The value of modelling and simulation in healthcare

A review of the evidence and some possible ways forward

Sally Brailsford
Jonathan H. Klein

Southampton Business School
University of Southampton
Executive Summary

This report is one of the outputs from the Festival of Evidence, a week-long conference in October 2014 organised by the Cumberland Initiative and funded partly by participants’ registration fees and partly by a generous grant from the Health Foundation. The Cumberland Initiative (http://cumberland-initiative.org/) is a collaborative group of UK universities, NHS organisations and commercial companies, whose bold vision is

“
To transform the quality and cost of NHS care delivery through simulation, modelling and systems thinking.
”
Overall conclusions

• Modelling and simulation are extremely useful techniques to improve decision-making, but whereas they have reached their potential in other domains, they have not yet done so in healthcare.
• There is an abundance of evidence, both national and international and in many forms, for the benefit of simulation and modelling in healthcare, but it is anecdotal, fragmented and diverse.
• Globally, there is little evidence of widespread successful ‘rollout’ of modelling interventions beyond the original client organisation.
• We believe that the best evidence is that which is composed of a mix of empiricist, rationalist and historicist sources. Assembling such evidence takes time, and we would generally expect the journey of evidential support to start from an empiricist, rationalist or historicist extreme, and move towards a more balanced profile.
• Evidence means different things to different people; and different people are convinced by different types of evidence.
• There are encouraging signs in the UK. For instance, modelling and simulation groups exist within several NIHR Collaborations for Learning in Applied Health Research & Care (CLAHRCs), a Welsh Health Board now employs five modellers under a joint initiative with Cardiff University, and several commercial organisations are supporting the use of modelling software by NHS users.

It is also essential that senior NHS management has an appreciation of the value of modelling and simulation, and there is evidence to suggest that a statutory requirement to demonstrate use of modern management approaches is helpful in bringing this about.
• Researchers (in the areas of both medicine and management) should take care to present their findings in ways which facilitate their adoption where appropriate by potential users.
• Evidence from modelling organisational innovation/interventions needs to be presented to clinicians and managers in a way that is different from how evidence for new treatments (RCTs or meta-analyses) is presented. This requires a culture change in the NHS.
• The modelling community needs stronger links with the implementation science community; why do some innovations spread and become common practice, while others never take off?

Research agenda

• We endorse the view which suggests that evidence-based practice should allow for more diverse evidence generation, and recognises the strengths of such generation methods.
• We need new ways of categorising and evaluating different types of evidence, especially for organisational interventions.
• We need more academic research into the ‘grey’ literature, and new, rigorous tools for analysing and interpreting its contents.
• We need to think about the nature of evidence for organisational innovation/interventions, and develop an equivalent to evidence-based medicine that has the same rigour and universal acceptability as the RCT.

Scaling up the use of modelling and simulation

• The cultural divide between the clinical and management points of view is a factor inhibiting the appropriate use of evidence in healthcare. Efforts should be made, not to eliminate the divide, but rather to create a pluralist ‘meta-culture’ in which each side appreciates the values and strengths of the other, and acknowledges its own limitations.
• For modelling and simulation to become part of the NHS management ‘toolkit’ it is essential that these skills are embedded within NHS organisations, and are not seen either as complicated techniques requiring a PhD, or as something that can only be done by specialist management consultants.
# CONTENTS

1. Introduction  
   1.1. What is modelling and simulation?  
   1.2. The Cumberland Initiative  
   1.3. The Festival of Evidence  
   1.4. Structure of this report  

2. Definitions of ‘evidence’ and framework(s) for understanding evidence  
   2.1. What is evidence?  
   2.2. The structure of evidence  
   2.3. Internal and external validity of methods for generating evidence  
   2.4. Hierarchies of evidence  
   2.5. Evidence-based practice and the primacy of the RCT  
   2.6. Different evidence for different stakeholders  
   2.7. Quality of evidence and strength of effect  
   2.8. Classifying evidence: the evidential triangle  
   2.9. Understanding evidence: discussions at the Festival of Evidence  
   2.10. Conclusions  

3. Analysis of evidence from the literature  
   3.1. Evidence from the academic literature  
   3.2. Other review articles  
   3.3. Evidence from the grey literature  
   3.4. Evidence from other work in the public domain  
   3.5. The Scenario Generator evaluation project  
   3.6. The challenge  

4. Analysis of evidence from the Festival of Evidence  
   4.1. Mike Carter, University of Toronto, Canada  
   4.2. Len Goldschmidt, Department of Veterans Affairs, USA  
   4.3. Carol Jagger, Newcastle University Institute for Ageing  
   4.4. Daryl Kor, Mayo Clinic, USA  
   4.5. Mark Mackay, Flinders University, Australia  
   4.6. Discussion sessions  
   4.7. The submitted evidence  
   4.8. Reflections and analysis of this evidence  

5. Overall conclusions and recommendations  
   5.1. Encouraging signs  
   5.2. Overall conclusions  
   5.3. Recommendations: scaling up the use of modelling and simulation  
   5.4. A research agenda  

6. References  

7. Appendices  
   7.1. List of keynote speakers  
   7.2. Pro-forma for collecting evidence from participants  
   7.3. Submitted evidence  
   7.4. Flipchart and Post-it feedback  
   7.5. Themes arising during the week  
   7.6. Attendees
1. INTRODUCTION

1.1. What is modelling and simulation?

Simply stated, modelling and simulation (sometimes known as operational research) is a ‘toolbox’ of analytical methods for improving decision-making within organisations, and achieving better system performance while reducing costs and risk. The toolbox contains a wide variety of methods, many (but not all) computer-based, and many of which have been in existence for decades. These methods come from a wide range of disciplines: mathematics, economics, psychology, computer science and management, to name but a few. They all have in common the concept of a model, which Michael Pidd (2009) defines as “… 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 in some way or other”.

It is not within the scope of this report to provide a comprehensive introduction to the whole toolkit of modelling and simulation. The RIGHT Workbook (http://www-edc.eng.cam.ac.uk/books/right/) provides an overview of most of the important simulation and modelling approaches as applied to healthcare systems: written in non-technical language for the lay reader, it includes examples and suggestions for further reading. Instead, we briefly present two particularly relevant approaches: systems thinking and simulation.

“A model is 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”

A map is a good example of a model. Neither of the two Google maps in Figure 1.1 are actually part of Slough: they are just images, “an external and explicit representation of part of reality”. Both of them leave out a lot of detail (especially the map on the left), but if “the person who wishes to use this model to understand... that part of reality” is faced with the problem of finding their way by road to the house marked by the red pin, then the map is a really useful aid. On the other hand, the map on the right is more useful to a person who is coming in to Southampton by plane and wants to spot a particular house. Thus all models are a simplification of reality, and the sort of model you choose depends on the problem or the decision the model needs to address. A key benefit of any model is that it allows the user to save time and/or money and reduce the risk of making a mistake in the real world, by first exploring the consequences of that decision in the ‘model world’.
Figure 1.1. Two types of model
At the heart of systems thinking, which is essentially a philosophy or practice rather than a single technique, is the idea that it does not make sense to look at any part of a system or problem situation in isolation, since the component parts of any system all interact and affect each other. It originated with the work of Ludwig von Bertalanffy in the 1940s (von Bertalanffy, 1969) and led to the development, by Jay Forrester at MIT (Forrester, 1960, 1961), of a modelling approach called system dynamics. The concept of systems thinking clearly has resonance in healthcare: firstly through the analogy of human physiology, and secondly from an organisational perspective. Human beings contain a gastrointestinal system, a cardiovascular system, etc., composed of interacting components and organs which work together to perform a specific function required to sustain life. Moreover all these systems themselves interact, so that a drug given to treat a cardiovascular problem can have gastrointestinal side effects. Likewise, the NHS is a massive, highly complex organisation consisting of a vast range of different services and locations, all of which are themselves systems consisting of people, resources and facilities. All of these sub-systems can affect each other; so for example the four-hour waiting time ‘target’ in hospital A&E departments can affect the performance of the Ambulance Service, and difficulties in accessing a GP can affect the demand for A&E. A failure to understand these system effects can lead to unanticipated and sometimes undesirable outcomes elsewhere in the system – for interventions which appear entirely logical and beneficial within one part of the system.

