Methods for modelling health service delivery: approaches, shortcomings and emerging opportunities.

Martin Utley, UCL Clinical Operational Research Unit

Preamble

This vignette has been commissioned by the MRC Methodology Research Programme Advisory Group to inform their thinking and planning in relation to methods for modelling health service delivery. In sections 1 and 2 I introduce the definitions I’ve worked to and the scope of this piece respectively. In section 3 I discuss briefly methods and areas of expertise in modelling and Operational Research that have relevance to health service delivery, particularly in the context of publically funded services such as the UK. I then set out in section 4 some areas where modelling is well-developed but underused and in section 5 other areas where research towards further development of methods might be fruitful. I give a couple of thoughts on the potential relevance of big data to modelling in section 6 before summing up in section 7.

In preparing this vignette I tried to avoid conscious bias towards the areas where my own research interests lie. Reading through what I’ve written, I realise I may have failed.

1. Definitions

To help frame the scope of this piece, it is perhaps worth introducing some working definitions.

1.1 Working definition of health service delivery

For the purposes of this vignette, I adopt the following working definition for service delivery

Those features of a service not relating to the clinical content of single patient-system interactions but rather to the deployment of resources decided upon within the system in order to effect these interactions, including mode of delivery, the staff groups and/or technology involved, the location of physical infrastructure, organisation and scheduling of services, the patient groups granted access, targeted or prioritised etc.

1.2 Working definition of model

Once, at a gathering of people in early middle age, my partner overheard me respond “mumble mumble modelling mumble” to a question as to what I did for work. She helpfully broke in to clarify that I do “mostly before shots”. This was funny at one level and, unwittingly, fairly accurate on another (see section on lack of evaluation in operational research applied to health care). I mention it here because the terms model and modelling mean different things to different people and indeed to the same people in different contexts.

In some contexts, a “model of service delivery” relates to one way of organising the delivery of a service; in others it can mean a system of equations constructed to describe the
operation and impact of one way of organising the delivery of a service. For clarity, in this vignette I will reserve the term model for this latter use – adopting as a working definition that proposed by Mike Pidd of Lancaster Business School in his excellent book “Tools for thinking” [1], reproduced below.

A model is an external and explicit representation of part of reality as seen by the people who wish to use that model to understand, to change, to manage and to control that part of reality.

Pidd explicitly avoids using the term “improvement” in his definition to acknowledge that an improvement from one perspective may constitute a worsening from another, equally valid perspective. I’ll almost certainly drift into talking of improvements to services so should state here that by “improvement” I mean “change to some purpose”.

1.3 Working definition of modelling

I reserve use of the term modelling to mean …

… the iterative social and technical processes engaged in by model developers working alone or in collaboration with other academic disciplines, clinicians, managers, and service users in the development, parameterisation, implementation and use of models.

2. Scope

For the purposes of this vignette, I’m taking “methods for modelling health service delivery” to encompass approaches from operational research and related disciplines that have been or could reasonably be used for the purposes of

Constructing models of the operation and impacts of a defined way of organising and delivering a defined health service in instances where the model constructed is entirely external to and independent of the organisational processes involved in the design and operation of the service delivered. Such modelling can be valuable in the normative evaluation of proposed or actual changes to service delivery by providing a framework for exploring the operation of the system under a wide range of plausible scenarios, and extrapolating from empirical findings in terms of outcomes and the evolution of system performance over time.

or

Constructing models for use in informing the design, operation or formative evaluation of services in collaboration with the organisation(s) concerned. This use of modelling is similar in many ways to the above but with greater focus on the features of the system that are within the sphere of influence of decision makers at different levels within organisational hierarchies and with a greater emphasis on the need for models to be explicable to, acceptable to, and potentially used by clinicians, managers and analysts in the health system.
Constructing models to illustrate the impacts of delivering services in new ways, taking as input the features of the clinical service to be delivered and of the target population but deliberately not constrained by the detail of current approaches to service delivery. While this approach to modelling can be criticised for a lack of verisimilitude, it can be valuable in communicating the insights that operational research have to offer those responsible for service delivery in the health service, challenging current practice. This use of modelling also has a role in what could be termed “utopic evaluation”, providing theoretical bounds on system performance as context for measured performance of current or proposed service delivery, or the performance estimated through other modelling approaches.

