

## INTERCONNECTED DES MODELS OF EMERGENCY, OUTPATIENT, AND INPATIENT DEPARTMENTS OF A HOSPITAL

Murat M. Gunal  
Michael Pidd

Department of Management Science  
The Management School  
Lancaster University  
Lancaster, LA1 4YX, UNITED KINGDOM

### ABSTRACT

National Health Service (NHS) performance targets in England have put pressure on hospital management to reduce waiting times. The stochastic nature of emergency patient arrivals creates problems for capacity planning for elective patients. We present a whole hospital model which can be used at policy level to investigate cause and effect relations, such as effects of increased emergency arrival volumes on elective waiting times. A typical general hospital can be abstracted in three main units; Accident and Emergency (A&E) department, outpatient clinics, and inpatient units. In real life these units are coupled and share hospital resources. We developed three discrete event simulation (DES) models for each unit to form a whole hospital DES model. We present our models conceptually and our main discussion is on the level of detail in these three models.

### 1 INTRODUCTION

Providing high quality health care is never cheap and costs a significant proportion of GDP in many countries. In the UK, most health care is provided through the National Health Service (NHS), which is financed through taxation at an annual cost of some £90Billion. In some countries, health care consumption is, effectively, rationed through a price mechanism, backed up by some form of insurance. Since NHS care is free at the point of need, there is no price mechanism and hospital care is mainly rationed through waiting lists for elective admission.

It is well-known from simple queuing theory that, when demand is stochastic, some form of waiting for service is inevitable unless the system is resourced at levels well above those needed for average demand. Further, the seasonal nature of the demand for elective care means that providing resources to meet average demand will lead to

further queues for service. Hence, planners of health care in non-price rationed systems like the NHS face a choice. Either provide enough resource to meet peak demand – which will lead to slack for much of the time, or accept that some form of waiting for treatment is inevitable.

Fifty years after the creation of the NHS in 1948, waiting lists and waiting times for many types of elective hospital treatment had grown to outrageous lengths and the UK government introduced a performance assessment framework and extra funds to bring waiting times down to more reasonable levels. In England, this required hospitals to monitor waiting times and waiting lists and to meet targets. Over time, these targets have increased markedly in severity but with more focused management and extra resources most hospitals have made great improvements in their performance. However, the stochastic and seasonal nature of demand means that continued improvement gets harder each time the performance screw is tightened. For example in 2005, some of the waiting time targets were as follow;

- Maximum wait for an Outpatient Appointment must be 13 weeks.
- Maximum wait for inpatients must be 26 weeks.
- Cancelled operations must be admitted in 28 days.
- Emergency bed-days must be reduced by 5%.
- Emergency patients via A&E must be admitted in 12 hours.
- Patients in A&E departments must spend 4 hours or less.

The first three of these targets concern elective care and the latest target in 2007 which combines these three is that no patient will spend more than 18 weeks from referral to treatment by 2008 known as the 18-week Referral-To-Treatment (RTT) target. This target is ambitious for many hospital trusts in England.

In this paper, we describe the conceptual development of discrete event simulation models of whole hospital to investigate the feasibility of performance targets, their interactions and their effect on hospital resources. We focus on two issues: first, finding a plausible level of detail for our models, and secondly, the development of generic models that can be parameterised to fit a range of hospitals.

As in any simulation model, the first key question is: What is the intended use of the simulation? This in turn allows consideration of two more questions: Where will the system boundaries be drawn? What level of detail is appropriate? Our models are intended for use by policy makers who wish to know whether such performance targets are feasible and by local managers and clinicians who wish to explore different options for performance improvement.

Our approach is based on the development of generic models that can be populated by data from or about individual hospitals. In the UK, data about each in-patient and out-patient episode is collected in a standardised form on a national basis (Health Episode Statistics – known as HES data). The idea of our modelling is that, as far as is possible, this HES data should be used to populate the models; supplemented, as necessary, by locally collected data. The core concept of the models is the simulation of individual patient flows as they change state during their time in hospital. In essence, their hospital stay becomes a connected series of input-output processes in which patients spend time during treatment.

DES has been widely used in health care analysis and improvement for many years. However, most applications tend to be highly focused with a microscopic scope on single services such as emergency departments. By contrast, ours is a whole hospital simulation model to be used to assess and improve their performance in a holistic way. A review of the related literature can be found in Jun et al (1999), Fone et al. (2003), and Gunal and Pidd (2005).