Simulation literally means pretending: it is a modelling approach which will be familiar to anyone who has played a computer game such as Sim City. An artificial version (a model) of the real world is created on a computer; as in computer games, powerful graphics can make such models look incredibly realistic, almost like watching a movie of the real system in operation. These models are then populated with data from the real-world system so that, for example, a simulation can represent the layout of an A&E department, with nurses, doctors, treatment rooms and patients. Simulation models use probability distributions to capture the individual variability between patients (and staff): no two patients, even if they are same age and gender and have exactly the same diagnosis, will have an identical experience in A&E. When the model runs, it is possible to see where queues build up and where there are bottlenecks in the system; the next step is to try out alternative strategies, e.g. for staffing or routing patients round the system, in order to overcome these problems. The key benefit of simulation is that such proposed changes can be tested out in a safe environment (a computer) rather than going ahead and trying them out for real via a pilot study, which is what the NHS tends to do now.

From the above brief description of modelling and simulation, it seems obvious that these approaches can be very useful for many organisational problems. Indeed, they have been successfully applied over the past 50 or 60 years in both the public and private sectors, in areas including manufacturing industry, transportation systems, service industries, government departments and the military (Jahangirian et al., 2010). Most large organisations have a functional department which over the years has been known under many names: operational (or operations) research, business intelligence, business analysis (or analytics), operations analysis… the fact that this function is seen by such organisations as sufficiently valuable to employ staff to perform it is a form of evidence that modelling and simulation are useful in practice. It is unthinkable for example that a motor manufacturer such as Ford would make any changes to its production lines without first building a simulation model to see the best way to implement this, and Ford employs a group of modellers whose job it is to build such models.
1.2. The Cumberland Initiative

The Cumberland Initiative (http://cumberland-initiative.org/) is a collaborative of universities, NHS organisations and commercial organisations, founded at Cumberland Lodge, a conference venue in Windsor Great Park, in July 2010. The director of The Cumberland Initiative is Professor Terry Young from Brunel University.

The participants at this first meeting came from more than a dozen universities and from a broad disciplinary background that includes design, engineering, management science, modelling, operational research, problem structuring and simulation. They recognised the potential for these disciplines to make a greater impact on healthcare in its hour of need. Their bold vision is “… to transform the quality and cost of NHS care delivery through simulation, modelling and systems thinking” (CI website). The Cumberland Initiative now includes around 30 academics from over 15 institutions in the UK, seven or eight commercial organisations offering either IT services or modelling and simulation consultancy services, and more than a dozen NHS partner organisations. The Cumberland Initiative has a building in Slough with a conference area which can be used for meetings, workshops, seminars and training courses, and a ‘warehouse’ area which can be used for major gaming simulations and mock-ups.

“to transform the quality and cost of NHS care delivery through simulation, modelling and systems thinking”

The Cumberland Initiative believes that transforming the quality of care through radically better processes and systems will save money through step changes in provision that also produce much better outcomes. These savings should dwarf those available through better selection of drugs or technology alone, and the impact should readily exceed 10 per cent of the total NHS budget.
1.3. The Festival of Evidence

One of the aims of the Cumberland Initiative is to develop evidence of what works. Randomised controlled trials (RCTs) are as old as the NHS and have transformed the selection of drugs and devices. Today’s infrastructure includes academic medical schools, the FDA, the Cochrane Collaboration, and in the UK, the HTA and NICE. Nothing comparable to an RCT exists for evidence related to care delivery or patient experience. There is little consensus as to what constitutes evidence. Britain also lacks systems to relate evidence to setting measures or measuring performance. This is a huge gap in our understanding and practice.

The Festival of Evidence, funded partly by participants’ registration fees and partly by a generous grant from the Health Foundation, was a week-long conference in October 2014, organised by the Cumberland Initiative and held at Runnymede in Berkshire. The aim of the conference was to begin to address the gap described above, by bringing international experts from the USA, Canada and Australia together with UK academics, and including contributions from commercial companies and NHS organisations, to analyse the evidence for the benefits of using systems thinking, modelling, OR and other risk-management tools in care delivery. Its purpose was to bring together people who have access to such evidence, and people who need to know how much of it exists in order to commission better research and services. The objectives were to:

- Gather evidence which is widely spread in a variety of places – the literature, the grey literature, commercial companies’ archives and people’s heads.
- Sift and analyse the evidence. There is no standard way of analysing service improvement, most of which is justified in terms of cost savings. Questions remain as to how real and sustainable many reported savings are.
- Engage the main stakeholders in the debate about using modelling tools in healthcare improvements.
- Demonstrate feasibility of concept: how could such tools be adopted at scale by the NHS?

1.4. Structure of this report

Section 2 considers the nature of evidence itself. How has evidence been understood and interpreted in the domain of healthcare? We discuss a number of frameworks or hierarchies of evidence found in the academic literature, and reflect on the new frameworks that emerged from the Festival of Evidence, especially in the context of evidence for the value of modelling and simulation in healthcare.

In Section 3 we review the evidence for the value of modelling and simulation from a number of perspectives: the academic literature, the ‘grey’ literature and other sources in the public domain, and our own personal experience based on a collaborative project on the use of modelling and simulation in the NHS, carried out on behalf of the (then) NHS Institute of Innovation and Improvement and the Simul8 Corporation. Section 3 concludes with a reflection on the nature of the key challenge: large-scale implementation of modelling and simulation within the NHS.

Section 4 presents a summary of the insights emerging from the week-long Festival, and the thoughts of the five international keynote speakers. We also discuss how some of the 23 examples submitted to the Festival of Evidence would fit into the frameworks described in Section 2.

Section 5 concludes the discussion and presents our thinking about the future: how might the use of modelling and simulation be scaled up across the whole NHS?
2. DEFINITIONS OF ‘EVIDENCE’ AND FRAMEWORK(S) FOR UNDERSTANDING EVIDENCE

2.1. What is evidence?

Evidence is material which supports a proposition or belief. Within the field of healthcare, evidence thus constitutes any material which supports a proposition or belief concerning healthcare.

Evidence in healthcare is usually discussed within the context of evidence-based practice or policy, which developed from the evidence-based practice movement. In evidence-based practice, decisions about what to do in particular situations are based – as much as is practically possible – upon evidence which indicates that the chosen action or policy at any level, ranging from the operational to the strategic, is the right thing to do to bring about desired effects and avoid undesirable ones. Cartwright and Hardie (2012) characterise the problem of using evidence as one of getting from “it worked there” to “it will work here” (p. 14). A wide variety of approaches and methods for generating evidence exist.

