

TUTORIAL: ADVANCES AND CHALLENGES IN HEALTHCARE SIMULATION MODELING

Sally C. Brailsford

School of Management
University of Southampton
Southampton SO17 1BJ, U.K.

ABSTRACT

Simulation has been used for modeling healthcare systems for over forty years. In many respects it is the ideal approach for addressing healthcare issues, yet the relatively small number of successful implementations would suggest that (outside academia) it has been underused in the health sector, compared with manufacturing industry or defense. In this paper we present a review of applications of simulation in healthcare, focusing on successful implementations, and we discuss some possible reasons why simulation has arguably failed to fulfill its potential. We describe recent advances in the area and identify opportunities for further research and new developments.

1 INTRODUCTION

Operations Research (OR) has been applied in the domain of healthcare for more than 40 years. The UK OR Society and the UK National Health Service (NHS) held a joint Colloquium on hospital appointment systems as far back as 1962 (Jackson, 1964). Since the 1960's OR models have been successfully used to assist clinical decision-making, facility location and planning, resource allocation, evaluation of treatments, and organizational redesign. Simulation is one of the most commonly used OR approaches, and is regarded by many as the technique of choice in healthcare (Davies and Davies, 1994). There are three main reasons for this. Firstly, healthcare systems are characterized by uncertainty and variability, requiring a stochastic approach. Secondly, healthcare organizations can be hugely complex and therefore require a modeling approach capable of dealing effectively with complexity. Thirdly, the key role played by human beings in healthcare systems requires an approach which allows interaction and communication between modeler and user or client. These features are all strengths of simulation and help to explain why this approach has been so widely used in healthcare applications.

In this paper we mainly focus on patient-level models. While simulation has been used at a physiological, pharmacological or microbiological level, we will only briefly discuss this type of model, and will chiefly be concerned

with organizational models or disease models, in which the unit of analysis is the patient.

We begin by briefly describing the two main simulation approaches used for these types of model, namely discrete-event simulation and system dynamics. We discuss why these approaches lend themselves to healthcare modeling.

We next propose a taxonomy of healthcare models, and illustrate this with examples of the use of simulation in each category. We then describe some recent advances in simulation in healthcare, and argue that the unique nature of healthcare problems has provided the catalyst for many theoretical developments in simulation.

Next, we consider some of the challenges facing healthcare simulation. Not the least of these is the issue of successful implementation. While academic publications in this field abound, there are relatively few documented success stories compared with other domains such as manufacturing industry. We discuss the reasons for this and some possible solutions. We conclude with a glimpse at the future.

2 SIMULATION APPROACHES IN HEALTH

Even the most cursory look at the literature shows that without doubt, the most widely used simulation approach in health is discrete-event simulation (DES). However system dynamics (SD) has been gaining in popularity in recent years. These are the two approaches on which we shall focus, although Monte Carlo simulation and agent-based approaches have also been used.

DES, in which individual entities flow around a network of queues for services, appears to be tailor-made for hospital systems in which patients join waiting lists for appointments, investigations and treatments. In DES, entities have characteristics which determine their pathway through the network, in exactly the same way as patients have individual characteristics which determine their pathway through the hospital system. This pleasing analogy contributes greatly to the enduring popularity of DES as a modeling approach. Psychologically, DES is appealing because it enables the modeler to give the entities all the

necessary human characteristics of age, gender, diagnosis, blood group, disease status, sexual preference, hair color, or whatever you please.

From a mathematical perspective, DES has many advantages too. Unlike Markov models which take no account of medical history, “service time” distributions can be dependent on individual characteristics and previous history, and any parametric or empirical distribution can be chosen to model activity durations. Complex logical rules can be used to determine patients’ routing through the system, or the outcome of a treatment. We can take account of randomness, variability and uncertainty, as long as we perform enough simulation runs to obtain statistically significant results. Indeed, depending on the flexibility of the software chosen, virtually anything can be modeled using DES.

A final advantage of DES is the wealth of software packages available, at a huge range of prices, most of which have some kind of graphical facility which enables the user to visualize the model. As a communication aid with non-numerate healthcare professionals, this can be invaluable.

System dynamics, on the other hand, lacks many of these desirable features, yet it makes up for this lack by having distinctive virtues of its own. In SD the individual entities are lost, and patients become an indistinguishable mass which flows around the model like water in a central heating system, accumulating in “stocks” corresponding to tanks or radiators, with inflows and outflows governed by valves or “rates”. Psychologically, this is definitely less attractive, especially to healthcare professionals who by training are people-focused and do not like the idea of reducing human beings to computer bytes. Moreover, there is no longer a stochastic element (SD models are deterministic) and despite noble efforts by software vendors to introduce other distributions, the basic assumption in SD is that length of stay in a stock is exponentially distributed. Moreover, in general, SD software does not have the appealing graphics of DES.

However SD has many key features which DES lacks. The fundamental principle of SD is that “structure determines behavior”, so that if we understand the structural relationships between the elements of a system, we will understand the emergent behavior of that system as a whole. In a DES model, we often cannot see the wood for the trees – we are so obsessed with detail (maybe, because we can model it?) that we lose sight of the big picture. In SD, we cannot model this detail, but we can gain understanding of the dynamic complexity of the system. SD is concerned with feedback and unanticipated effects. It has both quantitative and qualitative aspects. Unlike DES, SD models are not dependent on vast quantities of high-quality data, and so can be used at a more speculative or strategic level, for larger populations and longer time-horizons. A key advantage of SD is that the models generally run very fast (and

of course do not require multiple iterations), so can be run interactively in real time with decision-makers.

For excellent surveys of applications of DES and SD in healthcare, see Jun, Jacobson and Swisher (1999), Fone et al (2003), Dangerfield (1999) and Royston (1999).

Of course, simulation has been widely and successfully used for many years in many fields, including defense, manufacturing industry, service industries, finance and for training purposes. Therefore it is interesting to speculate whether healthcare is in any way different from these areas. The entities in healthcare systems are vulnerable human beings, with feelings and emotions, often in life-threatening situations. Healthcare systems are safety-critical and often highly technological. There are different cultures too: the clinical hierarchy has its own professional boundaries, while managers can be seen as interfering outsiders concerned only with costs and targets!

Although patients are obviously not inanimate widgets in a production process, the entities in a model of an airport terminal or a call center are human beings too, so we have to take account of human feelings and reactions in many other settings. Many of the characteristics of healthcare settings are equally true of military models, and culture clashes occur in all organizations.