## 2 MODELS

At a conceptual level, we consider an abstraction of a typical general hospital in three main parts; Accident and Emergency Department (A&E) for emergency patients, Outpatient Clinics for elective patients, and Inpatient wards/units for both elective and emergency patients if

they are admitted to a ward. All three of these components are within the physical boundaries of a hospital. Hence, movement between these is wholly endogenous. Another significant source of emergency patients for direct admission, is emergency General Practitioner (GP) referrals and these, along with arrivals at the A&E and elective admissions form the exogenous data that drives the hospital performance. Our overall model consists of three DES sub-models, which are presented in following sections. The A&E sub-model and Outpatient sub-model, which act as emergency and elective patient generators, are connected to the Inpatient model. All inpatient admissions share beds from the same pool, available across the hospital wards. The three sub-models are separately validated using HES and hospital specific data. All three sub-models are implemented in Micro Saint Sharp (2007).

### 2.1 A&E Sub-model

This is a self-contained simulation model of a hospital's A&E department. It simulates the main patient pathways. Upon arrival, walk-in patients are registered, assessed, and treated, and ambulance arrivals are taken directly to a treatment cubical. Our high level A&E representation in Figure 1 resembles Healthcare Commission's (2005) national review of A&E departments and in this respect it is applicable to other A&E departments, at least in England.

The model can run individually to be able to analyze only A&E performance, or can be connected to the inpatient model to feed hospital beds with emergency patients. We incorporated the following inputs to the model, most of which can be found in hospital admission databases. However, doctor's service times and test percentages were not available and special data collection was necessary. Details of this particular model including its validation can be found in Gunal and Pidd (2006).

- Patient arrival volumes and patterns for walk-in and ambulance patients
- Number of staff in each hour, by staff role (senior doctor, junior doctor, nurse)
- Number of cubicles
- Service time distribution parameters, by triage category and by doctor type
- Test and X-Ray percentages by triage colour

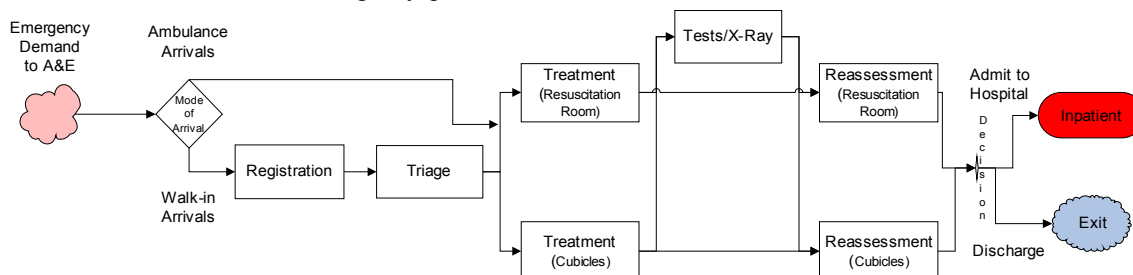


Figure 1: Layout of the A&E Model.

- Patient population's triage colour distribution

## 2.2 Outpatient Sub-model

This model is intended to be used as elective patient generator for the inpatient model. Figure 2 shows the main stages in patients' journey from first referral to admission. Patients are referred to specialists (or consultants) by GPs or by other consultants from other specialties. Historical analysis shows that most referrals are made by GPs. At the first outpatient appointment, the consultants may require diagnostic tests and, requiring further outpatient appointments prior to deciding whether inpatient admission is needed. In addition, many patients need follow-up appointments in outpatients following surgery. Hence, there are three types of outpatient appointment: first, pre-op follow-up and post-op follow-up.

As stated earlier, the current RTT target is that all patients must spend no more than 18 weeks from GP referral to treatment by 2008. The treatment can be as simple as some medication or as complex as a series of surgical operations. If inpatient admission is not required, the RTT measurement clock stops when the consultant commences the treatment in the clinic. However if inpatient admission is required, the treatment starts, and hence the RTT measurement clock stops, when patient is admitted for the first operation.

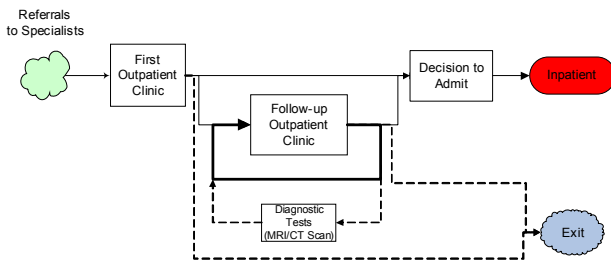


Figure 2: Layout of the Outpatient Sub-model.

