Lancaster University Management School Working Paper 2005/061

Operational research modelling to support healthcare operations management

Dave Worthington

The Department of Management Science
Lancaster University Management School
Lancaster LA1 4YX
UK

The LUMS Working Papers series can be accessed at http://www.lums.lancs.ac.uk/publications/
LUMS home page: http://www.lums.lancs.ac.uk/

Operational Research Modelling to Support Healthcare Operations Management

Dave Worthington, Department of Management Science, Lancaster University Management School, Lancaster, LA1 4YX.
Email: d.worthington@lancaster.ac.uk

Abstract

One often proffered guide for OR model building is that the model should be made as simple as possible subject to meeting its purpose.

In healthcare operations management the purpose of an OR model is often to persuade the appropriate stakeholders to modify a system in a way that (hopefully) turns out to be an improvement. Hence the choice of model can be very context-specific, depending very much who needs to be convinced of what.

This general issue is examined via a brief description of five different modelling approaches that have been used to support operations management of outpatient clinics. The extent to which the approaches might be 'fit for purpose' in different is discussed.

Introduction

This paper is written for healthcare OR practitioners and researchers (as might attend ORAHS31). It aims to address the important general issue of model choice in healthcare modelling (and perhaps beyond). It is well recognised that different aspects of healthcare activity often require very different modelling approaches, see for example Vissers & Beech [2005] which includes descriptions of modelling approaches for:

- hospital production and capacity planning;
- planning capacity of a new hospital unit;
- master scheduling of medical specialists;
- business planning for a surgical specialty;
- scheduling outpatient clinics;
- hospital admission planning and patient mix optimisation;
- balance of care for elderly patients.

However it is perhaps less well recognised that choices of modelling approaches exist <u>within</u> healthcare activity areas. The point of this paper is to highlight and address this issue by demonstrating a clear range of choices <u>within</u> one particular well-known healthcare operations management problem area. This device is used to prompt discussion of the important factors other than the healthcare activity being studied that might well influence model choice in practice.

Outpatient Clinic Appointment Scheduling.

The area of healthcare activity selected for this paper is Outpatient Clinic Appointment Scheduling. The problem of outpatient and general practice appointment systems will be well known to many healthcare OR practitioners and researchers from accounts in the literature, involvement in studies or indeed from personal experience. The fundamental problem is the case where patients are given appointments to see one or more doctors. The patients may arrive on time, early, late, or not at all and are seen by a doctor in order of arrival (or in order of appointment). The two primary performance measures are patients' waiting time and doctor's idle time.

Problems were recognised in health care appointment systems as early as a century ago. In 1908 when opening a new outpatient department at Cardiff Infirmary, Osler said:

> *"As one who has had a long hospital experience, may I mention one essential virtue for the members of the outpatients department staff to cultivate - namely, punctuality. It is not, of course, always possible, but it is remarkable how greatly it facilitates the work of an institution when men put in an appearance at the stroke of the hour.*

Figure 1 shows the architypal performance of an outpatient clinic. The situation depicted here is where patients are given appointments more frequently than they can (on average) be seen by the doctor. This is a very common occurrence in a health system which wants to avoid the medical staff becoming idle because patients do not attend (DNA), or consultations for some patients happen to be shorter than average. As can be seen from the Prob(busy) line, the doctor is almost certain to be busy throughout the clinic (except during the planned 10 minute break 80 minutes into the clinic) until the point is reached when the last patient has arrived. At this point the doctor will have seen all of his/her patients when they become idle. The unfortunate consequence of this virtual guarantee that the doctor never has to wait is that there is often a substantial number of patients waiting to see the doctor. In this case, on an average day, this number would typically grow from 1 to 6; and for a 1 in 10 busy clinic the number would typically grow from 2 to 8.
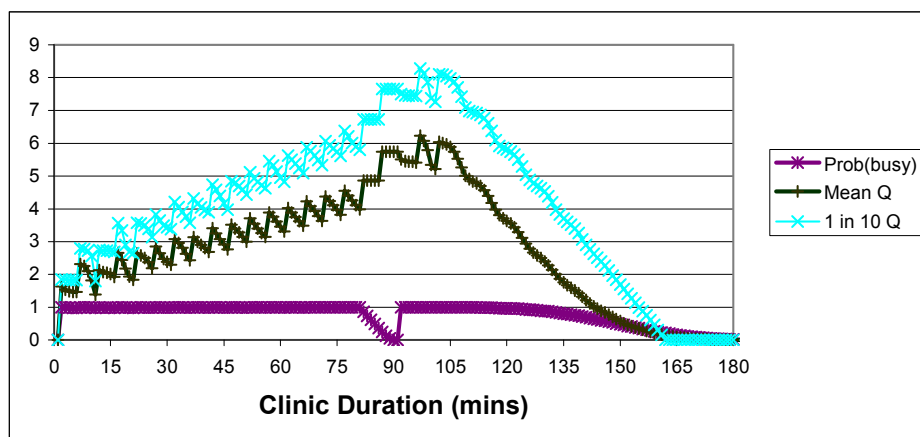