3 A TAXONOMY OF HEALTHCARE MODELS

For the purposes of this paper, healthcare models are classified into three groups. The first, termed Level 1 models, are models of the human body (see Figure 1). These are frequently called “disease models” but they can also represent biological processes in healthy individuals. Models can be at the system or organ level, or even at a cellular or microbiological level. They are often used for studying the clinical effectiveness or cost-effectiveness of some intervention. For example, by simulating the progression of breast cancer in the female population it is possible to compare the effects of different screening policies for early detection.

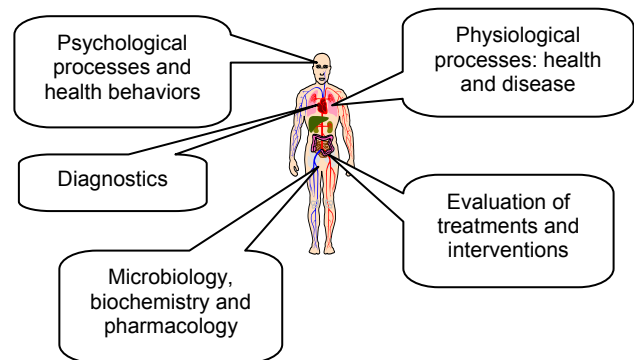


Figure 1. Models of the human body

We can also model people's health behaviors, for example the effect of a smoking cessation campaign on the prevalence of coronary heart disease. We can model the spread of infectious diseases such as chlamydia or HIV/AIDS. Simulation has also been used to model the biochemical effects of drugs, a study known as pharmacodynamics. Examples of these are given in section 4. Level 1 models are arguably the most interesting, as they often require creative modeling approaches or theoretical innovations.

The next group, Level 2 (Figure 2), denotes operational or tactical models at the healthcare unit level. By this we mean a clinic, a ward or hospital department such as the operating suite or emergency room. These usually (although not always!) still model individual patients, but are

not concerned with modeling the physiological or clinical processes going on inside them. Rather, they are concerned with modeling the flows of patients around this system, and identifying and eliminating bottlenecks. These models are used for capacity planning, resource allocation and process redesign. For example, how many ICU beds are needed to be 95% sure there will always be a free bed when required? How many nurses and doctors are needed to ensure that no patient spends more than 4 hours in the ER? Should inpatients or outpatients take priority for diagnostic imaging services? These are classical areas for the application of operations research in general and discrete-event simulation modeling in particular, and the academic literature contains hundreds, if not thousands, of Level 2 models. Some examples are given in Section 5.

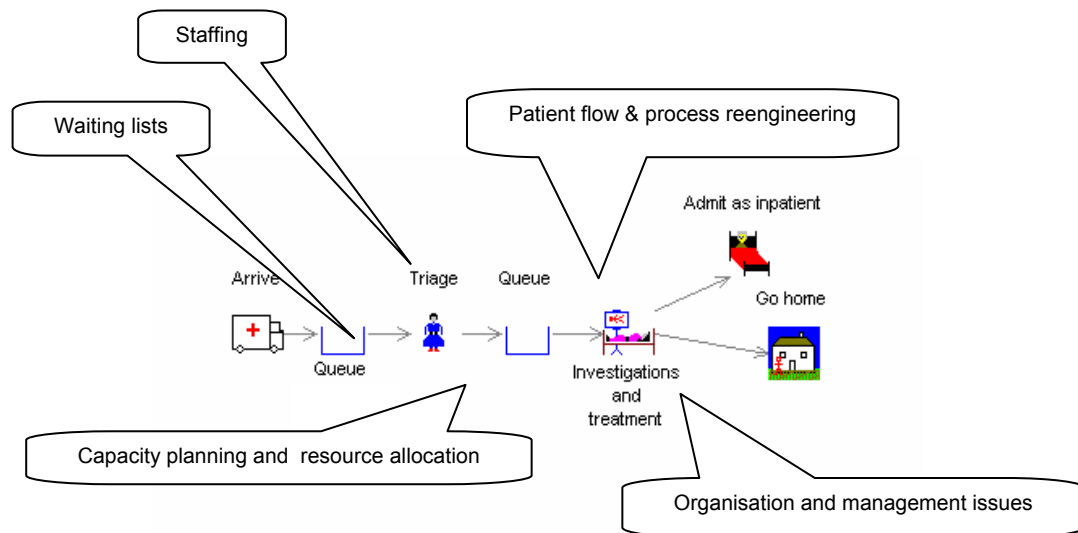


Figure 2. Operational models of healthcare units

Finally, Level 3 or strategic models are system-wide models which often do not model individual patients at all. They are also comparatively few in number in the literature, possibly reflecting the relatively low use of operations research for strategic planning. Unlike Levels 1 and 2, where DES is usually (but not always) the chosen approach, Level 3 models almost always use SD, as it lends itself to answering more long-term, broad-brush questions.

For example, how should a large city configure its services for emergency and unscheduled care? How might the Electronic Health Record impact upon the health of the next generation of children? How many doctors will the US need in 2030? How can we improve the way health professionals work alongside Social Services? Two examples of Level 3 models are given in section 6.

