to him, the consultant is responsible for admitting, treating and discharging patients of his department. Third, the subjective element in some aspects of the practice of medicine is very great: as has been observed, different consultants treat similar patients with the same diagnosis in quite different ways\[58,76\]. Of course, the argument is not put forward that any emerged patterns in the prospectively collected data should be assumed unquestionably as standard and repeatable. But, in our opinion, it would be preferable to have some information about the patient demand for the OW of the particular A/E department under study, than no such information.

In collecting patient arrival data, it was agreed with the consultant that the A/E department physicians would mark the record cards of those patients who, in their judgement, would have been accommodated in the OW if one had existed (actual disposals are discussed below). The consultant undertook the tasks of briefing the A/E department physicians on the OW
admission policy, that would have been followed if such a unit had existed, as well as of examining the cards for physician adherence to that policy. The data collection period was fixed a priori at eight weeks (3 September-28 October 1984). Although this period might be judged relatively short for planning purposes, the additional work involved in validating the patient cards and the fact that 'answers' were required relatively quickly made it prohibitive to consider a longer time period.

During the period of eight weeks, 198 patients were recorded as potential OW users out of a total of 5429 patients seen in the A/E department during that period. That is, 3.65% of all patients seen in the A/E department during those eight weeks would have been accommodated in the OW if one had existed. The corresponding percentages reported by Pike et al [122] (based on one-year data) and in the Leeds report [2] (based on two-month data) are 2.38% and 1.4%, respectively.

Figure 6.1 shows a collapsed plot of the patient arrivals by hour of the day; it can be easily seen that there is a significant time of day effect. In estimating the mathematical form of the patient arrival process we carried out a statistical analysis based on the actual times of patient arrivals (i.e. the times patients were first seen by physicians), recorded on the patient cards, using the method described by Lewis [97]. The instantaneous patient arrival rate was estimated to be
\[ \lambda(t) = 0.128 \exp \left\{ 0.759 \sin \left( \frac{2\pi t}{24} + 0.5 \right) \right\} \] (6.1)

On the basis of the harmonic analysis results, presented in chapter 3, it might be argued that such a time-dependent patient arrival rate would have been expected for the actual system. Also, the non-homogeneous Poisson process, with time-of-day dependence, has been found a satisfactory fit in many other studies where patient arrivals are analysed [26,97]. There is strong evidence, therefore, that a non-homogeneous Poisson patient arrival pattern is highly likely to occur in the real system.

Since a major use of an OW is to serve as an aid to physicians for making sound decisions on patient disposals, information on the current disposal of the prospective OW patients would be useful for planning purposes. Table 6.1 shows the disposal patterns for the OW patients, for all patients seen in the A/E department during the data collection period and for all patients seen during the year 1984. No statistical analysis is required to see that the pattern for the OW patients differs from the all-patient patterns, especially with respect to admissions.

Six percent of current emergency admissions to inpatient departments, during the data collection period, would have been accommodated in the OW. This implies that 6% of current emergency admissions to inpatient departments would have possibly been saved if an OW had existed. If one assumes that the same pattern will occur during the operation of the
unit (on an annual basis) and that each of these patients will occupy an inpatient bed for one day, then about 469 (i.e. 6% of the emergency admissions during the whole 1984) inpatient days will be saved for use by other patients. This is an important factor to consider when contemplating the development of an OW in the A/E department. In addition to this classification, the collected data were also classified with respect to other factors (e.g. age, sex, diagnosis) to provide a likely pattern of the future utilisation of the unit.

6.4.2 Patient lengths of stay

For patient lengths of stay, the problem of collecting data prospectively seemed unresolvable since the length of stay in the unit would be a matter of hours and predictions of such accuracy would be unreasonable to expect even from the most experienced physicians. Also, the use of more sophisticated subjective methods (e.g. based on Bayesian forecasting or multiple regression analysis), as described by Gustafson[70] for lengths of stay measures in days, were judged impractical due to the amount of physician time and training required. Notably, Gustafson reported that Bayesian forecasting, which gave the best results in his study, required 30 mins physician time per patient. In any case, the problem of predicting patient lengths of stay is quite difficult on its own and no widely acceptable method exists as yet[58,67].

These facts led us to consider (continuous and stationary)
theoretical distributions as likely representations of patient lengths of stay. Although the actual shape of such a distribution is not particularly relevant in an M/G/∞ representation for bed capacity planning purposes (only the mean is involved), it may be noted that the negative exponential and lognormal distributions have been widely used (either as fitted to actual data or as assumed) in studies on bed capacity planning[59,74,83]. Also, Hannan[75] has observed that these distributions provide good fits of the lengths of stay of OW patients while Pike et al[122] fitted satisfactorily to their data the geometric distribution which is essentially the discrete version of the negative exponential[131].

Two points should be made about the use of theoretical distributions in this study. First, the fact that patients needing observation have more or less similar conditions (in terms of severity) implies that, at least approximately, the patient lengths of stay would be fairly homogeneous. However, it seemed unlikely that patient discharges would occur during certain hours of the day (e.g. early morning hours). Although such situations of time dependent lengths of stay are frequently ignored in deriving analytic tools for use in bed capacity planning[59,135,136,137,155], it is noted here that the results obtained by using the analytic queueing model are subject to this limitation and, therefore, they should be interpreted with caution. For an existing system, a simulation approach might be preferable, if the additional effort required was justifiable, since such a model could
incorporate a greater degree of reality. Second, the consultant mentioned that it was unlikely that patients needing observation would be detained in the unit for less than a certain amount of time (e.g. 4 or 8 hours) or for more than 24 hours. Therefore, it would be most appropriate to stipulate a minimum and a maximum number of hours for the patient lengths of stay. It is noted that the concept of setting a maximum number of hours to lengths of stay for planning purposes has also been used by Hannan [75] and by Duncan and Curnow [58]. In particular, Hannan cites evidence from the experience of three hospitals that, on average, patients stay in the unit for about 60% of the stipulated maximum time.