2.2. The structure of evidence

Cartwright and Hardie (2012) frame their characterisation of evidence in terms of cause and effect. “It worked there” implies that the adoption of a particular action or policy caused a specific effect in a specific situation; “it will work here” implies that the same action or policy will cause the same effect in the current situation. Cartwright and Hardie liken evidence to a three-legged stool, in which the three legs are: (1) an indication that the proposed cause-effect relationship works somewhere; (2) an argument to indicate that the same cause-effect relationship is valid here; and (3) an indication that other factors necessary for the cause-effect relationship to produce the effect are present here. If one (or more) of these legs is missing, then any evidential prediction that the cause will produce the effect collapses. Thus, they see evidence as having two necessary components: a causal principle, and support factors. The former is the direct causal effect being invoked when an action or policy is chosen to bring about a specific end; the latter are the contextual factors which are required for the former to work successfully.
2.3. Internal and external validity of methods for generating evidence

Many authorities, including Evans (2003), underline the importance of the internal and external validity of methods of generating evidence. **Internal validity** refers to the extent to which a study definitively relates observed outcomes to particular actions or policies in the particular context under study. Randomised control trials (RCTs), in which, ideally, all possible causes except the one under investigation are controlled for, score very highly on internal validity. A well-conducted RCT establishes a cause-effect relationship between action and outcome, although it does not explain the causal relationship.

**External validity** refers to the generalisability of a study to other cases. In comparing alternative methods for generating evidence, there is often a trade-off between internal and external validity. For example, while a well-conducted RCT may have high internal validity, inasmuch as it demonstrates a causal relationship between action and outcome under the specific conditions of the study, it will typically exhibit low external validity. This is because so much effort has gone into controlling conditions (to raise internal validity) that transferring results to cases where different or less well-controlled conditions prevail (as we might expect most real cases to be) might be suspect.

The extent to which the results of an RCT can be extrapolated beyond the immediate conditions of the study can vary from discipline to discipline. For example, it is generally understood that the validity of the results of a physics experiment will be transferable from the particular laboratory in which they were generated to just about any other location in the entire universe. The confidence with which physicists are able to state this is not due to the experimental methodology, but to external factors: essentially the combined experience of physicists (they have empirically observed such transferability) and their theoretical understanding (they have models of the physical world which explain why such transferability should be the case). Thus, within the domain of physical science, experience and theoretical understanding indicate that there are methods of generating scientific evidence in which both internal and external validity are high.

“...prototype of a large machine, which worked perfectly when tested in the workshop in which it was constructed, failed comprehensibly and entirely unexpectedly when placed in the environment in which it was intended to function. Retrospectively, engineers were able to theoretically account for this failure, and plan the extremely expensive redesign that was ultimately required. This sorry history is a vivid lesson in the dangers of inappropriately assuming external validity.”

Even within the world of physical science, however, such reasoning can sometimes let people down. One of the authors (JHK) recalls an instance many years ago in which he was asked to assist an organisation in which a highly complex and sophisticated prototype of a large machine, which worked perfectly when tested in the workshop in which it was constructed, failed comprehensibly and entirely unexpectedly when placed in the environment in which it was intended to function. Retrospectively, engineers were able to theoretically account for this failure, and plan the extremely expensive redesign that was ultimately required. This sorry history is a vivid lesson in the dangers of inappropriately assuming external validity.

Within even the comparatively ‘hard-science’ world of medical science, the external validity of RCTs (the preferred method of generating evidence in medicine) can be controversial. The extent to which the results of a drug trial RCT, for example, can be extrapolated to conditions other than those of the RCT (such as co-morbidities, social circumstances, or different ethnicities) may not always be clear from either experience or theoretical argument.

Cartwright and Hardie (2012) distinguish between the trustworthiness of evidence and its relevance. Most rating schemes for methods of evidence generation, they contend, focus on the trustworthiness of the evidence (how likely it is to be correct, or internally valid) but the relevance of the evidence (its external validity, how likely it is to be generally transferable to other contexts and, in particular, the one in which you happen to be interested) is often sidelined.

This lies at the heart of the RCT dilemma – a well-designed RCT is highly trustworthy, but its design may necessarily limit its relevance. Lewith et al (2002) make a similar point in considering research in complementary and alternative medicine, identifying rigour and relevance as the two major issues in treating material as evidence. While rigour is concerned with “the management of biases that threaten the valid conduct and interpretation of data”, relevance relates to the “use to which information will be put by specific audiences”. They go on to observe that relevance “involves values placed on different types of information” and that often these value judgements fall outside the realm of science (involving ethics, for example) (p. 4). The above validity dilemma may to some extent be evaded using meta-analysis (the formal combination of the results of several RCTs): if a variety of RCTs carried out in different contexts tend to point in the same direction, this is clearly going to raise external validity without compromising internal validity. Systematic reviews which combine the results of different types of study can similarly raise external validity. Systematic reviews can also enhance the internal validity of studies where internal validity tends to be low: for example, if a number of observational studies (typical higher in external validity, but lower in internal validity) broadly concur in their findings, this has the effect of raising the aggregate internal validity of the studies considered as a whole.
Evidence-based practice has introduced the idea of rating schemes for evidence, with methods of generating evidence ranked in terms of their value. Based on an extensive review of the literature, Carter (2010) identifies a six-level hierarchy of methods (Table 2.1).

This hierarchy is fairly typical, though Evans (2003) notes that the systematic review is “starting to replace the RCT as the best source of evidence” (p. 78). Cartwright and Hardie (2012) identify both meta-analyses and systematic reviews as being the supreme sources of evidence. Meta-analyses formally combine evidence from RCTs, while systematic reviews bring together evidence from a broad variety of sources, though these sources should be vetted, in some way, for reliability.

Evans (2003) argues that a limitation of conventional hierarchies of evidence is their sole focus on effectiveness of interventions. He suggests (p. 79) that the scope of evaluation should be broader, and presents a four-level hierarchy of evidence with three dimensions (Table 2.2): effectiveness, appropriateness (concerning “the psychosocial aspects of the intervention… its impact on a person, its acceptability, and whether it would be used by the consumer”) and feasibility (concerning resources and implementability). In this extended hierarchy, methods which earn low ratings (or no ratings at all) when considered in terms of effectiveness can score high ratings in terms of appropriateness or feasibility.

By contrast, the GRADE scheme (Balshem et al, 2011) evaluates not individual trials, but rather a body of evidence in aggregate. The scheme has four levels: high, moderate, low, and very low. It considers bodies of evidence based either on RCTs or on observational studies. The baseline rating for RCT-based evidence is “high”; and for observational studies the baseline is “low”. These ratings can then be modified if various conditions prevail. For example, the rating of a body of evidence based on clinical trials would be reduced to “very low” if there were “very serious” inconsistencies in the results, and “serious” risk of bias. The rating of a body of evidence based on observational studies would be raised to “high” if the observed effect was “very large” (p. 404).

Cartwright and Hardie’s (2012) general criticism of these schemes is that while they are “actually good at identifying policies that work, that is, policies that work somewhere”, they fail to identify “what works here” (p. 137). They privilege studies which have high internal validity at the expense of those which are strong in external validity: the results from highly-scoring methods are generally highly trustworthy but can be of limited relevance to the needs of those

<table>
<thead>
<tr>
<th>Level</th>
<th>Evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level I</td>
<td>Well-conducted randomised control trials</td>
</tr>
<tr>
<td>Level II</td>
<td>Poorly-conducted randomised control trials</td>
</tr>
<tr>
<td>Level III</td>
<td>Prospective/retrospective cohort studies</td>
</tr>
<tr>
<td>Level IV</td>
<td>Case-control study; cross-sectional study</td>
</tr>
<tr>
<td>Level V</td>
<td>Case series and other non-comparative clinical design studies</td>
</tr>
<tr>
<td>Level VI</td>
<td>Expert opinion</td>
</tr>
</tbody>
</table>

Table 2.1. A hierarchy of evidence (Carter, 2010, p. 69)
who have to make decisions about actions or policies. Implicitly they are informed by a physical scientist’s view of the world, where the transferability of results is taken as read.