Figure 1: Archetypal Performance of an Outpatient Clinic

In 2003 a specific example of the outpatient clinic problem was faced by the Ophthalmology Department at the Royal Lancaster Infirmary, a major hospital in the north of England. It operated five or six clinics per week, involving two

Consultants, three Junior Doctors and a Senior House Officer (SHO). The numbers of patients booked in for the clinics varied greatly, with some sessions finishing early, resulting in a considerable amount of 'lost' clinician time, and others finishing late, leading to over-running clinics and patients being cancelled at the last minute. As can be expected, waiting times at the clinics also varied greatly. It was not uncommon for patients to be waiting for well over an hour for a consultation that could take as little as five minutes.

A study was undertaken to investigate the potential for re-organising the clinics so that more patients could be seen than at present. Although reducing patient waiting times at the clinics was not an explicit aim of the project, it was envisaged that an improved clinic schedule would also help reduce waiting times, or at least distribute them more equitably. Readers wanting more details of the study undertaken and the models used can refer to Worthington, Goulsbra & Rankin [2005] and Goulsbra & Rankin [2003].

MODELLING APPROACHES

The project at the Royal Lancaster Infirmary led to consideration of five distinct modelling approaches and the application of three of them (as indicated ** below).

1. Process Mapping & Bottleneck Analysis**

2. Steady-state queueing models

3. 'Partial' Simulation modelling**

4. Time-dependent queue modelling**

5. 'Full'/'Fuller' Simulation modelling

1. Process Mapping & Bottleneck Analysis

Process mapping and bottleneck analysis are well-established operations management tools, and their use has been recommended by the NHS Modernisation Agency via their series of six Improvement Leaders' Guides to encourage and facilitate this process. The guides cover a series of topics, establishing a framework for 'improvement leaders' to follow. The first three guides in particular look at how departments can identify areas for change. The topics covered are Process Mapping, Analysis and Redesign [NHS Modernisation Agency, 2002a]; Measurement for Improvement [NHS Modernisation Agency, 2002b]; Matching Capacity and Demand [NHS Modernisation Agency, 2002c].

Process mapping helps the operations manager and stakeholders to see and better understand the important features and stages of a perceived problem situation, and bottleneck analysis uses simple formulae to help identify problematic stages in the process. For example, in the Royal Lancaster Infirmary study simple formulae such as:

$$\text{Clinic Capacity} = \frac{\text{Clinic duration}}{\text{doctors} \times \text{mean consultation time}}$$

were used to give a clear indication of which of the clinic resources were likely to be a major constraint. In this case the bottleneck was clearly the consultation with the doctor.

In general terms processing capacity of the doctors can be increased if either utilisation of doctors' time is increased, or consultation times are reduced. In this case-study the interest was in the former rather than the latter. Discussion with those involved, observation of clinics and analysis of available data was therefore undertaken to ascertain what scope there was for increasing utilisation. This approach eventually focused on clinic durations, patient attendance (and non-attendance) and appointment systems:

- ♦ were too many clinics starting later than planned and/or finishing earlier than planned?
- ♦ could patient non-attendance for their appointments be reduced?
- ♦ how could revised appointment systems help achieve these aims?

In general this type of modelling enables stakeholders to appreciate the nature of the problem that they are facing, and to identify the type of solution that could improve the situation. However it gives no guidance on the precise form of the solution, nor on the scale of change in performance that can be expected. In addition a significant danger with this approach is that it tends to be rather naïve, for example as one bottleneck is removed another may be caused somewhere else in the system.

2. *Steady-state queueing models*

It is very rare to apply steady-state queueing models in this context of outpatient clinics these days. However early work using this type of model provided important generic insights for outpatient clinic management. In common with the bottleneck analysis, steady-state queueing models reinforce the message that improvements in waiting times will be achieved by:

- reducing consultation times

- increasing number of doctors

- modifying arrival rate of patients to better match clinic throughput.

In addition these queueing models also take specific account of the important effects of random variation and the need to reduce it as far as possible. For example they show that typically an appointment system is better than no appointment system, it is desirable to match appointment times to expected duration of consultations, and it is desirable to reduce the number of DNAs where possible, particularly early in a clinic.