Statistically, the setting of a minimum and a maximum length of stay implies considering the corresponding truncated versions of the selected distributions. The c.d.f. \( G \) of a truncated distribution at the points \( u \) and \( v \) \((u < v)\) is given by

\[
G(t) = \begin{cases} 
0, & t < u, \\
\frac{F(t) - F(u)}{F(v) - F(u)}, & u \leq t \leq v, \\
0, & t > v,
\end{cases}
\]

(6.2)

where \( F \) is the c.d.f. of the untruncated distribution with the same parameter values. The assumptions underlying the use of truncated distributions are as follows. Starting with a particular distribution governing the lengths of stay of those A/E department patients who need observation in the hospital, it is assumed that short-stay (i.e. less than \( u \)) patients are detained in the A/E department (and then
discharged, perhaps); long-stay (i.e. more than \( v \)) patients are admitted to inpatient departments after they have been accommodated in the OW. In other words, it is assumed that the only observation-patient lengths of stay which are relevant to the OW, are the ones in the central part of the distribution.

6.5 Queueing modelling

In the preceding section it was argued that the observation service through the A/E department could be adequately represented by an infinite server queue and that, under the circumstances of this study, the analytic M/G/\( \infty \) queueing model would be a feasible alternative to simulation. For an M/G/\( \infty \) system in steady state, the number \( N \) of patients in the system follows a Poisson distribution with mean \( r = \lambda w \), where \( \lambda \) denotes the arrival rate and \( w \) denotes the mean length of stay. Therefore, the consequences of selecting any number \( b \) of beds for the unit could be assessed in terms of the mean hourly occupancy \( r \) and of the cumulative probability

\[
P(N \leq b) \geq \beta
\]

(6.3)

where the parameter \( \beta \) is to be chosen by the planner. Roughly speaking, this probability means that the patient bed demands are met for at least \( \beta \% \) of the time or, equivalently, that the bed complement \( b \) is capable of handling \( \beta \% \) of patients on the average.
These criteria were evaluated for various arrival rates and for various mean lengths of stay for each pair of truncation points selected (i.e. the mean value was, obviously, within the range defined by the truncation points). The results were presented in tables which demonstrated the number of beds which would be required to satisfy the hourly bed occupancy at 95% and 99% of all hours; it was borne in mind that the results obtained would underestimate the number of beds required in the unit under the policy of not discharging patients during certain hours of the day. Table 6.2 shows the results obtained with hourly arrival rates 0.10, 0.15,...,0.30 and mean length of stay equal to 16 hours (u=8 and v=24). It is seen that the occupancy increases with the demand, for both 95% and 99% coverage, as expected.

6.5.1 The non-stationary case

In deriving the above mentioned results we have only given consideration to the possible size of the demand, based on the information on patient arrivals derived from the prospectively collected data; the estimated pattern of the demand was ignored. The reasons for this are that the arrival data available were not sufficient to perform forecasts of future demand and that it was expected that the numbers of beds suggested by the stationary model results would not differ substantially from the ones suggested by the non-stationary one, since the unit would be relatively small. However, since the latter reason is a statement of
'expectation', it is worth while examining the impact of the assumption of homogeneous arrival pattern on the average occupancy and on bed shortage. Such an examination is presented below where we assume both the stationary, $M/G/\infty$, and the non-stationary, $M(t)/G/\infty$, models for a particular example. We first present some theoretical results from the literature on the $M(t)/G/\infty$ queues.

Shanbhag[134] has shown that, for the $M(t)/G/\infty$ system with single arrivals, the number $N(t)$ of patients in the system at any time $t$ is a non-homogeneous Poisson process provided that the system is empty initially. This condition is obviously satisfied for a future system. The Poisson mean is given by

$$r(t) = \int_0^t \lambda(x) [1 - H(t-x)] \, dx , \quad (6.4)$$

where $H(t)$ denotes the c.d.f. of the lengths of stay and $\lambda(t)$ is the patient arrival rate.

In the particular case where the patient arrival rate $\lambda(t)$ is a periodic function of time and the distribution of lengths of stay is negative exponential, it has been shown[40,131] that $r(t)$ is also periodic with period equal to that of the arrival rate. In the Appendix of this chapter we show that, for sufficiently large $t$, $r(t)$ is in fact approximately periodic for any (stationary and continuous) distribution of lengths of stay. The practical value of this result is that the variation of the mean occupancy $r(t)$ with time can be adequately described by evaluating $r(t)$ at a finite number of
time points within the daily cycle after convergence has been reached to a satisfactory degree. The probability (6.3) associated with the values of \( r(t) \) at these time points can then be calculated by using the Poisson c.d.f.

The unit of measurement for the OW service is the patient hour when a maximum patient length of stay of about 24 hours is postulated. Therefore, \( r(t) \) was evaluated in this study at each hour of the day (hereafter called 'hourly occupancy level'). The interpretation of the probabilities (6.3), where \( N \) is replaced by \( N(t) \), is that the bed complement \( b \) is capable of handling at least \( \beta \% \) of patients during any hour of the day or, equivalently, that the patient bed demands are met for at least \( \beta \% \) of each hour of the day, when the occupancy is taken on the hour. This interpretation is similar to the one given to 'midnight' occupancy for inpatient departments. The parameter \( \beta \) is called the 'hourly service level'[75] and, in the non-stationary case, denotes also the desired degree of protection against peak periods[67]. In what follows we discuss the non-stationary results (and the corresponding stationary ones) in terms of a specific example.