We might also observe that they place the emphasis on empirical results. For example, in both Carter’s and Evans’ hierarchies, the only method that seems to acknowledge the value of theoretical evidence is “expert opinion”, rated at the lowest level in both hierarchies. It is true that theory is likely to influence choice of studies in the first place: for example, in medicine, researchers are unlikely to go the expense of trialling a drug with an RCT study unless there is already support, possibly in the form of theory, for the hypothesis that the drug will be effective. Nevertheless, Hjorland’s (2011) citation of the “historical swing of the pendulum between empiricism and rationalism in medicine” (p.1306) has resonances for us: at present the emphasis seems firmly placed on the empiricist position.

The attractiveness of the RCT is precisely that a well-designed and well-conducted RCT obviates the need for any theoretical account of the observed cause and effect relationship. Such an RCT rules out the possibility of any factors other than the factor under trial as possible causes of the observed effects, without any need to know anything about what those other factors might be. Cartwright and Hardie (2012) observe that there are other methods of inference, such as Bayesian network analysis, econometric techniques, and process tracing, which can deliver evidence from non-experimental data as rigorous as that from an RCT, but that for this to be so several assumptions about the nature of the data under analysis must be met. These requirements are not present when data is generated by RCT.

<table>
<thead>
<tr>
<th></th>
<th>Effectiveness</th>
<th>Appropriateness</th>
<th>Feasibility</th>
</tr>
</thead>
<tbody>
<tr>
<td>Excellent</td>
<td>Systematic review</td>
<td>Systematic review</td>
<td>Systematic review</td>
</tr>
<tr>
<td></td>
<td>Multi-centre studies</td>
<td>Multi-centre studies</td>
<td>Multi-centre studies</td>
</tr>
<tr>
<td>Good</td>
<td>RCT</td>
<td>RCT</td>
<td>RCT</td>
</tr>
<tr>
<td></td>
<td>Observational studies</td>
<td>Observational studies</td>
<td>Observational studies</td>
</tr>
<tr>
<td></td>
<td>Interpretive studies</td>
<td>Interpretive studies</td>
<td>Interpretive studies</td>
</tr>
<tr>
<td>Fair</td>
<td>Uncontrolled trials</td>
<td>Descriptive studies</td>
<td>Descriptive studies</td>
</tr>
<tr>
<td></td>
<td>with dramatic results</td>
<td>Focus groups</td>
<td>Focus groups</td>
</tr>
<tr>
<td></td>
<td>Before and after studies</td>
<td>Non-randomised control trials</td>
<td></td>
</tr>
<tr>
<td>Poor</td>
<td>Descriptive studies</td>
<td>Expert opinion</td>
<td>Expert opinion</td>
</tr>
<tr>
<td></td>
<td>Case studies</td>
<td>Case studies</td>
<td>Case studies</td>
</tr>
<tr>
<td></td>
<td>Expert opinion</td>
<td>Studies of poor</td>
<td>Studies of poor</td>
</tr>
<tr>
<td></td>
<td>methodological quality</td>
<td>methodological quality</td>
<td>methodological quality</td>
</tr>
</tbody>
</table>

Table 2.2. A hierarchy of evidence (Evans, 2002, p. 79)
2.5. Evidence-based practice and the primacy of the RCT

Hierarchies of evidence such as those above have generally been generated in the context of evidence-based practice (EBP). EBP grew out of the evidence-based medicine (EBM) movement, the formal origins of which were in the 1990s. According to Champagne et al. (2004), “while rapid acceptance of the practice of evidence-based medicine has been premised on the positive relationship between the use of clinical evidence and improved clinical health outcomes, a broader approach to evidence-based healthcare considers the use of clinical and non-clinical evidence for improving healthcare at multiple levels” (p. 3).

Mullen and Streiner (2004), reviewing EBP, explain that “EBP has been both heralded as one of the major advances in healthcare… promising to revolutionise both policymaking and practice… and excoriated as a development that will reduce professionals to mindlessly (and soullessly) following recipe books” (p. 111). However, they have a broadly optimistic view of the future of EBP as it matures. Interestingly, their discussion is based very largely on RCTs as the central pillar of EBP, citing limitations of EBP as being a shortage of RCT-generated evidence, and the fact that RCT evidence is statistical and therefore cannot guarantee a successful outcome in the individual case. However they conclude with an acknowledgement that “the original claims that practice must be based on the conclusions of RCTs and only RCTs have been softened in the face of reality to the use of the best available evidence” (p. 119) – in other words, RCTs are still at the top of the hierarchy, but evidence generated by other methods is at least acceptable if nothing better is available.

Others are more critical of EBP, at least as it is currently conceived. Gabbay and le May (2004), for example, observed that in a study of GPs clinicians “rarely accessed, appraised, and used explicit evidence directly from research or other formal sources”, and in practice developed what they termed mindlines, “collectively reinforced, internalised tacit guidelines, which were informed by brief reading, but mainly by their interactions with each other and with opinion leaders, patients and pharmaceutical representatives and by other sources of largely tacit knowledge that built on their early training and their own and their colleagues’ experience” (p. 3).

Wye et al. (2015) found a similarly catholic attitude to evidence in their study of the practice of UK clinical commissioning groups (CCGs). CCG commissioners claim to use a wide variety of sources of evidence to inform policy, and have a broad definition of what constitutes evidence. Evidence from local sources, however, can have more impact on commissioning decisions than material from other sources, suggesting that commissioners are particularly influenced by knowledge of what works locally. For a variety of reasons, commissioners tend not to use traditional academic research, finding it difficult to apply in practice. Indeed, there is a suggestion that traditional academic research can inhibit innovation, since research is often presented in a manner that is both context-free and somewhat equivocal. Rather, the kind of evidence that is often used to build cases for particular decisions includes doctrine (NICE guidelines, for example), and evaluations of local practice are considered valuable. The authors suggest that the demand for research is there, but it is not produced in a form which makes it easily usable by commissioners. The implication is that the onus is on researchers to present their work in a more policy-applicable manner. Although many of their recommendations concern presentation of their work, with a greater emphasis on rich narrative and contextualisation, they also suggest a broadening of research methods away from the emphasis on RCTs.