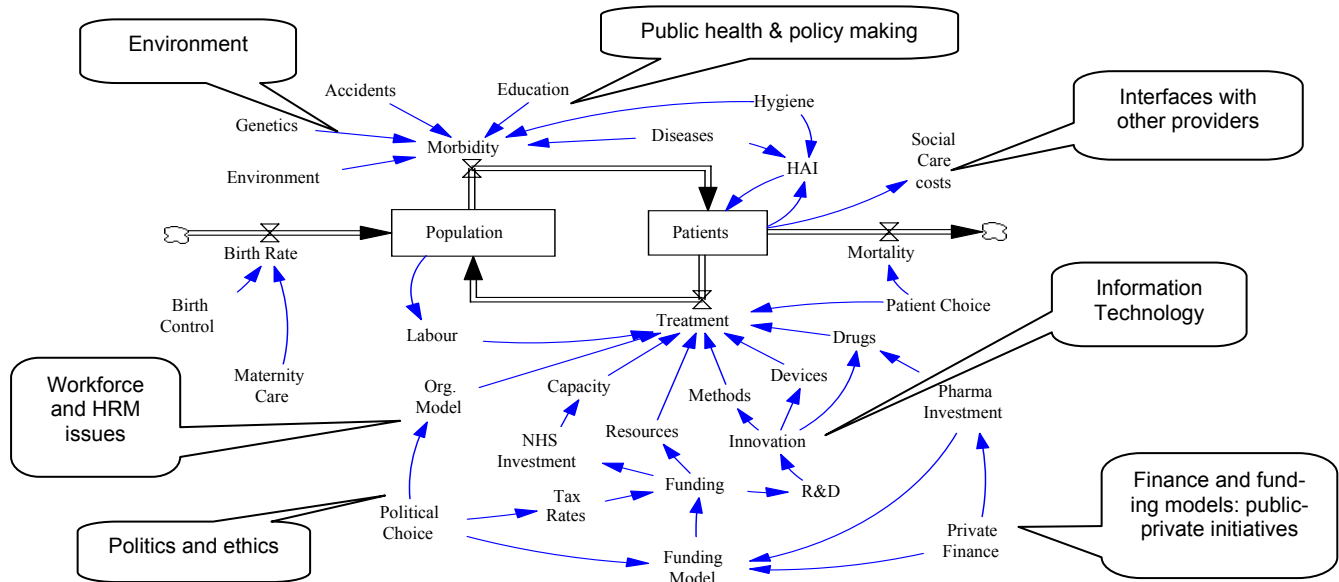


Figure 3. Strategic, system-level models

4 LEVEL 1: MODELS OF THE HUMAN BODY (DISEASE MODELS)

We first describe two DES models, one for screening diabetic patients for eye complications, and another to address the issue of mother-to-child transmission of HIV in developing countries. DES was the method of choice for these two models, but both required some technical modifications to capture adequately the real-life processes involved. We then describe an SD model for the sexually transmitted infection Chlamydia.

4.1 Screening for Diabetic Retinopathy

The term screening means testing an at-risk population for the early signs of a disease, generally because early detection leads to an improved outcome for the patient. Diabetic patients are at risk of a condition called diabetic retinopathy (DR), which can lead to blindness if untreated. However, the early signs of DR can be detected before the patient is even aware of any problems, and the condition can be successfully treated by lasering, preventing future sight loss. Screening can be carried out using various methods and in a variety of settings – in hospital, at an optometrist or in a primary care facility – and while it is clearly beneficial, since the tests are relatively cheap and easy, there is no consensus about the ideal setting, method or frequency of screening.

Davies et al (2000, 2004) developed a DES model which was used to address this problem. In fact two mod-

els were developed, one for Type 1 diabetes (the more acute form, which affects younger people and usually requires insulin injections) and the other for Type 2 (adult-onset, which can be controlled by medication or diet alone). The two types of diabetes cause different types of retinopathy, and although Type 1 is generally more serious, Type 2 is much more common, accounting for 85% of all cases. Diabetes is increasing in developed countries due to lifestyle changes and is a major item of health expenditure (UK Department of Health, 1991).

The modeling approach used was developed by Davies et al (1993) and is termed Patient Oriented Simulation Technique, or POST. The key feature of POST is that it allows entities to be in more than one place at a time. This is important in healthcare models because entities (patients) need to take part in several simultaneous activities, or queue in more than one queue at the same time. Disease progression can be modeled as a virtual queuing system by regarding the disease state dwelling times as activity durations, where the “activities” are assumed to be unconstrained (i.e., no servers are required, and hence there are no actual queues). At the same time, however, patient entities may also be taking part in a genuine, resource-constrained, queuing system, for example waiting for hospital admission, medical treatment, or a screening test. Thus, in a realistic model of a healthcare system, patient entities may be participating in several concurrent activities or queues. Moreover, these activities and queues are interdependent: if a patient changes disease state, his/her treatment may need to be changed, s/he may no longer re-

quire a screening test, and hospital appointments may need to be rescheduled.

For example, the stages of DR are modeled as a series of activities, where each activity denotes a change of state, and therefore a patient will be scheduled to change state at some future point in time. However in addition to this, the patient will be scheduled to attend for his next screening appointment. Also, his time of death will be scheduled. Thus in effect this patient will appear three times on the next-events list, a situation which most standard DES software cannot handle. However, POST overcomes the problem by a clever use of pointers and linked lists, which enable these three scheduled future activities to be linked to the same entity. So if the patient should happen to die before his next screening appointment, all his future events can be rapidly removed from the events list.

The models were coded in Borland Delphi and were populated with data about the natural history (the untreated progression) of DR from one of the world's largest studies, the Wisconsin Epidemiologic Study of Diabetic Retinopathy (Klein et al, 1985). Various screening (and treatment) policies in current use in the UK were tested, with current cost data (NHS, 2001). Detailed results can be found in Davies and Brailsford (2004). Interestingly, no striking differences were found between the policies tested, although there was a preference in terms of cost-effectiveness for community-based rather than hospital-based screening. This was essentially because community methods achieved higher levels of attendance. This turned out to be a key issue, and led to my personal interest in modeling health-related behavior.

4.2 Mother-to-Child Transmission of HIV

HIV/AIDS affects over 40 million people world-wide, and more than 70% of these people live in Africa. Mother-to-child transmission of HIV, during childbirth and subsequently via breastfeeding, accounts for over 90% of all HIV infections in children under the age of 15, and has a devastating socio-economic impact. However it is not clear whether bottle-feeding strategies, in combination with HIV treatment at childbirth, are beneficial for low-income countries. On the one hand, bottle-feeding might inhibit HIV transmission, but on the other hand bottle-fed babies are likely to have a higher background mortality risk compared with breastfed babies, because of infant diarrhea caused by poor water supply (Coutsoudis et al, 2002). Moreover, implementing HIV prevention policies in Africa is extremely difficult because of the poor medical and socio-economic infrastructure.

Rauner et al (2005) developed a DES model to evaluate the relative benefits of two potentially affordable interventions for sub-Saharan Africa, namely anti-retroviral treatment (ART) at childbirth and/or bottle-feeding strategies. The model also used the POST methodology and used

data from rural Tanzania. A population of individuals is created, each with their own personal characteristics: age, gender, HIV status, disease stage, and maternal stage (pregnant or breastfeeding). The "life histories" of these individuals are then simulated as time progresses. They grow older; the females conceive, give birth to, and breast-feed children; people die, both from AIDS and from other causes; children grow up and in turn bear children themselves who begin the whole process again. Moreover, individuals acquire HIV, and then progress through the stages of HIV infection, possibly also transmitting the virus to other people. Individuals therefore take part in many simultaneous activities, for which the POST software is ideally suited. These activities comprise the ageing process, the processes of acquiring HIV infection, the disease progression process, the process of mortality from causes other than AIDS, the processes of conception, childbirth and breast/bottle feeding, and (potentially) the processes of HIV testing and treatment.

However, the unique feature of this model was the way in which mothers and babies were linked. This was necessary in order to model HIV transmission accurately, since the probability of maternal transmission depends on the disease state of the mother. Each woman was linked (by pointers in the entity structure) to all her children and also to her own mother. Therefore, if for example a woman progressed to AIDS while she was breastfeeding, it was possible to trace her baby and change the probability of that baby acquiring HIV. The model required a long warm-up period (from 1990 to 2002) in order to create this family structure in the baseline population. The simulation experiments were then run for a further 12 years.