Consider first the case where patient arrivals occur according to the non-homogeneous Poisson process with rate function given by (6.1) and that lengths of stay follow either the truncated negative exponential (\( \mu=0.0625 \)) or the truncated lognormal (\( a=0.33 \)) distribution at the points \( u=8 \) hours and \( v=24 \) hours. These distributions have means
equal to 14.688 hours and 15.153 hours, respectively, while the corresponding untruncated distributions have equal means of 16 hours.

Figures 6.2(a-b) show the forms of the hourly occupancy level $r(t)$, evaluated numerically at the time points corresponding to each hour of the day, with the above truncated distributions (hereafter called 'exponential' and 'lognormal'). With the exponential distribution, the minimum value of $r(t)$ is 1.597 (on the 22nd hour) and the maximum value is 2.715 (on the 10th hour); with the lognormal distribution the corresponding values of $r(t)$ are 1.629 (on the 23rd hour) and 2.798 (on the 11th hour).

Figures 6.3(a-b) exhibit the cumulative probabilities of the left-hand-side of expression (6.3), associated with the above values of $r(t)$ with the exponential and lognormal distributions, respectively, and for various numbers $b$ of beds. It is seen, for instance, that in order to ensure that at least 95% ($\beta = 0.95$) of the patients would be admitted to the OW when required, irrespectively of the time of day, or that a bed would be available in the OW for at least 57 mins of every hour of the day, when the occupancy is taken on the hour only, six beds should be provided in the unit under both of the above length of stay distributions. On the other hand, when there are six beds in the unit, the hourly occupancy level of day would be within 26.6% and 45.3% with the exponential and within 27.2% and 46.6% with the lognormal.
We now consider the sensitivity of the above results with respect to the homogeneous Poisson distribution with estimated hourly mean equal to 198/1344 = 0.147 and with the above length of stay distributions. The mean hourly occupancy is equal to 2.164 with the exponential and to 2.232 with the lognormal. The deviations of these values from the ones obtained with the non-stationary model are within 1.6% and 26.2% with the exponential and within 0.3% and 27.0% with the lognormal. Also, the stationary model results imply that five beds would be adequate to provide at least 95% coverage for any hour of the day. The resulting average occupancies with five beds would be 43.3% with the exponential and 44.6% with the lognormal. In the non-stationary case, five beds would result in about 94% coverage for the 10th and 11th hours, with the exponential, and for the 9th-13th hours, with the lognormal. It is noted that low occupancies are expected in units where the demand cannot be scheduled and turning away any part of the demand is not a 'admissible' policy.

The above results confirm our prior expectations that the stationary model is a feasible alternative to the non-stationary one, when the latter is a closer representation of reality, for long range planning of small hospital units; and that the small errors involved might not justify the additional computational effort required in the non-stationary case. The results suggest also that the exponential distribution could be used instead of the lognormal if the small errors involved are not judged
significant. A plausible explanation for the closeness of the results obtained with the truncated exponential and lognormal distributions can be attributed to the fact that the corresponding variances are upper bounded by $(u-v)^2/4$. Finally, it may be noted that the non-stationary model, $M(t)/G/\infty$, (or its discrete equivalent) could be useful for short-term bed allocation in (large) inpatient departments according to seasonal fluctuations of the demand.

6.6 Conclusions

In hospital planning, administrators face the problem of determining the least costly arrangement of facilities while achieving acceptable levels of service to the community. This problem requires considering a variety of options and evaluating their consequences according to specific criteria. An obvious option is to provide for alternative and less expensive facilities for the same type and level of service. One of these facilities is the observation ward in the A/E department. The study described in this chapter was concerned with the prospective development of an observation ward in the A/E department St. Peter's hospital; the prospect was to improve the quality of patient care in that department and to reduce the number of emergency admissions to inpatient departments.

Most studies emphasise a certain degree of modelling precision while using historical data to derive the inputs
and to validate their models. In this study, much of the required information was unavailable or impractical to obtain. This necessitated accepting model simplicity in an attempt to provide information which could be incorporated into planning decisions concerning the number of beds in the prospective observation ward in the A/E department. The results of this study were presented by the consultant of the A/E department at a subsequent hospital meeting on the observation ward; the decision was taken to provide for a six bed unit.
APPENDIX

Lemma

If the arrival rate \( \lambda(t) \) is a periodic function of \( t \), period \( T \), and \( H(x) \) is a c.d.f. of a continuous random variable with finite mean \( \mu \), then the function \( r(t) \) given by

\[
  r(t) = \int_0^t \lambda(x) \left[ 1 - H(t-x) \right] \, dx \quad (1)
\]

becomes periodic in the long-run with the same period \( T \).

Proof

Expression (1) implies that 0 \( \leq r(t) \leq \max\{\lambda(x)\} \) \( \mu < \infty \). We shall prove that \( r(t) \) approaches \( r(t+T) \) in the long run.

From (1), we have

\[
  r(t+T) = \int_0^{t+T} \lambda(x) \left[ 1 - H(t+x) \right] \, dx. \quad (2)
\]

Let \( y = x - T \). Then (2) becomes

\[
  r(t+T) = \int_{-T}^{t} \lambda(y+T) \left[ 1 - H(t-y) \right] \, dy = \int_{-T}^{t} \lambda(y) \left[ 1 - H(t-y) \right] \, dy
\]

\[
  = \int_{-T}^{0} \lambda(y) \left[ 1 - H(t-y) \right] \, dy + \int_{0}^{t} \lambda(y) \left[ 1 - H(t-y) \right] \, dy
\]

\[
  = r(t) + \int_{-T}^{0} \lambda(y) \left[ 1 - H(t-y) \right] \, dy.
\]
That is,

$$r(t+T) - r(t) = \int_{-T}^{0} \lambda(y) \left[ 1 - H(t-y) \right] dy > 0. \quad (3)$$