I do not cover modelling approaches that are largely or wholly inductive or statistical in nature such as clinical risk modelling nor, given the focus on modelling service delivery, do I cover the extensive use of modelling in health economic evaluations of health interventions within Health Technology Assessment, or the use of modelling in clinical decision support such as radiotherapy treatment planning.

3. Operational Research methods relevant to modelling service delivery

In their valuable taxonomy and structured review, Hulshof et al [2] classify the Operational Research and management science modelling methods adopted in addressing problems in health care management as falling mainly into the categories given below.

3.1 Simulation

Simulation modelling involves the construction of a modelled system where the operation or behaviour of each component is specified by mathematical expressions or a set of rules, as are the relationships between components and the flows between components of people, information or other entities. It has had wide application in health care. [3] Typically (but not exclusively) simulation is used when the nature of activities modelled and the relationships between components are such that the behaviour of the whole system over time and under different circumstances cannot be derived analytically. Learning about the system is generated through running the model and analysing the behaviour it exhibits, qualitatively or quantitatively through the statistical analysis of model output.

Deterministic simulation approaches such as system dynamics can be used to identify negative and positive feedback effects in complex health care systems. Stochastic simulation approaches involve the sampling of empirical or parameterised distributions to give instances of variable such as ambulance response times, patient inter-arrival times, length of stay etc. The simulation is then run a large number of times, sufficient for statistical analysis of model output across runs to discern between genuine behaviour intrinsic to the system and artefacts of sampling.
3.2 Mathematical programming

Mathematical programming is a set of optimisation techniques. The goal of the system concerned is considered to be described by an “objective function” – a mathematical expression relating to one or more key features of the system (for instance the number of patients treated, the cost of providing a service or the number of ambulance requests met within the target time). The decision maker is considered to be able to control various features of the system – called decision variables - with different sets of values for the decision variables giving potentially different system performance as defined by the value of the objective function. A set of equations relate the decision variables to the variables that feature in the objective function and link other aspects of the system considered important, with this set of equations constituting the model of the system concerned. Mathematical programming identifies the values of the decision variables that act to maximize or minimize the objective function with the choice of these values typically restricted directly or indirectly by a set of constraints (for instance fixed clinic hours, a minimum ratio of qualified to unqualified staff or the number of ambulance stations available).

3.3 Heuristics and heuristic search

Heuristic approaches to problem solving involve devising a set of rules to apply, often iteratively, to obtain acceptable solutions to problems, typically without any guarantee of optimality. They have particular value in areas of scheduling, creating staff rosters, and transportation in instances where mathematical programming approaches are impractical.

Heuristic search is an alternative approach to optimisation in which a set of rules is applied to guide an iterative search of the “solution space” of all possible combinations of decision variables for a particular problem. The intent is to find a sufficiently good solution within acceptable computational time, again without guarantee of optimality.

3.4 Markov modelling and other forms of compartmental modelling

Compartmental modelling approaches such as Markov modelling are frequently used to model the progression of patients through different health states over time. Similar models can also be used to model the flow of patients through states that represent physical systems or the progress of systems that evolve through a set of stochastic “birth and death” processes such as patient arrivals, transfers and discharges.

Although computationally efficient, the validity of Markov models for some processes in health care can be challenged and other classes of compartmental models that have less restrictive assumptions are available [see for example 4].