Some are robust in their criticism of RCTs. Cartwright and Hardie (2012) consider that their results by their very nature are incomplete, demonstrating only that an action was effective in one context, but not in another, and observing that in any case “in many social policy cases, you can’t do RCTs” (p. 132). Klein (2000) points out that an RCT “carried out in a pioneering institution does not necessarily tell us very much about its effectiveness in different settings” (p. 65). Carter observes that “not every RCT is high quality” (p. 77), and that RCTs can often be regarded as “ill-suited to meet evidentiary needs – specifically the comparison of effective interventions among patients in typical patient care settings, with decisions tailored to patient needs” (p. 78). She argues that in practice the evidence generated by observational trials may often be as or more valuable than that from RCTs.
2.6. **Different evidence for different stakeholders**

Different types of evidence may appeal to different stakeholders. Lewith et al (2002) identify five categories of stakeholder, and their differing evidence needs. Although their discussion is specifically in the context of complementary and alternative medicine, their observations seem to have more general applicability and are summarised in the following table:

<table>
<thead>
<tr>
<th>Stakeholder</th>
<th>Evidence need</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patients (and their families and carers)</td>
<td>Accounts of particular treatments working effectively: anecdotes, stories, case reports</td>
</tr>
<tr>
<td>Practitioners</td>
<td>Statistically valid evidence related to practical context: clinical outcomes data, systematic case series, practice audit evidence</td>
</tr>
<tr>
<td>Clinical investigators</td>
<td>Statistically valid evidence: comparative trials, RCTs</td>
</tr>
<tr>
<td>Laboratory scientists</td>
<td>Explanation of relationship between cause and effect to support clinical evidence: basic science evidence</td>
</tr>
<tr>
<td>Providers</td>
<td>Definitive proof that a treatment is effective: systematic reviews, meta-analyses, expert consensus reports</td>
</tr>
</tbody>
</table>

Table 2.3. Stakeholders and their evidence needs (summarised from Lewith et al, 2002, pp. 5-6)

Some commentators attribute the difference in the needs of stakeholders to fundamental differences in cultural assumptions. For example, Walshe and Rundall (2001) contrast the lack of evidence-based practice in healthcare management with the enthusiasm with which it has been adopted in clinical practice, and attribute the disparity to differences between the two contexts in terms of culture, the nature of research and evidence, and how decisions are made. They observe that, whereas the clinical culture is “highly professionalised, with a formal body of knowledge that is shared by all members of the profession”, healthcare managers “are a highly diverse group drawn from different professional and disciplinary backgrounds” and “personal experience and self-generated knowledge play a much larger part in determining how managers approach their jobs” (p. 439). Indeed, many clinicians are also researchers in some way or another, while only exceptionally would we expect to find managers who were. Walshe and Rundall characterise clinical decision-making as conforming to a rational mode of decision-making, whereas their characterisation of management decision-making suggests that it is far more oriented towards identification of goals and procedures for attaining them (in terms of a classification scheme developed by Choo, 2006). They are in little doubt that it would benefit from a movement towards the rational mode, but, although they advocate a change of culture which could in principle be engineered, it is less clear that the other sticking points – the nature of research and evidence, and the nature of management decision-making – are addressable.
2.7. Quality of evidence and strength of effect

The intrinsic quality of studies is potentially variable: in assessing an evidential source, it is important to take into account such quality. Carter (2010) lists a number of factors that can affect the quality of an RCT, and their likely effect on results. The presence of quality-reducing factors is not necessarily the fault of the investigators: for example blinding in RCTs is sometimes not practically possible. Carter cites evidence that lack of blinding “overestimates treatment effects by 15%” (p. 76).

Many medical studies fail to distinguish adequately between the statistical significance of an effect and its strength – Carter reports that “authors do not consistently provide an interpretation of the clinical importance of their results”, often failing to compare the strength of the effect they have demonstrated with the “minimally clinical important difference”, defined as “the smallest treatment effect that would result in a change in patient management, given its side effects, costs and inconveniences” (p. 74). More generally, research results are often reported in a fashion which makes it difficult for would-be users to relate it with confidence to the situations in which they might want to apply it (Wye et al, 2015).

2.8. Classifying evidence: the evidential triangle

We have developed, following terminology used by Hjorland (2011), the concept of the evidential triangle, the apices of which represent different types of evidence (Figure 2.1). Empiricism privileges empirically-based data beyond all other forms of evidence, while rationalism emphasises the importance of theoretical understanding, and historicism underlines the importance of the social context of knowledge. Lewith et al (2002) make the point that, by virtue of the values they attach to different kinds of evidential support, different interest groups privilege different kinds of evidence: patients, for example, may attach importance to anecdotal evidence that particular treatments work, whereas clinical investigators are likely to privilege RCTs and laboratory scientists want to understand the science underlying observed data. In terms of our evidential triangle, these varying perspectives indicate inclination towards, respectively, the historicist, empiricist, and rationalist apices. A shift towards the historicist perspective seems implied by Evans’ (2003) suggestion that appropriateness and feasibility should be included in evaluating interventions, as well as effectiveness.
While Hjorland draws attention to the empiricist emphasis which current EBP adopts, we might suggest that within the practice of management the evidential emphasis is on historicist material. Management practices are often, we believe, adopted because they were apparently effective elsewhere, with little or no understanding of why they might have been effective or, indeed, any questioning of whether their apparent effectiveness might have been due to some other, completely different, cause. It is interesting that while Cartwright and Hardie’s (2012) book is mostly directed against the limitations of RCTs as sources of evidence, one of their most vivid examples (that of a successful nutrition education policy for pregnant mothers in Tamil Nadu being transferred to Bangladesh, where it failed) is an account of how the success of a single case was inadequately understood – a failure of a historicist source of evidence poorly supported by theoretical (i.e. rationalist) understanding.

The empiricist apex of our triangle is concerned solely with supporting data that a given action or policy works. The RCT is the epitome of this apex; however, the flaw with the RCT is that, in Cartwright and Hardie’s (2012) terms, RCT evidence demonstrates merely that an action has worked somewhere. The rationalist apex of the triangle is concerned with a theoretical understanding of the situation; here, the drawback is that suitably robust theory may be absent and, until put into practice, is untested. A theoretical model of a cause-effect relationship would be rationalist evidence. The historicist apex of the triangle is concerned with the action in its full and unique context. Such evidence would be realised in a richly described case study.

Cartwright and Hardie (2012) discuss in detail strategies for getting from knowledge that a policy worked somewhere (as might be generated by an RCT or a case study) to confidence that a similar policy will work in the target context. In summary, their strategy amounts to detailed thought and deliberation about the transfer to elucidate both the likely causality of the relationship between policy and outcome, and the required contextual support factors that need to be present for the causality to be effective. In other words, they seem to be arguing for backing up empirical data with theoretical understanding. If we expand (as they arguably assume implicitly) that this understanding should stretch beyond the material world and into the social and personal worlds of stakeholders, then we are surely suggesting that, ideally, evidence should occupy the centre of the triangle; that actions and policies should be evidenced by empirical, rational and historic evidence (all pointing in the same direction, presumably).

We would therefore suggest that the ideal situation for using evidence is one in which several sources of evidence result in aggregate evidence for the effect brought about by an action or policy occupying the centre of the evidential triangle. In practice, we suggest, the development of a strong and robust evidential case for an action or policy is a journey to the centre of the triangle from a point near an apex. For example, one or more case studies (historicist) may suggest an effect which might then be supported by a theoretical model (rationalist), before being tested by one or more RCTs (empiricist). Thus, the development of the case in this example describes an arc from the historicist apex to the centre of the triangle, via the historicist-rationalist region.

We believe it is worth explicitly considering the role of modelling and simulation in the context of the evidential triangle. Broadly, we consider a model to be a theoretical construct which supplies rationalist evidence. However, we would suggest that within the characterisation of models as theory, there are different flavours. For example, a system dynamics simulation might have some empiricist flavouring as it might enable effects to be related to causes without any clear understanding of the nature of the causal links. It might also have some historicist leanings as it might include the richness of detail of a particular case, and might even reproduce real events. But all this would be knitted together within a broad network of relationships between variables which would constitute a theoretical model of the system under consideration. Thus we consider modelling and simulation to be rationalist evidential instruments.