Now, since \( \lim_{t \to \infty} H(t-y) = 1 \), the difference on the left-hand-side of (3) can become arbitrarily small, as was to be proved.
Table 6.1  The disposal patterns for the prospective OW patients arrived between 3 September and 28 October 1984, for all A/E department patients arrived during the same period and for all A/E department patients arrived during the whole 1984.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Discharged</td>
<td>72 (36.4%)</td>
<td>2653 (48.3%)</td>
<td>15311 (46.6%)</td>
</tr>
<tr>
<td>Referred to GP</td>
<td>23 (11.6%)</td>
<td>546 (9.9%)</td>
<td>2439 (7.4%)</td>
</tr>
<tr>
<td>Review clinic</td>
<td>15 ( 7.6%)</td>
<td>853 (15.5%)</td>
<td>5122 (15.6%)</td>
</tr>
<tr>
<td>Outpatients</td>
<td>21 (10.6%)</td>
<td>333 ( 6.1%)</td>
<td>2176 ( 6.6%)</td>
</tr>
<tr>
<td>Admitted</td>
<td>67 (33.8%)</td>
<td>1110 (20.2%)</td>
<td>7817 (23.8%)</td>
</tr>
<tr>
<td>Total</td>
<td>198 (100%)</td>
<td>5495 (100%)</td>
<td>32865 (100%)</td>
</tr>
</tbody>
</table>

* includes also modes of disposal which did not appear for prospective OW patients.
Table 6.2  Bed requirements in the observation ward to satisfy 95% and 99% of the hourly patient demand, and resultant percentage bed occupancies, for various arrival rates and for mean length of stay equal to 16 hours.

<table>
<thead>
<tr>
<th>Arrival rate/hour</th>
<th>0.10</th>
<th>0.15</th>
<th>0.20</th>
<th>0.25</th>
<th>0.30</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of beds (95%)</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
</tr>
<tr>
<td>Percentage occupancy</td>
<td>40.0%</td>
<td>48.0%</td>
<td>53.3%</td>
<td>57.1%</td>
<td>60.0%</td>
</tr>
<tr>
<td>No. of beds (99%)</td>
<td>5</td>
<td>6</td>
<td>8</td>
<td>8</td>
<td>9</td>
</tr>
<tr>
<td>Percentage occupancy</td>
<td>32.0%</td>
<td>40.0%</td>
<td>40.0%</td>
<td>50.0%</td>
<td>53.3%</td>
</tr>
</tbody>
</table>
Figure 6.1 Collapsed hourly plot of arrivals of patients who were selected as suitable for the prospective observation ward in the A/E department of St. Peter's hospital during an eight week period.
Figure 6.2 Time-dependent mean number of patients in the system (queue-size).

(a) Neg. exponential distribution of lengths of stay ($\mu = 0.0265$) truncated at the points $u=8$ hours and $v=24$ hours.
(b) Lognormal distribution of lengths of stay \((a=e, b=0.33)\) truncated at the points \(u=8\) hours and \(v=24\) hours.
Figure 6.3  Time-dependent cumulative probability that the number of patients in the system does not exceed b (b=1, 2, 3, 4, 5, 6).
* Stationary probabilities (λ=0.147) are given in parentheses.
(a) Neg. exponential distribution of lengths of stay (μ=0.0265) truncated at the points u=8 hours and v=24 hours.
DISCUSSION

The National Health Service (NHS) is an organisation which is by any standard an important component of the British economy. The NHS employs about 1 million people and has a budget of approximately £14,500m, representing over 6% of the Gross Domestic Product or some 11% of the total public expenditure. In England, the NHS is administered as a hierarchy involving 14 Regional Health Authorities (RHAs) and some 200 District Health Authorities (DHAs).

In recent years several official reports and individual research studies have been published with a primary concern to develop methods for improved efficiency within the NHS[11,51,52,53,54]. The kernel of these studies is that improved efficiency derives from effective planning and management which, in turn, are inextricably linked to the availability and use of pertinent and accurate information. Particular emphasis has been placed on planning at the District level and recommendations have been made towards assisting DHAs in planning and monitoring hospital inpatient and outpatient services[53,54,55,56].

Hospital resource allocation policies are formulated by DHAs to provide the facilities demanded by the general public (as defined by professionals) in accordance with overall RHA guidelines (derived from RHA plans). Among such facilities,
the A/E department of the General District Hospital is of acknowledged importance for the community health. This thesis has been based on the view that resource allocation and management within the A/E department affects to a substantial degree the managerial effectiveness within the hospital. In turn, effective management in the A/E department demands effective information services.

It has been pointed out[67] that 'the success of health care plans must be measured in terms of the impact of the health care system on individual patients'. It follows that each facility should have available the information required to estimate resource demands. Without a good information system (or patient record system as called in this thesis to distinguish between information at the facility level and information at the hospital or wider system level) it is virtually impossible for the consultant of the A/E department to establish managerial policies, make reasoned and informed management decisions or recommendations, motivate his staff and evaluate performance; in fact do any of those things advocated in the Griffiths report[56]. In other words, such a system will provide the opportunity to shift from intuition to a willingness 'to let the data speak'; from reliance on 'off-the-shelf' functions to reliance on functions developed from data and validated in the range of the A/E department. Chawner[37] draws attention to the consultant's role in NHS management and argues that 'the consultant should be cost conscious'. The Korner report[53,55] appreciates that accuracy, relevance and timeliness are more likely to be
achieved if, wherever possible, information is generated from
data collected at facility levels for operational purposes;
it also attempts, in considerable detail, to define the
information requirements at various levels.

The primary purpose of the research described in this thesis
was to provide information for policy formulation on certain
aspects of the resource allocation and management process in
a DGH. The focus of attention was the hospital's A/E
department; the function of the A/E department was seen in
the framework of the total hospital system; and the approach
followed was based on the view that 'description must precede
prescription'. The argument is not advanced that, in
attempting to provide this information, we achieved as high a
degree of objectivity as a detached observer might find
desirable. The main reasons for this have been made explicit
throughout this thesis.