3.5 Queueing theory and related analytical models of patient flow

Queueing theory is the mathematical analysis of service systems characterised by the features of the population arriving to the system and the pattern of arrivals, the number of servers performing a particular task, the distribution of time taken for particular tasks, by whether some groups of patient are given priority over others and the other rules (imposed by the system or behavioural) by which any queues in the system are governed. Equations
are derived that allow the performance of a specified system to be explored in terms of queue sizes, the waiting times experienced by different priority groups, the utilisation of resources and the number of customers that baulk at the size of the queue or who join a queue but leave before being served. The analysis explicitly incorporates the uncertainty in arrival times and the times taken for the processes offered by a service, treating these as random variables governed typically by parameterised distributions.

For many real world systems, the assumptions necessary in terms of inter-arrival time and service time distributions to obtain tractable mathematical results do not hold. While this places some limitation on the validity of queueing theory results related to these systems, and perhaps a greater limitation on the acceptability of queueing theory to those working in the service, queueing theory remains a powerful tool to explore intrinsic trade-offs between different aspects of service performance in the face of patient-to-patient variability and uncertainty in terms of arrivals, treatment times and length of stay [5].

As an aside, it is worth commenting on the concept of the “objective function” that underpins mathematical programming and other approaches to optimisation such as heuristic search as this can be considered central to the world-view of operational researchers. In addition to the reductionism expected in any form of quantitative modelling, there is in the concept of the objective function an implicit assumption of model completeness – that the purpose of the system modelled can be expressed mathematically using only variables contained within the model. Perhaps more importantly, the concept of the objective function presupposes that the purpose of a system is or can be made explicit and is or can be agreed by the group of decision makers responsible for the design and operation of that system.

For operational researchers working closely with decision makers in the construction of a model, discussions around the objective function for the service considered can be an excellent vehicle for illustrating that systems designed to perform well on one desirable criterion may perform badly on another, that before making changes to a system it is a good to have a clear and agreed vision of what good looks like in terms of the intrinsic trade-offs between different criteria and for making plain to all involved in the models development and use the reductionist nature of the process and any consequent limitations to the use of model output in decision making. If working without that close engagement, modellers risk fundamentally misconceiving the problem they are working on and the decision processes they seek to inform.

4. Well defined problems, well-established solutions, little implementation

The adoption of OR solutions in healthcare is, in general, disappointingly low [3]. Within this context, there are some particular areas of modelling relevant to distinct, concrete, well-established activities in health service organisation and delivery where well-established modelling techniques and tools exist with clear potential to improve day-to-day service delivery on a substantial scale but where widespread uptake of tools has not been achieved in the UK NHS. I give some examples below before commenting on this lack of adoption of modelling.
4.1.1 Rostering

As one example here, the “nurse scheduling problem” is that of how to allocate members of nursing staff to shifts whilst ensuring that each shift has adequate cover and appropriate grade-mix, that rules on minimum rest periods between shifts and acceptable sequences of day and night shifts are respected and that the non-availability of certain staff for certain shifts, considerations of equity in terms of nights, weekends and bank holiday shifts and softer preferences among staff are accounted for. This has been studied by operational researchers for decades [6].

Using mathematical programming techniques and, where not possible, heuristic approaches, researchers have demonstrated the capability to produce higher quality rosters than are produced manually and in a fraction of the time taken. Yet each month thousands of health care staff devote several hours to this intrinsically difficult and arguably thankless task.

4.1.2 Scheduling

The appeal of scheduling lies in the scope for improving the performance of a system “simply” by changing the times at which appointments are planned [7,8], the order in which patients start a series of tests, or which patients are operated on when [9]. Given the possibility of using scheduling techniques (often based on mathematical programming formulations or heuristics) to deliver improvements with few explicit costs apparent (at least not to modellers) and the expansive literature in these areas, the lack of adoption of intelligent scheduling in health care should be a source of shame for operational research as a discipline.

With growing demand for services and a pressure to keep costs down, the potential gains associated with finding ways of successfully implementing better scheduling are huge. That said, it is clear that new approaches to this modelling work (if not new models) are required.