For all the intervention scenarios, cost-effectiveness-ratios were calculated by dividing the number of additional live HIV-negative children in the intervention scenario compared with the baseline scenario, by the additional costs of the intervention scenario compared with the baseline scenario. The results were sensitive to assumptions about the prevalence of HIV and the baseline infant mortality rate, which both depend on local conditions, and on the efficacy of ART. As HIV prevalence increases, bottle-feeding strategies became more cost-effective. In districts with poor water supply, bottle-feeding is not recommended.

4.3 Screening for Chlamydia

Chlamydia is the most common sexually transmitted infection in the UK 2003 (Health Protection Agency, 2004) and is a major public health problem. Most cases of chlamydia infection are asymptomatic and easily treated with antibiotics, but if untreated the infection can have serious long-term consequences (known as *sequelae*) including pelvic inflammatory disease, tubal infertility and ectopic pregnancy. Screening programs have been shown to be effec-

tive, but there are concerns that blanket screening of the whole population at risk will add extra burden to the overstretched health economy. Unlike diabetic retinopathy, where the population at risk is relatively small and well-defined, in this case all sexually active people are at risk. In 2003 the UK Department of Health introduced screening of all people between the ages of 16 to 25 in ten centers, with the view to extend this program to the rest of the country within the next few years as part of the National Chlamydia Screening Programme (UK Department of Health, 2005).

Evenden et al (2005) developed an SD model which demonstrated that certain high-risk sub-groups within the general population are critical in the infection dynamics, and improved targeting of these high-risk populations achieves greater cost-effectiveness. The model was based on an earlier model developed for the UK Department of Health by Townshend and Turner (2000), but used data from an opportunistic screening trial held in the Portsmouth area from October 1999 to September 2000 (Harindra, 2000).

The SD approach is ideally suited to modeling infection rates and population movements within infected and susceptible states. A particularly relevant aspect was that the repeated re-infection mechanism was captured, along with the increased risk of sequelae this creates. Figure 4 shows the model structure.

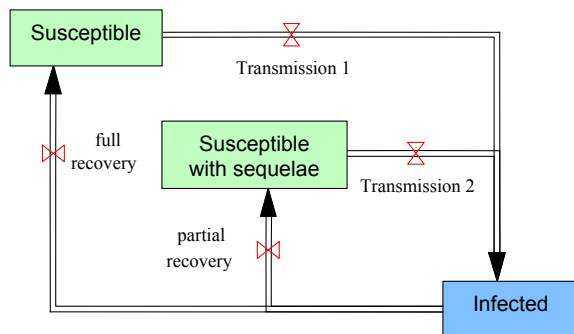


Figure 4. Structure of the Chlamydia model

These models, which were developed in Vensim, were replicated for two risk groups of patients. Based on the findings of Townshend and Turner (2000), 2.5% of the population were assumed to be high-risk. Various assumptions were made about the rates of sexual contact within and between these groups, about the prevalence of infection (5%, 8%, and 10%), and about the screening rates in each risk group. The low risk screening rate was varied from 0% to 10% of the total population, and the high-risk rate varied from a minimum value equal to the low risk rate, to a maximum of 120% of the low risk rate. The model results showed that the more targeted screening became, the more cost-effective it was. Therefore, rather than

increase the general population screening rate by 1%, which would be hugely expensive, it was far more advantageous to identify a few more high-risk people and screen them.

In a further paper, Evenden et al (2007) describe the use of statistical risk-group clustering techniques with the Portsmouth data to identify indicators that are strong predictors in determining high-risk status. The results are combined with geomapping techniques which visually display prevalence geographically across the region, thus identifying high prevalence postcode clusters and informing public health planners where to target intervention and screening strategies. These findings are then combined with the results from the SD model to provide a unique holistic view of the problem.

4.4 Summary

These models are all examples of Level 1 disease models. Although two use DES and one SD, they share several features which are common to this class of model. Firstly, all three models were developed in close collaboration with clinical experts (in the case of the HIV model, one of the co-authors had worked for several years in Tanzania and had expert knowledge of the local healthcare infrastructure). All three models used clinically recognized stages to model disease progression. Secondly, all three models were used to address a public health policy issue, and therefore were aimed at regional or even national decision-makers rather than local hospital managers. Thirdly, all three (even the SD model) required a lot of detailed data. This was derived from a variety of sources, mainly the literature but also (in the case of the Chlamydia model) data collected for a previous study. Moreover, some model parameters were unknown and therefore multiple scenarios had to be run exploring a range of potential outcomes, depending on the value of these variables. Finally, all three models required some degree of non-standard technical innovation in order to adapt the chosen simulation approach to the specific needs of the modeled disease.

5 LEVEL 2: OPERATIONAL MODELS OF HEALTHCARE UNITS

In this section we present, rather more briefly, two DES models developed to address organizational issues of resource allocation and capacity planning.

5.1 Hospital Bed Capacity Modeling

Harper and Shahani (2002) describe a simulation model developed for an 800-bed hospital in Reading, UK. This model, like POST, was also developed in Borland Delphi but used a simulation engine called TOCHSIM. The model

has a generic structure in which patients flow through “care units”, which can be a ward, a group of wards, a specialty bed-pool or even the whole hospital. The model structure is highly flexible and very detailed, and requires a great deal of data input, for example user-defined patient groups, arrival rates (hourly, daily and monthly), length of stay (LOS) distributions, bed numbers, admission rules, deferral rules (in the case when no bed is available) and priority listings.

Because of the need for so much input data, the model was designed to interface with the routine patient management and reporting systems used by the hospital. The fitting of arrival and LOS distributions was fully automated through a system called Apollo, a classification and regression tree (CART) method which allowed the construction of homogenous and clinically meaningful patient groups, thus reducing the need for input from hospital staff. For the groups thus constructed, arrival and LOS distributions were then fitted.

The model was used in three areas in this hospital. For example, the model was used in adult medicine for estimating the future seasonal bed requirements. A key learning for the hospital planners was that occupancy rates and their corresponding refusal rates are linked in a highly complex way, and are a function of the casemix, the number of beds available and the variability in LOS. Previously the hospital had used a rule of thumb of average LOS multiplied by average demand – assumed fixed – and their views changed greatly after this modeling exercise.

5.2 Intensive Care Units

Another popular area for the use of capacity planning models is the intensive care unit (ICU). This is a very expensive hospital resource both in terms of highly skilled staff and costly specialist equipment. Therefore it is vital to provide the right number of ICU beds and staff. (It is difficult to separate staff and beds, since most ICU patients require at least a 1:1 nursing ratio). Having too many beds can have serious financial consequences, but having too few means at best canceled elective surgeries and at worst, loss of life.