Insofar as random variation cannot be completely controlled, Jackson [1964] also recommended traffic intensities for clinics of between 0.85 and 0.95.

However, whilst steady-state queueing models can be used to give valuable insights into the general effects of certain changes to appointment systems, consultation durations and numbers of doctors, they are not capable of making accurate predictions because clinic behaviour is typically very time-dependent.

3 'Partial' Simulation modelling

Given the stochastic nature of the problem, simulation can (in principle) be used to improve understanding of system behaviour and to predict the impact of alternative appointment systems. The term 'partial' simulation is used here to contrast with the more ambitious use of simulation which typically endeavours to model the 'full' system incorporating all the important stochastic aspects of the problem. A 'partial' simulation focuses on a particular aspect of the problem, knowingly ignoring other stochastic aspects, but nevertheless aiming to improve understanding.

In this Royal Lancaster Infirmary case-study partial simulation was used to address two different issues, to discover the practical clinic capacity and to test a proposed revised appointment system.

Clinic Capacities

A simple spreadsheet programme to simulate the sums of randomly generated consultation times (& DNAs) shows that the simple bottleneck equation is inadequate when trying to calculate an actual clinic capacity, even after some allowance has been made for an average DNA rate. In particular :

$$\text{Clinic Capacity} \neq \frac{\text{Clinic duration}}{\text{doctors} \times \text{mean consultation time}}$$

$$\text{Clinic Capacity} \neq \frac{\text{Clinic duration}}{\text{doctors} \times \text{mean consultation time} \times (1 - \text{DNA rate})}$$

The results of this 'partial' simulation of a 12-patient consultant clinic are shown in Figure 2, which highlights the probability distribution of clinic duration (assuming the doctor never has to wait for a patient). Clearly the average duration will be about 165 minutes, but there is a 1 in 20 chance that it could be as short as 130 minutes and a 1 in 20 chance that it would be as long as 205 minutes.
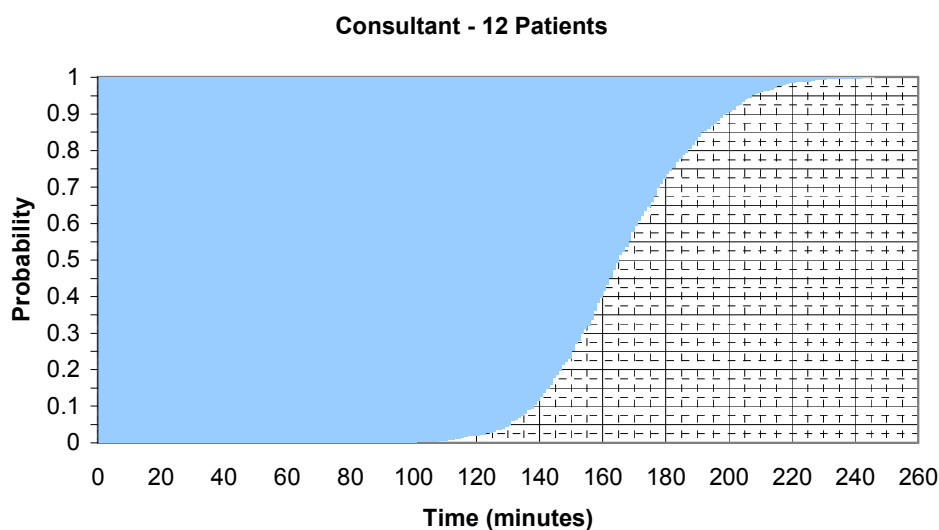
Consultant - 12 Patients

Figure 2: Cumulative probability distribution of (12-patient) consultant clinic duration.

Testing alternative appointment systems

Whilst the analysis above was very effective in highlighting the impossibility of arranging for clinics to finish at the same time each week, it was not seen as a sufficiently accurate model for testing out alternative appointment systems. A different 'partial' simulation was therefore carried out in which five sets of real patients with randomly generated DNA's and consultation times were (manually) put through current and proposed appointment systems. The results from this analysis are summarised in Table 1.

	Current System	Proposed System	Change
Average wait	16 mins	11 mins	-5 mins
No. waiting 31-45 mins	2.8 patients	0.6	-2.2
No. waiting 46-60 mins	0.4 patients	0	-0.4
Clinic duration	2hrs 55 minutes	2hrs 56 minutes	+1 min

Table 1: 'Partial' simulation results for current and proposed appointment systems.