The status of modelling and simulation as rationalist evidential instruments is interesting in the light of our arguments above, that both clinical and management practice lack this dimension. In the case of clinical practice, EBM strongly favours empiricist instruments while, in the case of management, EBP, stymied at the limited opportunity for use of RCTs due to practical circumstances, falls back upon the case study – the historicist dimension. Thus we argue that in the hybrid area of healthcare management, modelling and simulation has a lot to offer in terms of making evidential support for actions and policies three-dimensional.
2.9. Understanding evidence: discussions at the Festival of Evidence

The Festival of Evidence included discussions on the nature of evidence. Current practice concerning the way in which evidence is used to support actions and policies was characterised as a sequence of phases:

1. The sequence begins with awareness of a possible practice as a means of bringing about a desirable effect.
2. One or more individuals champion the practice.
3. These trials give rise to anecdotal evidence – stories, in effect – describing the successful application of the practice (assuming it was considered successful, at least by a champion).
4. The anecdotes give rise to corporate support.
5. Corporate support leads to the official adoption of the practice.
6. Eventually the practice becomes a cultural norm.

At different stages, different demands are made of evidence. Champions are inspired by evidence that is personally persuasive to them, and that in particular seems to suggest that a practice will solve real problems and enable them to exercise real influence on the system. Such evidence is generally clinical in nature. At the championing stage, projects to introduce new practices are likely to fail if the champion moves on for some reason.

Corporate support requires robust evidence of operational soundness and should, in particular, satisfy and convince non-clinical personnel of the policy’s effectiveness. Such evidence should appeal to at least one board-level potential champion. A mix of story and more formal evidence is needed at this stage to engage corporate support.

A variety of dimensions by which evidence might be characterised were proposed:

- **Stories versus statistics.** Well-constructed anecdotal evidence has strong persuasive power which can often sway decision-makers where more robust yet dryer statistical evidence can fail. In trying to capture the attention and support of policy-makers, it helps to back up statistical evidence with illustrative stories. The traditional source of gold-standard evidence is the RCT, which is regarded as the way in which irrefutable statistical evidence can be generated regarding cause and effect. Many other forms of evidence (such as accumulated case studies, for example) invoke weaker forms of statistical evidence. However, stories constitute anecdotal evidence and cannot be regarded as having any statistical validity whatsoever. But they can be very persuasive. Persuasiveness relates to two not entirely unrelated factors: the vividness with which the story is presented, and the extent to which the story makes sense. The latter factor appeals to rationality – however, it can be risky, since what appears to make sense is not always correct, and the reality may well be counter-intuitive.

- **In vivo versus in silico.** This dimension refers to the distinction between evidence that is carried as personal knowledge (often in implicit, non-expressed form), and that which is stored electronically (and explicitly).

- **Small problem well-done versus wicked problem partially-done.** The question here is, if you want to make a more persuasive impact in terms of evidence that demonstrates that a practice is effective, is it better to address a simple problem with a high degree of success (inviting the criticism that you are solving trivial, almost text-book problems, which have relatively little to do with reality) or a more complex problem with mixed, perhaps equivocal success?

- **Clinical versus management/operational.** Within healthcare, of course, this dimension is highly relevant. The point is often made that clinical and operational decisions are made within too completely different decision-making cultures, the former cleaving to principles of scientific
evidence, the latter more concerned with muddling through in recognition of the fact that it is not generally realistic to expect such evidence in real managerial situations. To some extent both cultures can be seen as having a limited view of the nature of evidence which, among other effects, serves to maintain the divide between them.

- **Effectiveness versus efficiency.** While effectiveness is about bringing about desired outcomes, efficiency focuses on the ratio of desirability of outcomes to consumption of resources. Thus a practice may be very effective but outrageously inefficient. Authorities also distinguish between the effectiveness of a practice, and its efficacy. An efficacious practice is one which is effective at bringing about a desired outcome when considered in isolation, but may not be so effective in the context of real situations.

- **Concrete answers versus confidence levels.** For many decision-makers, the problem with evidence is that it does not provide precise answers, but answers couched in terms of likelihood and confidence. (Even RCTs suffer from this perceived flaw, especially when they are small!) The problem is compounded by a frequent lack of appreciation of the difference between statistical significance and strength of effect. Thus, in extreme, a practice may have a highly statistically significant effect, but the effect may be so small as to be clinically trivial.

An ensuing discussion at the Festival of Evidence focused in particular on the importation of modelling from other domains into that of healthcare. Thus, there might be evidence of successful industrial modelling work using modelling methods not considered appropriate within healthcare. A few within the healthcare community might take on roles of champions to adopt such methods within healthcare. Assuming that such methods appeared to work, it might be gradually recognised that there could be something to them, and they might eventually diffuse within the healthcare domain and become accepted as healthcare modelling. For this process to take place, time is required, but also a strategy for communicating within the healthcare community. We might learn from successful communication strategies in other domains. The risk is that the methods that are being promoted will be seen merely as fads, and ultimately fail to gain widespread adoption.

The contrast was drawn between healthcare organisations which might have teams of modellers embedded within them, as distinct from those where little or no resources of this kind existed. In the former teams would work towards best clinical practice and patient outcomes, promoting both efficiency and effectiveness. However in the latter, modelling could be more risk-seeking, since as far as the future of modelling within such organisations was concerned, modellers had little to lose. Ad hoc piecemeal activity could be rather more speculative in its nature.

There was some detailed discussion of the contrast between stories and statistics, and the way they might relate to the transformation of organisations from those that were hostile to modelling to having a hunger for it. Ideally, statistical evidence should trump the evidence of stories. However, for some, stories may be the most effective way of convincing a hostile organisation that modelling might be worthwhile. We might hope that eventually organisations would come round to a less anecdote-based mode of decision-making.
2.10. Conclusions

This chapter has summarised much contemporary thought on evidence in healthcare and similar domains, and reported on the deliberations on this topic at the Festival of Evidence. Due to both resource and space limitations, we cannot claim it to be a full and comprehensive review of the literature, and some aspects of evidence have been treated in rather cursory fashion. Nevertheless, we believe it identifies the aspects of contemporary thinking on the topic that relate specifically to the purposes of this report. The conclusions of this chapter, which has summarised much contemporary thought on evidence in healthcare and similar domains, and reported on the deliberations on this topic at the Festival of Evidence, are as follows:

- The evidence-based practice (EBP) movement is strongly based on evidence-based medicine (EBM). However, much EBM is strongly biased in favour of the RCT as the preferred method of evidence generation. Even within the medicine, though, there exists dissent from this view, arguing that in many cases RCTs are either inappropriate for basing real medical practice on, or impractical to carry out. Within the field of EBP more generally, and in healthcare management in particular, such dissent seems even more appropriate. It would often be practically impossible to set up anything resembling an RCT, and to the extent that it would, such a trial might provide little credible evidence concerning practice in any specific real context.
- Thus, we endorse the view that suggests that EBP should be based on a view of evidence generation that allows for more diverse evidence generation, and recognises the strengths of such generation methods.
- In particular, we believe that the best evidence is that which is composed of a mix of empiricist, rationalist and historicist sources. To assemble such evidence takes time, and we would generally expect the journey of evidential support to start from an empiricist, rationalist or historicist extreme, and move towards a more balanced profile.
- In particular, we view modelling methods as providing a rationalist evidential perspective which can complement the empiricism of statistically-based evidence and the historicism of anecdotally-based evidence.
- It should be borne in mind, however, that evidence is assembled for different purposes and with different audiences in mind. Evidence is not just used to prove to stakeholders (e.g. clinicians) that a practice produces an outcome. It is also used to persuade stakeholders (e.g. policy-makers) that a practice should be adopted, and to reassure stakeholders (e.g. patients) that a practice is sound. Provided such persuasion and reassurance is done with integrity, we have no problem with it. Different types of evidence may be more effective in facilitating such uses.
- Researchers (in the areas of both medicine and management) should take care to present their findings in ways which facilitate their adoption where appropriate by potential users.
- The cultural divide between the clinical and management points of view is a factor inhibiting the appropriate use of evidence in healthcare. Efforts should be made, not to eliminate the divide, but rather to create a pluralist ‘meta-culture’ in which each side appreciates the values and strengths of the other, and acknowledges its own limitations.
3. ANALYSIS OF EVIDENCE FROM THE LITERATURE