Traditionally, most ICU simulation models use a fixed number of beds and determine experimentally the number of patients who either “queue” or are turned away if no bed is available. Such models generally assume that ICU beds are adequately staffed, and do not model nurses as a constraint. Griffiths et al (2005) developed a model for a large teaching hospital which used a more flexible approach to beds. Officially, the ICU had 14 beds, but in times of high demand extra beds would be made available elsewhere in the hospital. The problem addressed by Griffiths et al’s model was not how many ICU beds were needed, but how many nurses.

A fixed number of nurses, permanently based in the ICU and known as establishment staff, are rostered to work in the ICU for each 8-hour shift. In busy periods supplementary nurses may be required, and these can be either bank staff (establishment staff doing overtime) or agency staff. The latter are very costly and moreover may be unfamiliar with the unit. Thus for both cost and patient safety reasons, the hospital would prefer to keep the number of supplementary nurses to a minimum.

A DES model was developed in the software Simul8 (www.Simul8.com), using detailed data for the total of nearly 1100 patients admitted to the ICU in 2000. The model was highly complex. Patients were categorized by referral source and (for surgical patients) into elective or unplanned. Moreover, distinct day-of-week and time-of-day patterns were observed and a separate arrival rate was required for each of the 168 hours of the week. Length of stay distributions were fitted for all these patient types. Because it is nurses rather than beds which are the constraint in this model, a theoretical 30 beds were available, although a maximum of only 19 beds was used in 2000. The rostered number of nurses per shift was fixed for each experiment, but the number of supplementary nurses was assumed to be unbounded and was a performance measure.

In 2000, the actual number of nurses per shift was 14. The model showed that had 16 nurses been rostered, savings would have been made, equivalent to 690 times the cost of employing one establishment nurse for one shift. The model was extended to consider increased future demand and also a pioneering method in use at this hospital for the early detection of patients in general wards who might need ICU treatment at some future point. The physicians believed that by offering early ICU care, the eventual LOS in ICU could be reduced by between 10% and 20%. The model was able to evaluate the effect of this, concluding that a 20% reduction in LOS would result in one less nurse being rostered each shift, with further consequent savings.

5.3 Summary

These two models typify the use of DES in capacity planning. They are complex and detailed; they require a great deal of data, which has to be pre-processed in order to populate the model; and they require input from (and engagement with) hospital managers and clinicians. They produced very useful results for the collaborating hospitals, but it is not clear how widely their use has extended beyond the original hospitals for which they were developed.

6 LEVEL 3: STRATEGIC WHOLE-SYSTEM MODELS

These models are relatively rare in the literature, but are becoming increasingly popular in the UK. In this section we describe two applications of system dynamics, which is probably better suited to this area. The first is perhaps one of the best-known examples of SD modeling in the mainstream UK health OR literature. The second describes the entire emergency and unscheduled health care system in the city of Nottingham, a large industrial city in the center of England.

6.1 SD Modeling of Emergency Admissions

Lane et al (2000) developed a model designed to explore the relationships between waiting times in the ER and hospital bed closures. At the time a major London hospital (denoted pseudonymously “St Danes”) was planning to cut costs by closing beds, and intended to measure the impact of these bed closures by tracking waiting times in the ER. The hospital’s argument was that bed closures led to canceled elective admissions and this led to more people presenting in the ER, partly as a direct result of the deterioration in their health and partly as a behavioral response by primary care doctors wishing to get their patients admitted “by the back door”. It was therefore expected that the first sign of pressure as beds were closed would be that waiting times in the ER would rise because of this increase in demand.

The model was run for a range of different assumptions about the number of beds available, for various permanent increases in demand, and for a “crisis day” when there was a sudden increase of 13% in demand. The key finding was that the immediate impact of bed shortages was not observed in the ER at all, but was evident first in canceled elective admissions, so that using ER waiting times to measure the effect of bed shortages was misleading. In fact, the model showed that it did not make sense to look at any single measure in isolation, but that this was indeed a complex system where changes in one area could have unforeseen knock-on effects elsewhere, and so a holistic view was required.

6.2 Emergency Care in Nottingham

Brailsford et al (2004) developed an SD model as part of a research project which was itself part of a larger, ongoing project in Nottingham, known as the Emergency Care – On Demand (ECOD) project. In Nottingham, emergency hospital admissions had risen dramatically (by over 10% year on year for one of the two Nottingham hospitals) in recent years. The aim of the ECOD project was to look at the whole healthcare system to determine why demand was so high and to investigate what could be done to alleviate this

pressure. The ECOD project had a Steering Group which contained representatives of all the providers of emergency and unscheduled health care in Nottingham: the hospitals, the ER, the ambulance service, out-of-hours primary care services, social care, and two new services, NHS Direct (a national call centre which provides medical advice over the telephone) and the Walk-In Centre, a nurse-led facility designed to treat minor cases and (in theory) reduce inappropriate use of the hospital ER. Our contribution involved carrying out a system review and providing research support to the ECOD project.

The Nottingham study concerned a population of over 600,000 potential patients. Furthermore, although the specific pathways followed by individual patients were of interest, they were of less importance than understanding the major flows of people through the “front doors” to the system, and gaining insight into the general structure of the system and the relationships between its component parts. The problems experienced in the ER, for example, were not principally felt to be due to high variability in casemix or staffing levels, but more to the sheer volume of demand and consequent pressure on resources. Thus SD was selected as the modeling tool.

The study involved both qualitative and quantitative modeling. A system map and a set of influence diagrams were developed through a series of about 30 interviews with different stakeholder and providers. Participants were asked to amend the map and describe the historical, organizational or political factors which influenced flows of patients in their own area. The resulting final map of the system was used as the basis of a quantitative SD model, developed in Stella/ithink, which was populated with activity data for 2000-01 from all the providers. After the model was validated for the current year, a range of scenarios were run (suggested by the ECOD Steering Group). These looked at the effects of various interventions, including early discharge, increasing the proportion of patients who used the Walk-In Centre rather than the ER, reducing admissions for specific groups of patients (e.g. over 65’s) and providing additional diagnostic services in the community, since it was felt that some primary care physicians used the ER to bypass the normal referral process for diagnostics, so that patients ended up being admitted unnecessarily (the same effect as observed earlier by Lane et al (2000)).

The key findings of the model were that if current trends continued, both hospitals would be forced to cancel several hundred admissions for elective surgery every month within a couple of years. However relatively small changes in one part of the system had significant impact elsewhere. For example, it was much better to prevent admissions through the use of a community diagnostic facility than it was to reduce length of stay and discharge patients early. The model was run interactively with the Steering Group and the results used to inform a Stakeholder Day, at which focus groups developed a Local Ser-

vices Framework for emergency care. A Treatment Centre was subsequently built in the grounds of the larger hospital to provide community diagnostic services and to fast-track routine surgical cases.