Whilst five simulations is clearly not sufficient to claim a high degree of accuracy, the results were sufficient to demonstrate the nature of the overall impact of moving to a different appointment system. In particular patients would wait less long on average, with only a small proportion of patients waiting over 30 minutes. One added benefit of the manual simulations was the transparency of the model, in that clinic staff could easily be talked through each step of the model.

4. Time-dependent queue modelling

One major disadvantage of the hand simulations is that the relatively small number of randomly sampled patients may make the results unrepresentative. In addition, whilst the average measures of performance are perhaps the most important, they provide very little insight into the range of clinic performance levels to expect in practice.

In contrast a time-dependent queue modelling approach can be used to ensure that statistically representative results for the whole clinic duration are obtained very efficiently.

This modelling approach was implemented using CLINIQUE, software which allows the user to investigate the impact of different appointment schedules on numbers of patients waiting at different times in the clinic and their associated waiting times. CLINIQUE is purpose-built software for modelling appointment systems, and is based on the queue modelling research of Brahimi and Worthington [1991] and Wall and Worthington [1999]. It incorporates statistical information on patients' consultation times and overall DNA rates to predict system performance measures for appointment systems specified by the user.

Figure 3 shows that in an 'average' clinic under the existing appointment system the number of patients in the queue (i.e. in the waiting room) will range between 1 and 3 for much of the clinic, and so patients will experience waits between about 6 and 18 minutes (as consultations average about 6 minutes). Under the proposed appointment system the 'average' clinic will have considerably shorter queues that range between

0 and 1 for much of the clinic, i.e. waits between 0 and 6 minutes. In addition, as not all clinics are average, Figure 3 also shows the estimated queue lengths that would occur in the busiest 1 in 10 clinics. For the existing system queues range between 2 and 5 patients for much of the clinic, i.e. waits between 8 and 30 minutes; for the proposed system queues range between 0 and 2 patients for much of the clinic, i.e. waits between 0 and 8 minutes. Clearly on both the average and the 1 in 10 measures the proposed system is much better for patients than the existing system. However the model also indicates that the probability both doctors are busy is substantially lower in the proposed system, and hence that this particular proposal may have moved too far in favour of reducing patients' waiting times at the expense of under-utilising doctor time.
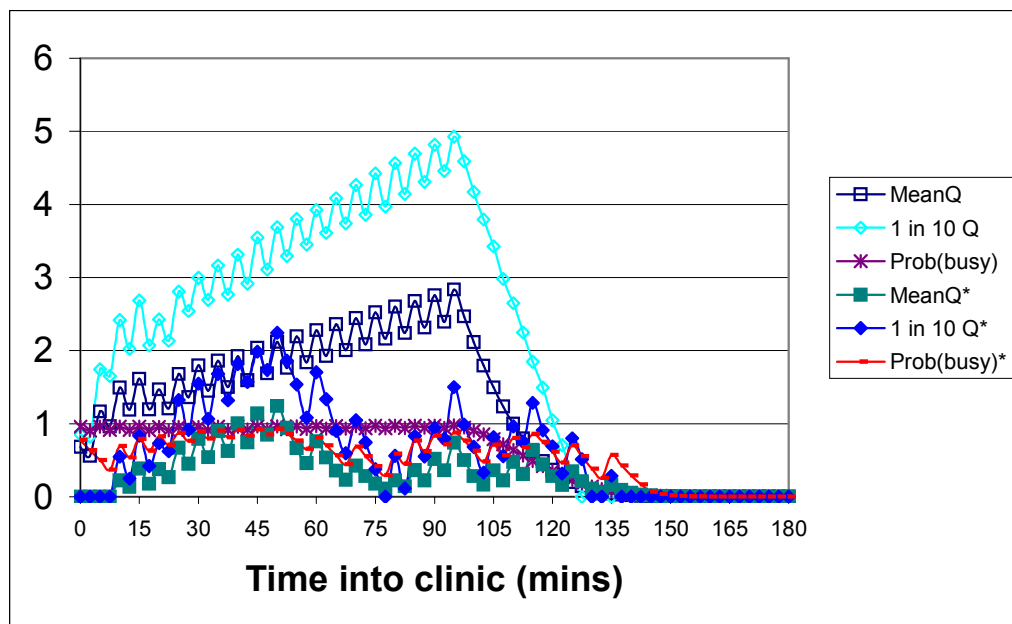

Figure 3: Performance of Registrar Clinics under previous and proposed appointment systems. (NB: * indicates proposed appointment system)

The limitation of the time-dependent queue modelling approach in some settings is that it may not be able to model potentially important features of the system. For example it could not deal with a network of patient flows, where a patient may be delayed significantly not only before seeing a doctor but also by waiting for other tests or treatments. At a finer level of detail it is not able to show, for example, the added benefits of scheduling patients with a higher chance of not attending later in the clinic.