We can simplify the delays in patients' journey into three phases;

- Delay between referral and first out-patient (OP) clinic appointment
- Delay between first OP clinic and follow-up OP clinic (or clinics).
- Delay between decision to admit (generally the last OP clinic) and admission.

The first two types of delays are directly associated with consultants OP clinic capacity. Intuitively, the more clinic slots there are, the less time patients have to wait. To reduce waiting times in these two phases, hospitals are either increasing their OP capacity (by putting weekend or evening clinics), or are cutting down unnecessary waits

(by having one-stop clinics which some diagnostic tests are done on the same day as OP clinic). The third delay is, on the other hand, associated with the bed capacity of hospital, or with theatre capacity. Based on expert advice about UK hospitals, we assume that theatre capacity is in line with bed capacity for non-day-case elective patients, and therefore do not include a theatre model.

We considered two options in modelling this system; first to use stationary distributions for delaying patients in the three stages of the journey, second, to build a model which takes clinic capacity into account. The first option has a drawback that there is no possibility of running what-if scenarios, such as "what if we have weekend clinics". Choosing the second option, on the other hand, leads to increased complexity. A model which takes clinic capacity into account should take every consultant's schedules, priorities in first/follow-up booking, and differences in service rates into account. Producing such a model in a way that is applicable to all hospitals may make too great a demand on hospital specific input data. This suggests that an intermediate approach is needed in which OP capacity is reflected in performance.

## 2.3 Inpatient Sub-model

This sub-model simulates patient journeys once they are admitted as inpatients. Patients are admitted to hospital as emergencies through A&E or direct GP referral, or as electives through OP clinics. We analyzed our pilot hospital's admission database and specifically focused on three points in our analysis; patient journeys within the hospital, patient duration of stays (DoS) in wards, and admission patterns. These are the basis for the sub-model inputs. Our analysis revealed the following facts for the pilot hospital, which are believed typical of mid-sized general hospital;

- 80% of patients are admitted as Emergency and the rest is as Ordinary Elective (non Day Cases). Therefore most of the beds are consumed by Emergency patients.
- Within the emergencies, half of the patients are admitted through A&E and the rest are referred by GPs as emergencies.
- Analysis of patient journeys revealed that
  - 60% of Emergency and 88% of Elective patients stay on one ward throughout their period in hospital,
  - 28% of Emergency patients are first admitted to an Assessment Unit and then transferred to another ward,
  - 12% of all patients stayed on more than 2 wards.
- Emergency patients stay longer than Elective patients. Apart from same day discharges, all wards

DoS distributions come from Exponential family, some with heavy tails.

- The number of admissions via A&E does not vary significantly by day or by week. However, Emergency GP Referrals drop in weekends and in holidays. Ordinary Elective admissions have similar attributes, e.g. no admission on Saturdays, very few on Sundays and drop in holidays.

A general process diagram for the inpatient simulation sub-model is given in Figure 3. It depicts the three types of arrivals, ward stays and patients journeys in hospital. The feedback loop represents transfers of patients from one ward (or unit) to another within the hospital. Patients stay in hospital until they are “dischargeable”. The two other simulation sub-models are shown as black boxes. Note that “Emergency demand from GPs” is not modelled in detail, and stationary distributions are used to generate this type of patient.

There are two queues in this abstraction. The first queue is between the A&E department and hospital wards and is sometimes known as the “trolley wait”, usually measured in hours. The second queue is the waiting list for elective care and patients wait several weeks in this state.

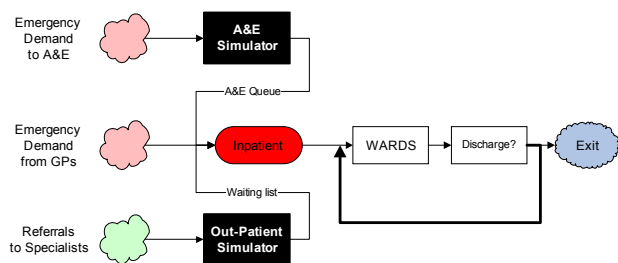


Figure 3: Layout of the Inpatient Model.