This model was used to inform strategic decisions. The numerical results of the model were indicative of trends and were relative rather than absolute. A DES model of this size would be very slow to run even a single iteration. Moreover, the qualitative insights which the model provided (and indeed the value of the mapping process as a communication tool) were found to be as useful as the numerical results. This approach is now being used by Brailsford and Lattimer in a national study of workforce change in unscheduled care, in which many of the same strategic issues arise.

6.3 Summary

It is no coincidence that these two examples both come from the field of emergency care. This is an area where it is almost impossible to draw a well-defined boundary around any subset of the modeled system. Emergency healthcare systems demand a whole-system approach. For this reason DES models of emergency departments, much as they appeal to the Operations Research modeler, are often doomed to fail. We may model the internal workings of the emergency department in great detail, but most of the time the problems do not lie within the ER but with its interfaces with other hospital departments, the ambulance service and all the other primary care and community services outside the hospital.

7 RECENT ADVANCES IN HEALTHCARE MODELING

We have seen how the need to model simultaneous activities, driven by the requirements of healthcare models, gave rise to the POST software developed by Davies et al (1993). Similarly, the need to automate the pre-processing of large datasets to provide the input for healthcare simulation models led to the CART tool developed by Harper and Shahani (2002) as a front-end for a DES model. Many other interesting advances have been made in recent years. Sometimes, new diseases have led to an upsurge in modeling developments: examples include the huge number of models developed in response to the AIDS epidemic, not to mention SARS and the potential avian flu pandemic. Advances have been made in combining simulation with other techniques, for example optimization or geographical modeling. Many off-the shelf simulation packages now have built-in optimizers. ARENA, for example, includes a tool called OptQuest which can use a variety of metaheuristics including scatter-search, tabu search and neural networks. Simul8 has an add-on tool called Optimiz which

used neural networks to guide simulation runs. Advances in the user-friendliness of software have meant that tools can now be placed in the hands of healthcare professionals, rather than being the sole domain of Operations Research modelers. We describe three models which illustrate these points.

7.1 Combining Simulation and Optimization

The model for diabetic retinopathy developed by Davies et al (2002) described in section 4.1 was used as the basis for an exploratory paper by Brailsford et al (2007) in which simulation was combined with ant-colony optimization (ACO). ACO (Dorigo, Maniezzo and Colomi, 1991) is inspired by a biological metaphor. In nature, ants are able to find shortest paths between their nests and food places by a collective behavior that is characterized both by random exploration and by the use of markers (chemical substances called “pheromones”) that are deposited on the ground and represent the “common memory” of the ant colony. This exploration/exploitation behavior can be mimicked in computational approaches to the solution of search and optimization problems, which has led to a large (and rapidly increasing) number of successful applications.

Our application used a variant of ACO called stochastic or s-ACO (Gutjahr 2003). The simulation is intricately embedded within the ACO procedure. The overall process is driven by s-ACO and involves successive invocations of the simulation to evaluate specific policies. Each ant represents a choice of screening policy (a set of parameters such as start age, end age, interval between screens, and so on), for which a solution - the value of the objective function - is calculated by simulation. Only a few iterations of the simulation are carried out for each screening policy, but as the optimization proceeds and better candidates are identified, a larger number of iterations are performed when comparing the current best solution with the global best solution. Thus s-ACO allows a more efficient and intelligent exploration of the solution space than traditional simulation experimentation. Multiple random restarts (“rounds”) are used in order to avoid local optima.

We compared these policies in terms of two objective functions: Min cost-effectiveness (minimum incremental cost per year of sight saved, compared with a no-screening baseline) and maximum effectiveness (years of sight saved). We chose the second objective merely to test the s-ACO algorithm against a known optimum, since obviously if money were no object, the most effective policy would be to screen the maximum number of people as frequently as possible with the most accurate (but expensive) test. It was interesting that the most cost-effective policy turned out to be to screen people between the ages of 30 and 60, at 30-month intervals, using the cheapest available test. Such a policy would be socially unacceptable. On a practical level, the model was very slow to run and while it was an

interesting experiment, we concluded that further work would be needed before this became a practical tool.

7.2 Modeling Human Behavior

We have shown that simulation has been widely used to evaluate the outcomes from medical interventions designed to improve patients' health. However in practice these outcomes can be greatly affected by patient behavior. For example, patients may not complete a course of a prescribed medication because they find the side-effects unpleasant. A study designed to evaluate this medication which ignores such behavioral factors may give unreliable results. Health psychologists have developed a number of models of health-related behavior, for example Becker's Health Belief Model (Becker, 1974) and Ajzen's Theory of Planned Behavior (Ajzen, 1991: see Figure 5). Schmidt (2000) developed an approach called PECS, which considered the physical, emotional, cognitive and social aspects of human behavior. Brailsford and Schmidt (2003) developed a DES model which combined PECS with the Health Belief model to model attendance for diabetic retinopathy screening, based on the model of Davies et al (2000).

This approach was extended in a model for breast cancer screening in Jenni Sykes' doctoral thesis (Sykes, 2007; Brailsford, Sykes and Harper, 2006). Sykes used the Theory of Planned Behavior to model a woman's probability of attending for mammography. She developed a DES model representing the natural history of breast cancer in a cohort of women. The three variables, attitude, perceived behavioral control, and subjective norms, join together in a linear regression equation to predict intention to attend. Intention to attend and perceived behavioral control then go on to predict the behavior itself with their own regression weights.

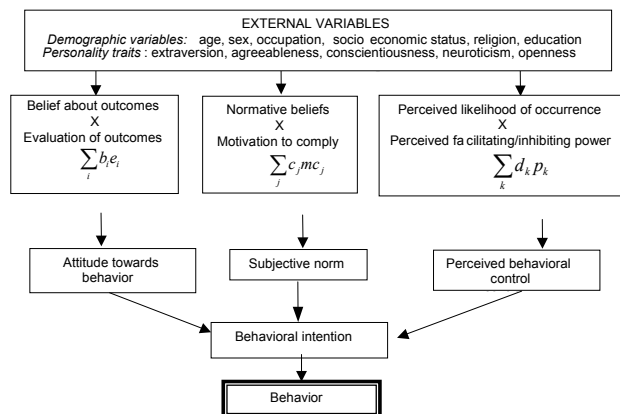


Figure 5: The Theory of Planned Behavior

The model can be used to investigate the effects of behavioral interventions, such as education campaigns designed to increase women's belief in their own perceived

behavioral control, or to raise awareness and change women's attitudes. Work is now in progress to consider how such changes could realistically be achieved in practice, and try to estimate the impact of interventions designed to modify behavior. By looking at the relative costs of such interventions, compared with changes in screening policy, we can assess whether it is more cost-effective to change the eligibility criteria for screening, or run a media campaign, for example.