In conducting these studies we had been aware of the fact
that resource allocation decisions within the NHS are not
based on quantitative information only; there is a
substantial qualitative component involved. The NHS exists
to meet needs as defined by professionals (human judgemental
process), not demands as expressed by the general public. In
turn, the definition of whose needs are going to get priority
in conditions of scarcity is, to a large extent, a political
process including professions, pressure groups and parties.
Within the overall framework set by such qualitative factors,
however, only suitable analysis in quantitative terms can
provide estimates of the consequences of changes which might be made in the resource allocation process within a hospital.
REFERENCES


82. JOINT CONSULTANTS COMMITTEE (1978). Medical staffing of accident and emergency services.


118. NORTH WEST SURREY DISTRICT HEALTH AUTHORITY (1983). Division of responsibilities and functions between Districts and Units.


Appendix: Doctors to Shifts in an Accident and Emergency Department

APPENDIX

PUBLISHED MATERIAL
Allocating Doctors to Shifts in an Accident and Emergency Department

G. VASSILACOPOULOS
Department of Mathematics, Royal Holloway College, University of London

A method is proposed for allocating doctors to weekly shifts in an accident and emergency department of a hospital. Two models, solved by dynamic programming, are used. The first finds the optimal solution to the problem of allocating doctors per hour of the week proportionally to the corresponding patient arrival-rate; and the second uses this solution to smooth out the hourly discrepancies in the number of doctors within a shift. The solution is then assessed by computer simulation.

INTRODUCTION

Accident and emergency (A/E) departments in the United Kingdom operate 24 hours a day, 7 days a week. At least one doctor must be present at any time, and all arriving patients are seen by doctors. Owing to time dependency of the patient arrival-rate, the problem is to allocate doctors to shifts so that the patient queue-sizes and waiting-times in the A/E department's waiting room are minimized.

The general problem of allocating manpower to shifts has been studied by many authors in the past. As an example, Baker studies the cyclic requirements for full-time and part-time staff in order to minimize a cost function; Kolesar et al. determine the number of police patrol cars required to meet demand for service by using queueing theory results as input to an integer-programming problem; and Zanakis and Lawrence propose a mixed integer-programming model for shift allocation of personnel in a military environment.

In the health operational research literature, we cite the studies by Maler-Rothe and Wolfe; and by Hershey et al., both dealing with scheduling of nursing staff. Reference to the problem of doctor allocation in A/E departments is made by Bolling and more recently by Carlsm et al.

In this paper, a method is proposed for determining the number of doctors per shift (on a weekly basis) which may be used routinely as part of the managerial planning process in an A/E department. An overview of the method is given below.

(1) Find a mathematical expression for the patient hourly arrival-rate.
(2) Determine, by dynamic-programming, the optimal number of doctors to be allocated per hour of the week so that the allocation is proportional to the corresponding arrival rate.
(3) Use the optimal solution obtained at step (2) to find, by dynamic programming, the optimal number of doctors to be allocated per shift.

The method has been implemented in the A/E department of St Peter's Hospital, Chertsey, Surrey, England. The data required for step (1) were obtained from its patients' records. The dynamic-programming algorithms are fast enough to be run in a suitable microcomputer system. Such systems are currently being installed in many A/E departments in the United Kingdom.

DISCUSSION ON THE PROBLEM

The patient arrival-process

The patient hourly arrival-rate in the A/E department concerned varies considerably with the hour of day and day of week. Vassilacopoulos used cross-classified data analysis and spectral analysis to study the effect of these factors upon the numbers of hourly and daily patient visits, for data collected over the summer of 1983. It was found that the patient hourly arrival-rate could be approximated by a Fourier series consisting of three harmonics, with periods 24, 12 and 8 h. A trend component, due mainly to a cyclical day-of-week effect, was also found to be statistically significant.
Subsequently, by analysing other data collected over homogeneous time periods (e.g. winter), it was also found that the hour of day and day of week effects explain a very large proportion of the variation in the numbers of patient arrivals. Therefore, it may be concluded that the patient arrival-rate for each hour of the week is constant during a homogeneous time period. However, should other effects (e.g. month) be found significant, they would have to be taken into account in the estimation of the arrival rate.

Modelling the A/E department as a queueing system

A theoretically similar system to an A/E department, namely the New York City Police Department, was modelled by Kolesar et al. as an $M/M/m(t)$ queueing system with time-dependent parameters. The required number of police patrol cars per hour was found by solving the resulting transient state difference-differential equations numerically; integer programming was then used to determine the required number of police patrol cars per shift.

Observation of the operation of the A/E department of St Peter's Hospital, however, led us to the conclusion that queueing-system modelling would only marginally represent reality. In support of this argument, we cite two reasons:

(1) There is no clear-cut queueing discipline in the system. Patients are, in principle, seen on a first-come first-served basis. The one official class of priority cases (such as those needing resuscitation) may be ignored at this point since they represent a very small proportion of the total workload. However, other patients may be given priorities to service unofficially (e.g. young children, old people and pregnant women, irrespective of the severity of their complaint, and other patients according to the apparent nature and severity of their complaint). Further, these priorities are not clearly ordered.

(2) The beginning of service to one patient does not necessarily coincide with another patient's departure. This is a very usual situation since patients may be temporarily accommodated in beds and seen periodically by a doctor. This fact complicates the calculation of the actual consultation (service) time of a patient. It was also noted that a patient may be attended to by more than one doctor at a time (e.g. severe cases).

Despite the above reasons, which complicate the modelling of an A/E department as a queueing system, the problem of allocating doctors to shifts is of a 'queueing nature'. This was clearly stated by the Consultant in charge of the A/E department as follows:

"Allocate the number of doctors to pre-specified shifts so that the overall patient waiting-times and the patient queue-sizes at any time $t$ are minimized. The solution should take into account that at least one doctor must be present at any time in the A/E department and that there is a limited number of doctor-hours available per week".