Categorizing patients according to admission method is surely not detailed enough to depict the complex treatment stages of patients. An identifier must be used to describe the medical condition on which the pathway will depend. Some examples include Hospital Resource Group (HRG), ICD10, Treatment Specialty or Consultant Specialty. We choose “First Consultant Episode Specialty” as patients’ medical condition identifier, since this is easily extracted from HES data. Thus, in the simulation, a specialty is assigned to each simulated patient and this stays unchanged until the patient is discharged.

Based on their specialty, generated patients take appropriate routes within the hospital based on “ward transition matrices”. These contain routing probabilities that can be calculated using hospital’s admission database. We used stationary distributions for each ward’s DoS distribution. For long stay wards with high variation, we fitted DoS distributions to hyper-exponential. The outputs of

this sub-model are the occupancy of wards, and total length of stay (LoS) of patients.

This model and other two sub-models form the whole hospital model once they are linked together. All three models run in Micro Saint Sharp, using a minute as the basic unit of time.

### 3 DISCUSSION ON LEVEL OF DETAIL IN THE MODELS

Simulation models are built for some purpose and no model is valid for all purposes. Modellers know that objectives are the driving factors of a project. What to include, what to exclude, how detailed the models are all depend on what the models are intended to do. Therefore level of detail in a simulation model is a function of model objectives.

The term “Level of detail” can have various meanings in a simulation model. It may be considered as number of inputs, number of processes, types of entities, types of attributes, number of outputs, or lines of codes. One can see a simulation model of a system, for example, with 10 processes and 3 output variables as “detailed” and another can see a model with 5 processes and 5 outputs as “not detailed”. However one thing is almost certain, that level of detail is directly proportional with complexity. More detailed models need more data for their inputs, more time to run, and are harder to understand.

The simulation literature has generally paid attention to “level of detail” issue with notions such as complexity, scope, and conceptual modelling. Examples include Brooks and Tobias (1996), Chwif, Barretto, and Paul (2000), Henriksen (2006), and Robinson (2006).

One of our aims is to find the appropriate level of detail for a whole hospital model. We divided a hospital into three top level parts and modelled them separately in a “divide and conquer” fashion. The resulting sub-models have different levels of detail, e.g. very detailed A&E model, less detailed inpatient model. Table 1 shows two factors affecting level of detail in our models. More factors can be added such as run-time length, number of lines of codes, and number of variables.

Table 1: Level of detail in our models.

| Models     | Level of Detail | Number of input types | Number of top-level processes |
|------------|-----------------|-----------------------|-------------------------------|
| A&E        | High            | 6                     | 8                             |
| Outpatient | Medium          | 5                     | 5                             |
| Inpatient  | Low             | 2                     | 4                             |

The most detailed of all three is the A&E sub-model. It captures almost all major processes in patients’ journeys in A&E. Model inputs are also detailed. Weekly, daily and hourly variations are taken into account for arri-

val patterns. Likewise, the number of staff by role (senior, junior doctors, nurses) is entered for a typical week, by day and by hour. Instead of using a sub-model for A&E arrivals, we could replace it with a black-box and use probability distributions derived from the A&E sub-model as patient generators to the inpatient sub-model. This would obviously decrease the level of detail. However doing this would remove the chance to experiment the models so as to examine the effects of within changes A&E on waiting time targets. For example it may be valuable to investigate the effect of fewer of A&E doctors on elective waiting times, or the effect of increased percentages of X-rays to radiology waiting times.

The outpatient sub-model is less detailed but depicts all major stages of patient journey. Likewise, we can not experiment with what-if scenarios without a sub-model of outpatients. If we had assumed elective arrivals as a black-box, we would not know the effects of booking policies on elective waiting times.

#### 4 CONCLUSION

Among other OR/MS techniques, simulation has many advantages in modelling complex systems such as hospitals. Waiting time related performance analysis which takes hospitals' stochastic features into account, such as variations in length of stay and arrival patterns, can be made using simulation models.

We presented a conceptual model of a whole hospital. It includes three sub-models, A&E, outpatient, and inpatient which interact with each other and can work all together. We specifically focused on conceptualizing main processes of these units to achieve a plausible level of detail. Our objective is to see affects of capacity changes on waiting times and validate performance targets. We are also interested in generalizing hospital processes to be able to populate these models to different hospitals.