Nevertheless the graph is able to show clearly and efficiently the benefits of moving to a better appointment system, not only on average, but also in terms of comparably busy days and comparably quiet days.

5. 'Full'/'Fuller' Simulation modelling

In theory simulation models can always be used to model time-dependent behaviour of 'simple' and 'non-simple' clinics, as they have the capability to represent a very wide range of stochastic situations. General purpose simulation software (for example

Witness and Simul8) have this capability, and specialist clinic simulation software also exists (for example CLINSIM, see Paul [1995]).

In practice such studies are always limited to some extent by time and by data availability, and this was very much the case in the Royal Lancaster Infirmary case-study. However when sufficient time and data are available, 'full' computer simulation models are sometimes built and run for large numbers of repetitions. Whilst the issue of the 'complete' validation of any model is always debatable, there is no doubt that such models and analyses can be very convincing devices to motivate change and to test alternative systems of management, see for example Dodds [2005].

Discussion

Improvement of healthcare operations is an important objective, and one where operational researchers have often claimed that models have an important role to play. In theory models should enable operations managers to:

- better understand problematic situations,

- generate potential improvements,

- evaluate those potential improvements.

In practice such modelling work often needs to happen within a complex situation, where stakeholders may have competing objectives, system behaviour is not fully understood, only limited data is available, and the situation may be changing over time.

Hence, in practice, the choice of modelling approach needs to be sensitive to these practicalities. If:
- stakeholders can agree on objectives,
- the system of interest is well understood,
- data is available or can be collected on all important parameters,
- and the situation is stable,
a 'full' simulation can be a very convincing tool and have the best chance of correctly predicting the future performance of potential improvements.

However, in less ideal modelling circumstances, the operational researcher needs to be aware of alternative modelling approaches and be prepared to choose between them. In particular the modeller needs to be prepared to use judgement and experience to adopt or recommend a modelling approach that has a good chance of helping the healthcare manager to move forward efficiently in their search to improve problematic situations. This could well involve modelling using approaches selected from a wide range of possible approaches.

Conclusion

The issue of model choice is well recognised in the OR literature, and one often proffered guide for OR model building is that the model should be made as simple as possible subject to meeting its purpose.

However it is rarely clear what this means in practice. Indeed, in the 'heat of battle' there is a tendency for operational researchers to revert to their 'weapon of choice'

rather than to consider the likely strengths and weaknesses of the available approaches.

This paper has sought to highlight the existence and importance of model choice by describing five distinct modelling approaches available for one particular healthcare activity, each of which can be seen to have a role in supporting healthcare operations managers to bring about service improvements.

References

- Brahimi M. and D.J. Worthington. "Queuing models for out-patient appointment systems - a case study". *Journal of the Operational Research Society*, Vol. 42, 1991, No. 9, pp733-746.
- Dodds S. "Designing improved healthcare processes using discrete event simulation", The British Journal of Healthcare Computing & Information Management, Vol 22, 2005, No. 5 , pp 14-16.
- Goulsbra R and J. Rankin. *"A Study for the Ophthalmology Department General Review Clinics at the Royal Lancaster Infirmary."* MSc in Operational Research Dissertation, Department of Management Science, Lancaster University, 2003, UK.
- Jackson R.R.P., J.D. Welch and J. Fry. "Appointment systems in hospitals and general practice". *Operational Research Quarterly*, Vol. 15, 1964, No. 3, pp219-237.
- NHS Modernisation Agency 2002a, Improvement Leaders' Guides (series 1, guide 1). Process mapping, analysis and redesign.
- NHS Modernisation Agency 2002b, Improvement Leaders' Guides (series 1, guide 2). Measurement for improvement.
- NHS Modernisation Agency 2002c, Improvement Leaders' Guides (series 1, guide 3). Matching capacity and demand.
- Paul R. "The CLINSIM simulation package". *OR Insight*, Vol. 8, 1995, No. 2, pp24-27.
- Vissers J and Beech R. "Health Operations Management", published by Routledge, September 2005.
- Wall A. and D. Worthington. "Using the Discrete Time Modelling Approach to Evaluate The Time-Dependent Behaviour of Queueing Systems". *Journal of Operational Research Society*, 50(8) 1999, pp777-788.
- Worthington D, Goulsbra R & Rankin J, 'Scheduling appointments in outpatient clinics', in 'Health Operations Management' edited by Jan Vissers and Roger Beech, published by Routledge, September 2005.