TACKLING THE PROBLEM

Doctor allocation per hour of week

The problem, as just stated, may be transformed to a discrete one if it is assumed that the numbers of doctors ($n_i$) may change on the hour only and that the patient arrival-rate ($\lambda_i$) is constant for each hour $t$ of the week over the period considered.

We first define the following notation:

$N = \text{number of available doctor-hours per week,}$

$M = N - 168 = \text{number of available doctor-hours after allocating one doctor per hour of week,}$

$r_i = n_i - 1 = \text{number of doctors in excess of one, allocated to the } t\text{th hour of the week} (t = 1, 2, \ldots, 168),$  

$b = \text{number of available doctors.}$

In order to allocate doctors per hour of the week approximately proportionally to the corresponding patient arrival-rate, we first determine the (real) numbers, $q_i$, by

$$q_i = \frac{M}{\sum_{i=1}^{168} \lambda_i} \lambda_i = \frac{M}{\sum_{i=1}^{168} \lambda_i} \lambda_i, \quad t = 1, 2, \ldots, 168,$$  \hspace{1cm} (1)
where $\lambda$ is assumed to be constant following the previous discussion on the patient arrival-process. Then, the integer numbers, $r_i$, are obtained by solving the following problem by dynamic programming:

$$\text{minimize } f_i = \max_{1 \leq t \leq 168} \{ |q_t - r_i| \}$$

subject to

$$P1: \sum_{i=1}^{168} r_i = M$$

$$0 \leq r_i \leq b - 1$$

$r_i$ integers.

Selection of this criterion for allocating doctors per hour of the week has certain advantages, which hold, within the constraints of the problem, under the fairly 'loose' condition that the weekly patient arrival-rate is constant over the period considered.

(1) It is suggested that the expected number of arriving patients corresponding to each doctor on duty is approximately constant throughout a week, which might be judged desirable from the doctor's point of view.

(2) It is ensured that the number of doctors present in the A/E department at any time of the week is closely related (in phase) to the patient arrival-rate.

(3) For a queueing situation, the utilization factor per hour of the week, defined by the expression

$$\rho_t = \frac{\lambda_i \mu}{n_i}, \quad t = 1, 2, \ldots, 168,$$

is kept approximately constant throughout a week, provided that the service rate ($\mu$) is constant. This is expected to reduce unduly long queues during those hours of the week when the department tends to be very busy.

Expression (4) might also be chosen as the criterion to minimize in order to determine the number of doctors to be allocated per hour. For our data, it was found that at least one of the optimal solutions suggested by (2) coincided with the optimal solution given by minimizing $\lambda_i/n_i$ for every hour of the week.

**Doctor allocation per shift**

Let $S = \{s_1, s_2, \ldots, s_k\}$ denote a feasible set of the hours of the week when the number of doctors present in the A/E department may change. This set is provided by the Consultant in charge of the A/E department.

We introduce the following additional notation:

- $l_t = s_t - s_{t-1}$; $t = 1, 2, \ldots, k$ = time period within which the number of doctors remains unchanged (shift).
- $d_i = \text{number of doctor-hours in excess of } l_i, \text{ allocated to time period } l_i; i = 1, 2, \ldots, k$ ($d_i$ divisible by $l_i$).

Minimize

$$f_2 = \max_{1 \leq i \leq k} \left\{ \left| d_i - \sum_{j=t-1}^{t} r_j \right| \right\}$$

subject to

$$P2: \sum_{i=1}^{k} d_i = M$$

$$0 \leq d_i \leq (b - 1)l_i$$

$d_i$ integers divisible by $l_i$.

It is easily observed that problem $P2$ may not have a feasible solution. Necessary and sufficient conditions for the existence of a feasible solution to problem $P2$ are given in the Lemma which is stated and proved in the Appendix. For the data used in this study these conditions are satisfied.
Computationally, end-effects are avoided by letting $s_0 \geq 168$, depending on the choice of $s_0$. It must be noted, in passing, that there may be more than one optimal solution to problems $P_1$ and $P_2$. It is, therefore, instructive to test the effect of implementing each of those solutions. A rough assessment could be obtained by simulating the A/E department as a queueing system $M/G/m(\lambda)$ with time-dependent arrival rate. The simulation results for an optimal solution are given below.

**IMPLEMENTATION OF THE METHOD**

The medical staff of the A/E department of St Peter's Hospital, Chertsey consists of one Consultant Surgeon, two Registrars and four Senior House Officers. The department has a subsidiary unit at Working Victoria Hospital, which is assigned one of the four Senior House Officers from 9 a.m. to 5 p.m. Mondays to Fridays only. After subtracting these doctor-hours, as well as those of the Consultant, there are 292 doctor-hours remaining to be allocated to the A/E department of St Peter's Hospital.

Table 1 shows a feasible set of shifts, together with the current and recommended doctor allocation patterns. By comparing the recommended doctor allocation patterns with the average arrival rate per shift, shown in Figure 1, it is seen that the advantages (1) and (2) mentioned earlier hold approximately.

The third, as well as a validation of the recommended solution for a queueing situation, is tested by simulating the A/E department as an $M/G/m(\lambda)$ queueing system with time-dependent arrival rate.

A time-dependent Poisson distribution of patient arrivals was incorporated into the simulation model as interarrival times, and the parameter of this distribution was changed during each hour (60 min) of simulation time. By this procedure were reasonably assured, and the results confirmed it, that the arrival pattern of patients in the simulation model would replicate the pattern of patient arrivals which would actually occur.