4.1.3 Transportation of staff, patients and equipment

With persistent interest in increasing the delivery of services at patients’ homes, the challenges of moving staff and equipment from location to location and the intrinsic “down time” associated with such deployment of these often costly and scarce resources would seem important.

Modelling methods adopted in the logistics and home service industries to plan routes and schedules for, say, the delivery of goods bought through an online super market or home visits by telecom engineers have potential application in planning routes and schedules for health service staff delivering home-based care. The literature contains examples of OR models adapted for use in home care services [see for example 10] but I am not aware of home care services adopting these approaches at scale in the UK.

There are a similar set of problems faced in patient transport services [see for example 11], with the scope to balance patient journey times, aggregated route lengths and the times between drop-off (pick-up) and the start (end) of patient appointments.
4.1.4 Location and allocation modelling

The geographical placement of health services can influence access, use and equity in health service delivery. Location and allocation modelling enables different options for the placement of clinics, ambulance stations etc. to be evaluated in terms of the average distance or time travelled by patients to access services or the proportion of demand a population that lies within a certain distance or travel time of a service. It also permits the optimal placement of one or more clinics within a region to be determined [12] although this is one area where it cannot always be assumed that an agreed and explicit objective function exists (see box). The availability of sophisticated GIS systems has enabled the development of location models that incorporate that incorporate time-varying patterns of demand and journey times.

Although notions of services having mutually exclusive geographical catchment areas are not consistent with current policies concerning patient choice, allocation modelling still has useful insights to offer, particularly for non-elective services and when patient journey times are considered to influence strongly choices by patients as to which services to use.

4.2 Thoughts on research and knowledge mobilisation activity required to improve adoption of well-established modelling solutions to well-established problems

In my view the key research challenges with relation to application of OR in the areas outlined above are not in the area of technical model development. Rather, research is required most urgently to understand how modellers can work in a way that promotes adoption of the techniques and solutions offered by Operational Research, to communicate more effectively the role of modelling, and to evaluate the impact on organisations of introducing these approaches.

There is some literature on ways of working that enhance the impact of Operational Research studies [see for instance 13, 14] but there would be value in operational researchers engaging with the literatures and research expertise in the areas of innovation and implementation in health care. The OR literature is heavily focussed on the technical content of models, with the modelling process given second billing at best. Partly this comes from the reward cycle at play in academic departments, partly from a desire to maintain that the models described can be applied generally with no further input from the academic team. Further work in the areas outlined above should perhaps only be done in collaboration with those who study and understand processes of innovation and organisational change and the role and limitations of quantitative analysis in decision making processes within health care.

Modelling remains an alien form of evidence to many in the health service and to other academic disciplines and creative work to improve the communication of what modelling is and how it differs from other research methods is needed. Specific work is needed in the area of how model output is presented, with there being a danger of misleading audiences when models don the clothes of statistical inference in reporting findings.

Evaluation of Operational Research models of service delivery is often restricted to validation – processes of mathematical proof to demonstrate that the equations that
underpin the model are correct within the assumptions that frame it, or showing that a computational model reproduces historic data or that the model is able to successfully predict behaviour of a system beyond the dataset used for calibration. More work is needed that evaluates the usefulness of modelling to those striving to improve service delivery, with OR models and the attendant modelling processes seen as a potentially complex intervention. This requires evaluations that build understanding of the organisational and professional barriers and facilitators to implementation as well as establishing an empirical evidence base on the costs and benefits of using modelling to inform change.

5. Emerging challenges and opportunities for modelling health service delivery

In addition to the areas suggested in section 3 where there are well-established solutions to well-established problems and the deficit is one of implementation, there are emerging problems in service delivery and methodological developments in OR that suggest areas where further research would be valuable. I give some examples below.