3.1. Evidence from the academic literature

Since modelling and simulation have been widely and successfully adopted in so many other domains, one might reasonably suppose that modelling and simulation are also highly relevant for healthcare organisations. These approaches have indeed been widely applied by academics, and a huge number of academic papers have been published over the years. As part of the EPSRC-funded RIGHT project (Research Into Global Healthcare Tools) led by Terry Young, Brailsford et al. (2009) explored this vast research literature. To give an indication of its scale, the paper begins with the following statement:

Undertaking a review of modelling and simulation in healthcare is without doubt a Herculean task. This is a literature which, having carried out searches on consecutive days using the Web of Knowledge bibliographic database and the search string “((healthcare or health care) and (modelling or modeling or simulation))”, was found to be expanding at the rate of about 30 papers a day. A search carried out on June 21, 2007 using the Ovid search engine and the same search string resulted in 176,320 hits.

(Brailsford et al. (2009), p.134).

The RIGHT study adopted a novel review methodology, similar in concept to the approach of stratified sampling, to analyse the relative frequency of use of a range of modelling approaches in healthcare, along with the specific domains of application and the level of implementation. Using this methodology, the RIGHT review reduced this enormous literature to a representative sample of 342 papers, which were classified in three categories:

- **Implemented**, meaning that the model recommendations were used by a collaborating organisation and the outcome was reported in the paper;
- **Suggested**, meaning that the study was carried out in collaboration with (or using data provided by) a real-world healthcare organisation, but no practical implementation was reported;
- **Conceptualised**, meaning that the study was carried out solely by academics with no input from healthcare practitioners.

Figure 3.1 presents the findings, broken down broadly by type of modelling approach. It can be seen that the proportion of implemented studies is very small. Overall the proportion of papers in each category was Conceptualised (44.7%); Suggested (50%); and Implemented (5.3%).
3.2. Other review articles

Of course the RIGHT study was by no means the first (nor the last) general review on applications of modelling and simulation in healthcare; numerous review articles over the decades have been written on the topic. The most recent is Hulshof et al. (2012), who present a taxonomy of healthcare-related applications, classified by application area, containing over 460 references. One of the earliest, and possibly the first, comprehensive review of healthcare modelling was by Fries (1976), who compiled a list of 188 articles and grouped them into 15 categories according to their area of application. Papers were selected only if they used what Fries describes as “mathematical methods of modelling and solving decision problems that form the core of OR”. This bibliography was later supplemented with an additional 164 articles to make a total of 352 references (Fries, 1979). The review covers more than a dozen mainstream OR journals of that time. The author does not provide details of the full list of journals searched nor the selection criteria, but it is likely that the 352 articles cited represent a large proportion of the body of healthcare modelling literature at that time.

Papers were selected only if they used what Fries describes as “mathematical methods of modelling and solving decision problems that form the core of OR”. This bibliography was later supplemented with an additional 164 articles to make a total of 352 references (Fries, 1979). The review covers more than a dozen mainstream OR journals of that time. The author does not provide details of the full list of journals searched nor the selection criteria, but it is likely that the 352 articles cited represent a large proportion of the body of healthcare modelling literature at that time.

Two separate review papers on computer simulation projects were published a year later in 1980 by Tunnicliffe Wilson (1980). One article focused on applications to healthcare population problems and the other on healthcare facilities. Between them, they covered over 200 articles. A follow-up paper by the same author (Tunnicliffe Wilson, 1981) focused on implementation issues as the author reported that among all these articles, only 16 reported recommendations that had been acted upon.

Towards the end of the 1980s, Smith-Daniels et al. (1988) reviewed the literature pertaining to acquisition decisions, for example sizing of facilities and facility location, and allocation decisions, for example inpatient admissions scheduling. A few years later, Klein et al. (1993) presented a bibliography that included medical decision-making and simulation modelling with a focus on planning models. Jun et al. (1999) surveyed articles on the application of discrete event simulation modelling to healthcare clinics and systems of clinics, for example hospitals, outpatient clinics and emergency departments. The paper does not discuss the adopted review methodology and thus it is not possible to ascertain how systematic and wide-ranging this review was.

![Figure 3.1: Evidence for model implementation in the academic literature (Source: Brailsford et al. (2009))](image-url)
Fone et al. (2003) produced a systematic review of computer simulation modelling in population health and healthcare delivery. This article was almost certainly the first to adopt a rigorous systematic review process described in detail, and involved the screening of 2,729 references that eventually were reduced to 182 using inclusion criteria. The focus is entirely on discrete event simulation, and articles are grouped into four application areas. The authors comment that although the number of modelling articles has grown substantially in recent years, very few reported on outcomes of model implementation and so the value of modelling requires further research: “… we were unable to reach any conclusions on the value of modelling in healthcare because the evidence of implementation was so scant.” (Fone et al., 2003, p. 333)

More recently, Brailsford and Vissers (2010) published a review focusing on European applications, which considered three different levels for modelling (patients and providers; units and hospitals; regional and national) and contrasted the findings from the RIGHT review (Brailsford et al., 2009) with a similar analysis of the papers presented over 35 years at the annual ORAHS conference, the European Working Group on OR Applied to Health Services. Since ORAHS is not exclusively an academic conference, but also includes presentations by practitioners and consultants, the authors expected to find significantly more evidence of model implementation in the ORAHS data. However, although there was a slightly higher level, the difference was not striking, and even ORAHS shows a depressingly low implementation rate (just over 6%). Surprisingly, a higher proportion (50%) of ORAHS papers fell into the “Conceptualised” category than was the case in the general academic literature (44.7%).

Katsaliaki and Mustafee (2011) adopted a similar approach to the RIGHT study and surveyed 201 academic papers on applications of simulation in healthcare, published between 1970 and 2007. The findings regarding implementation are similar: only 11 of the papers reviewed (5.4%) report implementation of the model findings. Katsaliaki and Mustafee reflect on some possible reasons for this, referring to a review paper published back in 1978:

"Relatively few of the published healthcare simulation articles reported significant effects that simulation had on the healthcare system being studied. This may imply that, although authors document the model, the issues they model and the model results, there are few real implementation results to report. England and Roberts (1978) implied that the reasons behind this are either inadequate models that cannot quantify the impact of the human factor, or the diversity of authority in healthcare facilities, which thwarts the simplicity of a single administrative decision to change the system. The latter problem lies mostly in the political sphere. However, governmental bodies and other national or local council/agency fund a considerable number of studies (43% in our review).” (Katsaliaki and Mustafee (2011), p.144)