As indicated earlier, the actual consultation (service) times were not easy to measure. However, since the simulation was carried out for comparison purposes only, it was considered meaningful to generate random service times from a continuous distribution of the exponential family. Three such distributions were tried, namely negative exponential, log-normal and gamma, with various values of the parameters.

The simulation program was written in FORTRAN 77, and it was run with both the current and the recommended doctor allocation patterns for the shifts shown in Table 1.

<table>
<thead>
<tr>
<th>Shift number</th>
<th>Duration</th>
<th>Current allocation</th>
<th>Recommended allocation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Mon. 0900-Mon. 1300</td>
<td>3 3</td>
<td>3 3</td>
</tr>
<tr>
<td>2</td>
<td>Mon. 1300-Mon. 1700</td>
<td>3 3</td>
<td>3 3</td>
</tr>
<tr>
<td>3</td>
<td>Mon. 1700-Mon. 2100</td>
<td>2 3*</td>
<td>3 3</td>
</tr>
<tr>
<td>4</td>
<td>Mon. 2100-Tues. 0900</td>
<td>1 1</td>
<td>1 1</td>
</tr>
<tr>
<td>5</td>
<td>Tues. 0900-Tues. 1300</td>
<td>3 3</td>
<td>3 3</td>
</tr>
<tr>
<td>6</td>
<td>Tues. 1300-Tues. 1700</td>
<td>3 2*</td>
<td>2 2</td>
</tr>
<tr>
<td>7</td>
<td>Tues. 1700-Tues. 2100</td>
<td>2 3*</td>
<td>3 3</td>
</tr>
<tr>
<td>8</td>
<td>Tues. 2100-Wed. 0900</td>
<td>1 1</td>
<td>1 1</td>
</tr>
<tr>
<td>9</td>
<td>Wed. 0900-Wed. 1300</td>
<td>3 2*</td>
<td>3 2*</td>
</tr>
<tr>
<td>10</td>
<td>Wed. 1300-Wed. 1700</td>
<td>3 2*</td>
<td>2 2</td>
</tr>
<tr>
<td>11</td>
<td>Wed. 1700-Wed. 2100</td>
<td>2 2</td>
<td>2 2</td>
</tr>
<tr>
<td>12</td>
<td>Wed. 2100-Thurs. 0900</td>
<td>1 1</td>
<td>1 1</td>
</tr>
<tr>
<td>13</td>
<td>Thurs. 0900-Thurs. 1300</td>
<td>3 3</td>
<td>3 3</td>
</tr>
<tr>
<td>14</td>
<td>Thurs. 1300-Thurs. 1700</td>
<td>3 2*</td>
<td>2 2</td>
</tr>
<tr>
<td>15</td>
<td>Thurs. 1700-Thurs. 2100</td>
<td>2 2</td>
<td>2 2</td>
</tr>
<tr>
<td>16</td>
<td>Thurs. 2100-Fri. 0900</td>
<td>1 1</td>
<td>1 1</td>
</tr>
<tr>
<td>17</td>
<td>Fri. 0900-Fri. 1300</td>
<td>3 2*</td>
<td>2 2</td>
</tr>
<tr>
<td>18</td>
<td>Fri. 1300-Fri. 1700</td>
<td>3 3</td>
<td>3 3</td>
</tr>
<tr>
<td>19</td>
<td>Fri. 1700-Fri. 2100</td>
<td>2 2</td>
<td>2 2</td>
</tr>
<tr>
<td>20</td>
<td>Fri. 2100-Sat. 0900</td>
<td>1 1</td>
<td>1 1</td>
</tr>
<tr>
<td>21</td>
<td>Sat. 0900-Sat. 2100</td>
<td>2 2</td>
<td>2 2</td>
</tr>
<tr>
<td>22</td>
<td>Sat. 2100-Sun. 0900</td>
<td>1 1</td>
<td>1 1</td>
</tr>
<tr>
<td>23</td>
<td>Sun. 0900-Sun. 2100</td>
<td>3 3*</td>
<td>3 3*</td>
</tr>
<tr>
<td>24</td>
<td>Sun. 2100-Mon. 0900</td>
<td>1 1</td>
<td>1 1</td>
</tr>
</tbody>
</table>
G. Vassilacopoulos—Doctor Shifts in an A/E Department

Figure 1. Average arrival rates per shift (the breakdown of shifts is shown in Table 1). Note: the patient arrival-rate per shift is not constant during a shift period.

Figure 2 shows the average queue sizes per shift, rounded off to the nearest integer, obtained with the current (thin line) and with the recommended (thick line) doctor allocation patterns. The random service times were generated from the negative exponential distribution with mean 25 min. The program was run for 100 weeks simulation time (expressed in minutes). It is seen that under the recommended pattern, the queue sizes are flatter and are reduced overall. Also, with the current pattern, the patient waiting-times had sample mean 77.1 min and sample standard deviation 121.7 min, whereas with the recommended pattern, the corresponding values were 67.1 and 104.8 minutes. The simulation algorithm, as well as formulae for the operational characteristics of the system, are described in detail in Vassilacopoulos.11

With the recommended doctor allocation pattern, the simulation results obtained were better for all other service time distributions used, irrespective of the parameter values. However, when the mean service time was taken to be more than 30 min, some queue sizes and the overall waiting times were found unrealistically large (e.g. for mean service time equal to 30 min, the average waiting times were 232.8 and 173.9 with the current and recommended pattern, respectively). For mean

Fig. 2. Average queue sizes per shift obtained by simulating the A/E department as a queueing system $M/M/m(t)$ with time-dependent arrival rate (the breakdown of shifts is shown in Table 1). Under current doctor allocation pattern. Under recommended doctor allocation pattern.
service time as small as 10 min, the recommended pattern was still better but the difference was very small.