Modeling of human behavior is not limited to the field of healthcare, of course. There is great interest currently in the defense arena in what is known as "man in the loop" simulation, where human factors are incorporated in military simulation models. In manufacturing industry, the impact of worker behavior can be significant, even in production industries such as automobile manufacture (Baines and Kay, 2002).

7.3 Combining Simulation with Geomodeling

Oral and maxillofacial surgery (OMFS) is a specialty based in dentistry, although it overlaps to some extent with plastic surgery and ear, nose and throat surgery. In the 1990's there was trend in the UK towards OMFS services being provided in a daycare rather than in-patient setting. In 2002, a reorganization of the Strategic Health Authority (SHA) boundaries in London provided the opportunity for a review of the provision of these highly specialized services, and rationalize the service across London by reducing the total number of OMFS centers from 42 to a hub-and-spoke model in which there would be 5 major hubs (one per SHA sector) offering in-patient care, and a number of spokes offering day surgery. Harper, Phillips and Gallagher developed a DES model incorporating a geographical information system (GIS) in order to evaluate potential locations for the 5 hubs. The historical data contained postcode information for every patient. A total of 36 different categories of surgery were defined. In the model, individual patients were sampled at random points in the map using the GIS, and their travel times to a chosen center were then sampled from a travel time distribution, depending on the sampled mode of transport. The model was used for a series of scenarios in which the decision rules for selecting the center were varied. For example, in some scenarios patients had to attend the center for the SHA sector in which they lived, but in others cross-boundary flows were permitted and they were allowed to attend the geographically nearest center. Clearly, reducing the number of providers of in-patient surgery to 5 would save costs overall, but at the expense of considerable increase in travel times for patients requiring these services.

This approach has advantages over classical facility location methods, in which demand is assumed to arise at discrete nodes rather than at individually sampled locations. The approach has great potential in other specialties

where cost savings to the provider in centralization and rationalization need to be balanced against inconvenience to the patient. Increasing numbers of surgical procedures are routinely carried out as day cases, which formerly would have required an in-patient stay.

7.4 Summary

This is by no means an exhaustive or even vaguely systematic review of recent developments in healthcare modeling. However the examples do illustrate the breadth of these developments and they also illustrate one of the strengths of healthcare simulation, namely that in every case, it was a real-world problem which drove the theoretical development, and not vice versa. The need to capture or reflect some aspect of reality in order to make the model useful in practice has been the catalyst for the development of a new methodological approach.

8 CHALLENGES FOR THE FUTURE

8.1 Implementation

A recent literature search using the keywords “health” and “simulation” found over 3,400 references, and over a thousand for Emergency Departments alone. However, despite this proliferation of academic publications, and unlike manufacturing industry where a similar literature exists, there has been no widespread take-up of simulation by the healthcare industry. Countless projects are carried out by academics and published in academic journals, but these models are not widely taken up by other health providers.

The problems of getting models implemented are not new. Forty years ago now, Watt (1977) wrote a paper tellingly entitled “Why won’t anyone believe us?”, describing the difficulties of using simulation models to influence policy-makers. Wilson (1981) surveyed 200 simulation projects in healthcare but only found 16 which reported successful implementation. Common factors in these 16 included at least one author who worked at the institution concerned, a problem of high priority to that institution, external funding, and a detailed description of data collection. Twenty years later, a systematic review of healthcare simulation models (Fone et al, 2003) found 182 papers published between 1980 and 1999, yet very few examples of implementation. The authors said

“... we were unable to reach any conclusions on the value of modelling in health care because the evidence of implementation was so scant.” (Fone et al, 2003, p. 333)

8.2 Generalizability

One possible barrier to implementation is that of generalizability. All healthcare modelers stress the importance of working closely with clinical or managerial practitioners, in order to gain buy-in and acceptance, but this can lead to the “not invented here” syndrome in which a model developed for one hospital is seen as being specific to that hospital and of no immediate applicability elsewhere. This leads to the situation in which we have over 1000 published simulation models for Emergency Departments. Yet how different can such departments really be?

An exciting area for future research is the development of generic models to overcome this problem. Many people are currently working on this, but nobody has cracked the problem yet. It is possibly more a social, cultural and educational problem than a technical one.

8.3 Combining DES and SD

Another exciting possibility is the integration of simulation approaches to combine their relative strengths. If it were possible to develop a methodology with the advantages of both – the detailed, stochastic, individual patient level approach of DES combined with the whole-systems, strategic view of SD – this approach would have benefits far beyond healthcare. However healthcare provides the settings where such an approach could be readily applied. For example, returning again to emergency care, there are aspects of this system (such as the ER) where variability plays a major role and an individual-level approach is required. On the other hand, as we have already mentioned, the overarching system is hugely complex with many interacting parts and needs an approach capable of capturing the inter-relationships and feedback dynamics within it.

8.4 Conclusion and Apologia

I have attempted to give an overview of the state of the art in healthcare simulation modeling; a view that is of necessity idiosyncratic and personal. Moreover, it has mainly used examples from the UK as illustrations, although all the approaches and methods described here have been applied worldwide. I apologize to the many people whose excellent work I have omitted to mention, and in mitigation can only say that I have attempted merely to provide a flavor of this exciting field and that a truly systematic review of simulation in healthcare would probably be the work of a lifetime.