5.1 Incorporating behaviour of patients and staff and differential or state-dependent performance

Typically within models of service delivery, the influence of staff and patients in determining system performance are restricted to their seniority and specialty or their clinical presentation. Variability in, say, consultation times, adherence to protocols or compliance with treatment is viewed as uncertainty stemming from random processes rather than the result of potentially mutable or intrinsic individual characteristics. In many circumstances, such modelling assumptions are reasonable; when the performance or behaviour of individuals is outwith the sphere of influence of the decision makers to be informed by the model and if the system behaves as if all variability is random, there can be little value in including such detail. However, there are circumstances when the inclusion in a model of the scope for individuals to display intrinsically different behaviour and performance or for behaviour and performance of individuals to be influenced by the state of the system can give a richer understanding of the ways in which a system would behave under different scenarios.

As a simple example of state-dependent behaviour, the referral thresholds applied by general practitioners may change if a waiting list initiative in the acute sector reducing waiting times. Models incorporating this effect show an initial improvement in waiting times degrade as the system finds a new balance as opposed to a the permanent improvement suggested by models that treat demand as a fixed entity.

Models including the impact of non-adherence to protocols, slower than average consultation or operating times will produce different estimates of system performance if such deviations are assumed to be random fluctuations intrinsic to the clinical process or if they are considered to affect some practitioners more than others. Similarly, the cumulative effects of non-adherence to treatment or non-engagement with services will differ in models that treat these as purely random events or incorporate different groups of patients that display different behaviours.
Analytical models are able to incorporate many of these effects but are less flexible than simulation approaches, particularly when it comes to modelling the state-dependent behaviour of systems. Simulation methods are able to include sophisticated behavioural models where agents (typically staff or patients) within the model have encoded rules of behaviour that determine how they interact with the system and with other agents, with the focus being on the evolution of such systems over time and the variation in this evolution between different runs of the model.

As well as linking these modelling capabilities with research on professional and patient attitudes and behaviour, research is required to develop an understanding of when inclusion of such effects is essential to the effective use of modelling in evaluating current or proposed ways of delivering services.

5.2 Incorporating clinical outcomes to models of patient flow

Models of the flow of patients through health care systems such as those based on queueing theory tend to focus exclusively on metrics of process such as throughput, waiting times and resource utilisation. Within such models the clinical outcomes achieved through delivery of service are considered as implicit but there are circumstances where explicit inclusion of outcomes would enhance analysis. For instance, in many emergency response models, patients are prioritised by severity of injury and ways of delivering service assessed by the waiting time for different priority groups. A focus on outcomes may lead to different conclusions on what constitutes an effective response, especially when priority groups have different capacities to benefit from treatment and have different treatment times. There are examples of flow models incorporating clinical outcomes [see for instance 15] but more work in this area is needed.

Augmenting patient flow models with analysis of outcomes could also bring a new perspective to evaluations of the outcomes achieved at a system level through different ways of delivering services. With comparisons of outcomes in health care are generally based on the proportion of patients that achieve certain outcomes, a focus on flows of outcomes could give a richer picture of system performance.

5.3 Modelling transition between current and proposed ways of delivering services

Models used to evaluate ways of delivering a service typically focus on the behaviour of one or more configurations of a service once it has reached “steady state” – essentially when it has bedded into the wider system and is performing as it would be expected to in the long term. This notion of stability often brings a smile to the lips of those working in the UK health system and there is a valid concern that comparing the performance of systems at steady state gives an incomplete picture. One area where new research would be valuable is in the development of models that can facilitate improved understanding of the transition between one configuration of service and another, capturing the effects of patients and professionals adapting to a new system and the additional costs associated with the uncertain patient flows as transition occurs. Research in this area would link well with the work of health economists and others looking at the true processes and costs of service reconfiguration and would be timely given the recently restated desire of the system to shift care from the acute sector to community and home settings.
5.4 Modelling flows and interactions between organisations and sectors within health and social care