When the optimal solution shown in Table 1 was presented to the Consultant Surgeon in charge of the A/E department, he thought that, at least initially, no changes should be made during the period from Friday to Monday inclusive. These additional constraints were then incorporated in the formulation of problems \( P1 \) and \( P2 \). The new optimal pattern obtained suggested that the only changes to be made were to shift one doctor from shift 6 to shift 7 and from shift 14 to shift 15 (see Table 1). The simulation results obtained with the new pattern were still better than those obtained with the current pattern, but the difference was small for all service time distributions (e.g. for negative exponential service times with mean 25 min, the average waiting time was 74 min and the standard deviation 117.6 min). The new pattern was implemented in the A/E department of St Peter's Hospital owing to its additional advantage, over the current one, that the number of doctors allocated is more closely related to the average arrival rate during the corresponding shifts.

CONCLUSIONS

A method is provided for determining the number of doctors to be allocated per shift in an A/E department, based on the proportional allocation of the numbers of doctors per hour to the corresponding patient hourly arrival-rate. The results obtained suggest that this is a suitable criterion for an A/E department situation.

Implementation of the method requires only routinely collected data, and it may be used on a microcomputer system. However, it suffers the disadvantage that it relies heavily on the estimates of the patient arrival-rate. Therefore care must be taken to recognize and eliminate or modify those observations (hourly counts of patient arrivals), which are due to non-systematic effects (e.g. disasters or epidemics). To this end, the use of robust estimation procedures such as Winsorization might be more appropriate than the simple arithmetic average.

The method, in conjunction with the simulation model, may also be used in order to examine the effect of changing the number of available doctor-hours, provided that the A/E department under study behaves approximately as a queueing system.

APPENDIX

Define the following quantities:

\[ M = \text{number of available doctor-hours (units) after allocating one doctor per hour of the week} \quad (M > 0), \]
\[ b = \text{number of available doctors}, \]
\[ k = \text{number of shifts per week}, \]
\[ l_i = \text{length (in hours) of the} \ i \text{th shift} \quad (i = 1, 2, \ldots, k). \]

**Lemma**

Let \( l \) denote the minimum of the \( l_i \) \( (i = 1, 2, \ldots, k) \).

Suppose that

\[ l_i = a_i l, \quad i = 1, 2, \ldots, k, \quad (1) \]

where the \( a_i \) are integers such that

\[ 1 \leq a_i \leq b - 1, \quad i = 1, 2, \ldots, k. \quad (2) \]

Then, a feasible solution to problem \( P2 \) exists if and only if \( M = h l \) for some integer \( h, h > 0 \).

**Proof**

Let \( (d_i)_{1 \leq i \leq k} \) denote a feasible solution. Then, from the definition of problem \( P2 \), there exist integers \( n_i \) \( (i = 1, 2, \ldots, k) \) such that

\[ d_i = l n_i. \]

Now, from (1) it is easily verified that

\[ M = h l, \quad \text{with} \quad h = \sum_{i=1}^{k} a_i n_i. \]

522
Let \((n^*_i)_{i=1}^k\) be integers, such that \(0 \leq n^*_i \leq b - 1\), which give the maximum allocation \(M^*\) of the available units to the \(k\) shifts. Then,

\[
M^* = \sum_{i=1}^k l_i n^*_i = \sum_{i=1}^k a n^*_i.
\]

In what follows we shall prove that \(M = M^*\) by contradiction.

Suppose that

\[
M = M^* + r, \quad r > 0.
\]

Then, from (4) and the assumption that \(M = h l\), it follows that

\[
r = \theta l, \quad \theta > 0, \quad \text{integer}
\]

and from (2) we have that

\[
l_i \leq l (b - 1) \quad i = 1, 2, \ldots, k.
\]

Note that \(l (b - 1)\) are the units allocated to the shortest shift since, otherwise, \(M^*\) would not give the maximum allocation.

Now, the fact that there are \(r\) unallocated units means that

\[
r < \min \{i; n^*_i < b - 1\}.
\]

Since

\[
M \leq (b - 1) \sum_{i=1}^k l_i,
\]

such \(l_i\)s always exist. Consider one of these, \(l_j\) say. Then, from (6), \(l_j \leq l (b - 1)\). Therefore, taking out \(l_j\) units from the shortest shift and adding them to the \(j\)th shift would give

\[
n^*_j = n^*_j + 1 \leq b - 1.
\]

Now adding \(l\), say, units to the shortest shift would result in a remainder of

\[
r' = r - l
\]

units.

The last result implies that the initial \(n^*_i\)s do not give the maximum allocation as assumed. Therefore, there exist \((n_i)_{i=1}^k\) which give zero remainder in (4).

A feasible solution \((d_i)_{i=1}^k\) to problem \(P2\) may then be obtained by taking \(d_i = l n_i\) \((i = 1, 2, \ldots, k)\).

Acknowledgements—Thanks are due to Mr P. S. J. B. Rana, F. R. C. S. (Ed) for allowing data to be collected in the A/E Department of St Peter’s Hospital, Chertsey. The author is indebted to Professor M. R. C. McDowell and Professor K. C. Bowen for helpful comments on a first draft of this paper. This study was carried out while the author was supported by the Greek State Scholarship Foundation.

REFERENCES

A simulation model for bed allocation to hospital inpatient departments

G. VASSILACOPOULOS

Department of Mathematics, Royal Holloway College, University of London, Egham, Surrey TW20 OEX, England.

Abstract: A general simulation model is presented for the problem of determining the bed complement in hospital inpatient departments to meet a predetermined demand for service subject to a set of operational constraints. The constraints considered in the version of the model presented in this paper are that emergency patients should be admitted without delay, that the occupancy should not fall below a prespecified level, and that the length of the waiting list should not exceed a prespecified level. Extensions and modifications of the model are briefly discussed. Three case studies on the General Medical, General Surgical and Orthopaedics departments of St. Peter's Hospital, Chertsey, Surrey illustrate how this model may be utilised.