REFERENCES

- Ajzen, A. 1991. The theory of planned behaviour. *Organizational Behavior and Human Decision Processes* 50:179-211.
- Baines T.S. and J.M. Kay. 2002. Human performance modelling as an aid in the process of manufacturing system design: a pilot study. *International Journal Of Production Research* 40: 2321-2334.
- Becker, J.M.H. 1974. The health belief model and sick role behavior. *Health Education Monographs* 2:409-19.
- Brailsford, S.C. 2005. Overcoming barriers to the implementation of OR simulation models in healthcare. *Journal of Clinical Investigative Medicine*, 28:312-315.
- Brailsford, S.C., V.A. Lattimer, P. Tarnaras and J.A. Turnbull. 2004. Emergency and on-demand health care: Modelling a large complex system, *Journal of the Operational Research Society* 55:34-42.
- Brailsford, S.C., M.S. Rauner, W.J. Gutjahr and W. Zeppelzauer. 2007 (to appear). A combined discrete-event simulation and ant colony optimisation approach for selecting optimal screening policies for diabetic retinopathy. *Computational Management Science*. Published online at <http://dx.doi.org/10.1007/s10287-006-0008-x>
- Brailsford, S.C. and B. Schmidt. 2003. Towards incorporating human behaviour in models of health care systems: an approach using discrete event simulation. *European Journal of Operational Research*, 150:19-31.
- Brailsford, S.C., J. Sykes and P.R. Harper. 2006. Incorporating human behavior in healthcare simulation models. *Proceedings of the 2006 Winter Simulation Conference, Monterey, USA, December 2006*. ed. L. F. Perrone, F. P. Wieland, J. Liu, B. G. Lawson, D. M. Nicol, and R. M. Fujimoto.
- Coutsoudis A, Goga AE, Rollins N, Coovadia HM. 2002. Free formula milk for infants of HIV-infected women: blessing or curse? *Health Policy Planning* 17: 154-160.
- Dangerfield, B.C.. 1999. System dynamics applications to European health care issues. *Journal of the Operational Research Society* 50:345-353.
- Davies, R. and S.C. Brailsford. 2004. Screening for diabetic retinopathy, in: *Handbook of OR/MS Applications in Health Care*, ed. F. Sainfort, M. Brandeau, and W. Pierskalla, International Series in Operations Research and Management Science, Kluwer, Norwell, Mass., pp 493-518.
- Davies, R., S.C. Brailsford, P.J. Roderick, C.R. Canning and D.N. Crabbe. 2000. Using simulation modelling for evaluating screening services for diabetic retinopathy, *Journal of the Operational Research Society*, 51:476-484.
- Davies, R. and H. Davies. 1994. Modelling patient flows and resource provision in health systems. *Omega*, 22:123-131.
- Davies R., O'Keefe R.M. and Davies H.T.O. 1993. Simplifying the modeling of multiple activities, multiple queuing and interruptions: a new low-level data structure. *ACM Transactions on Modeling and Computer Simulation* 3: 332-46
- Dorigo M, V. Maniezzo and A. Colorni. 1991. The Ant System: An autocatalytic optimisation process. Technical Report 91-016, Dept. of Electronics, Politecnico di Milano, Italy.
- Even den, D., P.R. Harper, S.C. Brailsford and V. Harindra. 2005. System Dynamics modelling of Chlamydia infection for screening intervention planning and cost benefit estimation. *IMA Journal of Management Mathematics*, 16:265-279.
- Even den, D., P.R. Harper, S.C. Brailsford and V. Harindra (2007, to appear), Improving the cost-effectiveness of Chlamydia screening with targeted screening strategies. To appear in *Journal of the Operational Research Society*.
- Fone D, Hollinghurst S, Temple M et al. 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: 325-335.
- Griffiths J.D., N. Price-Lloyd, M. Smithies and J.E. Williams. 2005. Modelling the requirement for supplementary nurses in an intensive care unit. *Journal of the Operational Research Society*, 56:126-133.
- Gutjahr, W.J. 2003. A converging ACO algorithm for stochastic combinatorial optimisation. In Albrecht A and Steinhöfl K (eds.) *Proceedings of. SAGA 2003 (Stochastic Algorithms: Foundations and Applications)*, Springer, Berlin, 10-25.
- Harindra V., J.M. Tobin and G. Underhill. 2002. Opportunistic chlamydia screening; should positive patients be screened for co-infections. *International Journal of STDs and AIDS* 13: 821-825.
- Harper, P.R., Phillips S. and Gallagher J.E. Geographical simulation modelling for the regional planning of oral and maxillofacial surgery across London. 2005. *Journal of the Operational Research Society*, 56:134-143.
- Harper, P.R. and A.K. Shahani. 2002. Modelling for the planning and management of bed capacities in hospitals. *Journal of the Operational Research Society*, 53:11-18.
- Health Protection Agency. 2004. HIV and other sexually transmitted infections in the United Kingdom in 2003. Annual Report, November 2004.
- Jackson, RRP. Appointment systems in hospitals and general practice.1964. *Operational Research Quarterly* 15: 219-237
- Jun, J.B., S.H. Jacobson and J.R. Swisher. 1999. Application of discrete-event simulation in health care clinics:

- A Survey. *Journal of the Operational Research Society* 50:109-123.
- Klein R et al. 1985. The Wisconsin Epidemiologic Study of Diabetic Retinopathy. A comparison of retinopathy in younger and older onset diabetic patients. *Advances in Experimental Medical Biology* 189: 321-35.
- Lane D.C., C. Monefeldt and J.V. Rosenhead. 2000. Looking in the wrong place for healthcare improvements: A system dynamics study of an accident and emergency department. *Journal of the Operational Research Society* 51: 518-531.
- NHS, NSC Diabetic Retinopathy Screening. 2001. Resources for a national diabetic risk-reduction programme, www.diabetic-retinopathy.screening.nhs.uk/costings.html
- Rauner, M.S., S.C. Brailsford and S. Flessa. 2005. The use of discrete-event simulation to evaluate strategies for the prevention of mother-to-child transmission of HIV in developing countries, *Journal of the Operational Research Society* 56:222-233.
- Royston, G., A. Dost, J. Townshend, and H. Turner. 1999. Using system dynamics to help develop and implement policies and programmes in healthcare in England. *System Dynamics Review* 15:293-313.
- Schmidt, B. 2000. The Modelling of Human Behaviour; SCS-Europe, Ghent, Belgium.
- Sykes, J. 2007. Behavioural Healthcare Modelling: Incorporating behaviour into healthcare simulation models: A breast cancer screening example. PhD thesis, University of Southampton.
- Townshend, J.R.P. and H.S. Turner. 2000. Analysing the effectiveness of chlamydia screening. *Journal of the Operational Research Society* 51, 812-824.
- UK Department of Health. 1991. The State of the Nation. HMSO, London.
- UK Department of Health. 2005. National Chlamydia Screening Programme. www.dh.gov.uk
- Watt, K.E.F. 1977. Why won't anyone believe us? *Simulation* 28: 1-3.
- Wilson, J.C.T. 1981. Implementation of computer-simulation projects in healthcare. *Journal of the Operational Research Society* 32: 825-832

AUTHOR BIOGRAPHY

SALLY BRAILSFORD is Professor of Management Science at the University of Southampton, UK. She received a BSc in Mathematics from the University of London, and an MSc and PhD in Operational Research from the University of Southampton. Her research interests include simulation modeling methodologies, system dynamics, health service research and disease modeling, and the modeling of human behavior in healthcare systems. She is on the editorial boards of Health Care Management Science, the Journal of Modeling in Management and the Journal of Simulation, and is currently Vice President of the UK OR Society. Her e-mail address is s.c.brailsford@soton.ac.uk