Either through the mechanisms whereby modellers form the collaborations necessary to work in health care or through the funding of projects by single organisations, the majority of models developed of service delivery in health care study the operation of a single organisation or of a single department within an organisation. With many important problems concerning the interface between different organisations and different sectors in health and social care and a perceived need to rethink the division of activity across sectors, there is a need for modelling work to build understanding of the opportunities for and impact of revising patient flows through a system. Such work should be done in conjunction with other disciplines including those health economists keen to understand the impact of proposed service reconfiguration on the constituent organisations within a system as well as on the system as a whole. Technical considerations aside, such research will need to address the question of how best to conduct modelling work with stakeholders whose organisational objectives may be mutually incompatible under prevailing funding mechanisms.

5.5 Patient and public involvement in modelling studies

It is fair to say that modellers working on problems of health service delivery are behind other groups working in applied health research in involving patients and the public in the design and conduct of their studies. Some work has been done to highlight the benefits of engaging with patients and public representatives early in the design of studies [16], such as ensuring that the choice of metrics by which a modelled system is to be assessed is informed by their perspective but more effort is required in this area.

Given that OR is still an alien form of evidence to many managers and clinicians working in health services, the challenges of communication that may impede meaningful patient and public involvement should not be underestimated. Also, the structure of modelling studies, where the study design process can be iterative and can overlap with any empirical work involved presents a different challenge to involvement than presented by research conducted to fixed prospective protocols. For these reasons, I think studies to understand how best to involve patient and the public in modelling studies would be valuable.

5.6 Health system resilience

In addition to potential uses in improving or evaluating health service delivery under normal operating conditions, modelling has a role to play in assessing the resilience of current or proposed ways of delivering services to surges in demand, staff shortages, losses of infrastructure or other plausible threats. Some simulation work has been done in the design-evaluation of emergency plans [17] but less attention has been given to understanding configurations of services that are intrinsically more or less resilient to seasonal or abnormal pressures. Of potential relevance here is the growing interest of applying to health care techniques of robust optimisation, intended to give solutions that work well over wide range of plausible scenarios.
6. A short section on modelling and big data

The prospect of large, interconnected data sets being more readily available for research purposes has many modellers salivating. For this reason, I include here a few words on why this excitement should perhaps be tempered. While access to linked data sets for use by modellers may become easier and the range of modelling studies considered feasible may broaden as a result, modellers will still need to work with those collecting and submitting data to understand the provenance of data and any limitations this places on the use of data within models; too often modellers forget that data are as much a product of current systems for delivering services as they are a reflection of underlying processes of disease, response to treatment, trends in demand etc.

That caveat aside, an increased use of modelling is arguably necessary (although not sufficient) for the availability of large, linked data sets to translate into better (or at least better informed) decisions about the design and operation of health and social care services. Much of the commentary on the benefits of big data are based on an implicit assumption that clinicians, managers, organisations and patients will know how to use bountiful linked data when it becomes available. In addition to models that can transform data into useful information, modelling work can help build the necessary understanding of current decision and organisational change processes that information feeds into and how these processes would need to change to make best use of the fruits of big data.

7. Research, Knowledge Mobilisation and Capacity Development

I hope that in this vignette I have managed to outline the different roles that modelling methods can play in the evaluation and improvement in health service delivery. For the areas of application outlined in section 4, my personal view is that what is needed most is a concerted effort on knowledge mobilisation, supported by interdisciplinary research into how operational research can be done in a way that facilitates effective knowledge mobilisation. The areas outlined in section 5 are where novel modelling approaches could assist the evaluation of and/or inform solutions to current or emerging problems in health service delivery.

Ultimately, the use at scale of operational research methods to improve services is reliant not only on research that is driven by that purpose and effective processes of knowledge mobilisation but on the development within the health care workforce of the capacity to engage effectively with, critically appraise and deploy modelling tools.

References