We concluded that, based on our main objective of evaluating NHS waiting times for a hospital, these three models should have different levels of details. The A&E model is detailed enough to evaluate A&E related targets such as the 4 hour target. It relates A&E capacity, doctors, nurses, cubicals etc., with waiting times of patients as measured in minutes. The outpatient model is less detailed than the A&E model and only picks patients main stages in their treatment journey. It relates clinic capacity, consultants, appointment types, booking system etc., with waiting time of patients as measured in days or weeks. The back-end model, the inpatient model, is the least detailed model and it links these two input generators to physical bed capacity. As consequence of bed shortages at the time of need for emergency and elective patients, waiting for a bed can be measured in hours or weeks.

In our future research, we want to analyze the affects of discarding a detail in the model on model outputs. For example, how will the outputs change when we use one Ward Transition Matrix for all specialties, instead of for each specialty? Or how faster the model will run? Likewise, for whole hospital model how will the model behave if we use black-box models, instead of using detailed A&E and Outpatient sub-models?

#### REFERENCES

- Brooks, R.J. and A.M. Tobias. 1996. Choosing the Best Model: Level of Detail, Complexity and Model Performance. *Mathematical and Computer Modeling*, 24(4), 1-14.
- Chwif, L., M. R. P. Barretto, R.J.Paul. 2000. On Simulation Model Complexity. *Proceedings of the 2000 Winter Simulation Conference*. J. A. Joines, R. R. Barton, K. Kang, and P. A. Fishwick, eds. Pp 449-455.
- Fone, D., S. Hollinghurst, M. Temple, A. Round, N. Lester, A. Weightman, R. Roberts, E. Coyle, G. Bevan, S. Palmer. 2003. Systematic Review of the use and value of computer simulation modelling in population health and health care delivery. *Journal of Public Health Medicine*. 25 (4), 325-35.
- Gunal, M.M. and M.Pidd. 2005. Simulation Modeling for Performance Measurement in Healthcare. *Proceedings of the 2005 Winter Simulation Conference*. M. E. Kuhl, N. M. Steiger, F. B. Armstrong, and J. A. Joines, eds. Pp.2663-2668.
- Gunal, M.M. and M. Pidd. 2006. Understanding Accident and Emergency Department Performance Using Simulation. *Proceedings of the 2006 Winter Simulation Conference*. L. F. Perrone, F. P. Wieland, J. Liu, B. G. Lawson, D. M. Nicol, and R. M. Fujimoto, eds. Pp. 446-452.
- Healthcare Commission. 2005. Acute Hospital Portfolio Review: Accident and Emergency.
- Henriksen, J.O. 2006. Taming The Complexity Dragon, *2006 Winter Simulation Conference*, Titans of Simulation.
- Jun, J.B., S.H. Jacobson, J.R. Swisher. 1999. Applications of discrete event simulation in health care clinics: a survey. *Jnl Opl Res Soc*. 50, 109-23.
- Micro Saint Sharp. 2007. Alion Micro Analysis and Design. available via <[http://www.maad.com/index.pl/micro\\_saint](http://www.maad.com/index.pl/micro_saint)> [as accessed May 2007].
- Robinson, S. 2006. Conceptual Modeling for Simulation: Issues and Research Requirements. *Proceedings of the 2006 Winter Simulation Conference*. L. F. Perrone, F. P. Wieland, J. Liu, B. G. Lawson, D. M. Nicol, and R. M. Fujimoto, eds. Pp 792-800.

## **AUTHOR BIOGRAPHIES**

**MURAT GUNAL** is pursuing his PhD at Lancaster University. He received his MSc degree from the same university in 2000. He is an experienced simulation modeller and his current research interest is to investigate the use of simulation on performance measurement in public services such as hospitals. He worked as a simulation analyst in the Turkish Navy where he was interested in military applications of operational research. Email to [m.gunal@lancaster.ac.uk](mailto:m.gunal@lancaster.ac.uk).

**MIKE PIDD** is Professor of Management Science at Lancaster University where his work spans simulation modelling and the development of improved simulation methods utilising current developments in computing hardware and software. His current application work focuses on simulation modelling for improvement in public services, especially in policing and healthcare. He is known for three books: Computer simulation in management science (in its 5th edition), Tools for thinking: modelling in management science (in its 2nd edition) and Systems modelling: theory and practice – all published by John Wiley. Email to [m.pidd@lancaster.ac.uk](mailto:m.pidd@lancaster.ac.uk).