Naturally, many academics have reflected on this lack of implementation and adoption, have discussed the challenges of healthcare modelling, and have tried to analyse the reasons for success in those few cases where a model has been used in practice. Brailsford and De Silva (2015) suggest that access to both policy-makers and clinicians is important, and that one of the prime reasons for the successful implementation of their model was that De Silva himself was not only a modeller but also a practicing healthcare professional who could “talk the talk and walk the walk”. Others have developed frameworks for successful implementation: for example, Harper and Pitt (2004) propose a project life-cycle for modelling in healthcare, based on their many years of experience.
3.3. Evidence from the grey literature

One of the features of sampled papers considered by the RIGHT review was the "funding source" for the work described in its sample of papers, and found that "... the academic literature shows a huge contrast with the ‘grey’ literature, since only 4% of studies were funded by a health service organisation" (Brailsford et al., p137). The authors go on to say "It seems clear that many references to healthcare modelling exist outside the domain of conventional journal publications. Commercial and promotional literature, website references and unpublished presentations, for instance, contain much of interest in this field. The challenge is to find a viable means of accessing and referencing these sources, which by definition are not recorded in conventional bibliographic databases. Despite this we believe that ‘grey literature’ may be centrally important in revealing lessons to be learned from the implementation of models in healthcare, an area which seems to be sorely absent in most of the research literature reviewed here" (Brailsford et al., p139). The grey literature consists of promotional material, websites, project reports, client testimonials, blogs, media reports, and so on.

In conclusion, it is evident that this is a vibrant and popular area of academic research. Moreover, although most commercial organisations – large or small – do not publish in the academic literature, it is apparent that outside of academia, modelling is being undertaken both by business consultancies and also by analysts employed within public sector organisations. For example, at the start of February 2015 the Government OR Service GORS employed 52 analysts working for or on behalf of the Department of Health, and NHS England was in the process of setting up a similar unit. Anecdotally, for example through personal contacts (e.g. former students now employed as management consultants) or via the news media, we know that the NHS employs business consultants for a range of services which include some modelling or analytics. For example the consultancy industry online portal consultancy.uk reported that “The NHS spent £640 million on private sector management consultants in 2014, up from £313 million in 2010. The new figure, obtained by Professor David Oliver in a Freedom of Information request, shows that, contrary to the government’s 2010 vow to reduce the external consultancy costs, management consultants at the NHS are having a ‘field day’. Whether or not external consultants actually add long term value, still remains an open and hotly debated issue.” (consultancy.uk, December 2014).

Since several smaller consultancies are partner members of the Cumberland Initiative, we do know first-hand about some of the modelling and simulation work being undertaken for NHS organisations. However it is very difficult to get an accurate estimate of the true volume of NHS work being undertaken by the bigger commercial consultancies. The grey literature gives some indication, but there is no rigorous methodology for searching it. Moreover such material has not been peer-reviewed in the way that academic papers are, so there is no way of quality-controlling such documents. We believe there is a need for innovative tools which can take the output of a Google search and perform the same function for the grey literature that bibliographic search engines such as Web of Knowledge, Scopus or Ovid perform for the academic literature.
3.4. Evidence from other work in the public domain

Another organisation with very similar aims to the Cumberland Initiative, although with a different ethos and culture, is MASHnet, the UK Network for Modelling and Simulation in Health. The MASHnet website (http://mashnet.info) states:

“MASHnet was established in January 2005. It was originally funded for three years by the Engineering and Physical Sciences Research Council (EPSRC) with a research network grant. Subsequently it received some funding from the UK Operational Research Society. In November 2011 MASHnet became a not-for-profit Community Interest Company. The primary aim of MASHnet is improve the application of modelling and simulation in health and social care. To achieve this aim, it strives to bring together all the parties engaged in health and social care modelling and simulation to ensure the more successful realisation of these approaches in the delivery of services. The three main professional groups identified as central are the health and social care services, the research community, and the commercial sector. MASHnet provides a range of services and organises activities designed to improve effective interaction and collaboration between these groups.”

There is inevitably some overlap in membership between MASHnet and the Cumberland Initiative, but the two groups have different ambitions and approaches to achieving them. The MASHnet website contains a number of case studies, which incidentally include some by Cumberland Initiative members. In addition, MASHnet was recently commissioned jointly by Universities UK and the Health Services Research Network to produce a booklet describing five successful case studies, entitled Change by Design: Systems Modelling and Simulation in Healthcare. In all of these case studies, a demonstrable benefit was achieved. Four of the case studies were undertaken as collaborations between academics and an NHS client organisation; the fifth was essentially consulting, although consulting with a twist (it was a Scenario Generator application; see section 3.5). A PDF version of this booklet is downloadable from http://mashnet.info/wp-content/files/2014/08/140610-Modelling-briefing-FINAL.pdf. The booklet contains several quotations from NHS users of modelling, such as:

- “Modelling made it possible for us to better match our service to patient need, free up valuable resources for reinvestment in new services, and challenge delivery teams to change the way they worked. In collaborating to achieve a common goal we were able to cut service referral times from four weeks to less than ten days.” Sasha Karakusevic, Chief Operating Officer, North Bristol NHS Trust
- “Asking clinicians to evaluate the model teased out information that proved as valuable as the model itself. The entire process of engaging with stakeholders to develop the system model proved the key to drawing out what the real problems were.” Colin Burton, Service Planning and Business Analyst, University Hospital Southampton NHS Foundation Trust
- “A process simulation can not only be used to predict the acceptability and performance of the proposed solution, it can also be used to predict the cost of the solution and therefore show if the solution will also meet financial constraints. This means that...the whole quality-performance problem can be addressed from start to finish.” Simon Dodds, Consultant Surgeon, Heart of England NHS Foundation Trust
3.5. The Scenario Generator evaluation project

Scenario Generator is a commercial modelling tool produced by Simul8 Corporation in 2009 as part of a joint venture with the then NHS Institute for Innovation and Improvement (NHS III). In a nutshell, its purpose was to assist commissioners (originally Primary Care Trusts and now Clinical Commissioning Groups) in their strategic service planning. Scenario Generator is a simulation tool which allows users to design care pathways for specific conditions or patient groups, and then ‘feeds in’ demographic and disease incidence and prevalence data, together with projections for the planning period. This enables planners to test out the consequences (and costs) of commissioning specific volumes of work from the providers in their area. Scenario Generator was designed to be easy to use and not to require a highly technical analytical background. Each PCT was provided with one license for the tool, populated with their own local population and incidence/prevalence data, and was given free training and a year’s free support. At the time when this year of free support was coming to an end, MASHnet was commissioned by the NHS III to evaluate the success of the initiative. The research team was asked to discover who was using Scenario Generator, and for what; what factors had led to successful take-up and use, and what barriers had prevented it; and more broadly, what capacity for simulation and modelling there was within PCTs.

This study was subsequently published (Brailsford, Bolt et al., 2013) and the findings are highly pertinent to issues of ‘scaling up’ or reusing models from one site to another.

The research team performed 28 in-depth interviews with NHS staff across the whole spectrum, ranging from people who were using the tool successfully through to people who had never got beyond the training (although they still claimed to be interested). There were examples of highly successful use where the model had played a key role in strategic planning, and some cases where people had never really engaged with the tool. Overall, a very diverse picture emerged in terms of the types of people involved, the types of problem the tool was used for, and the success or otherwise in terms of adoption. Some key common success factors were:

- **A champion** (who need not be a technical person, in fact it was often better if they were not) who was able to enthuse their colleagues and be almost evangelical about modelling;
- **A business-critical problem to work on** – otherwise key people would find excuses not to attend meetings and key discussions would not take place;
- **External factors encouraging PCTs to use modelling** – at the time, World Class Commissioning was in force and this explicitly mentioned modelling as a necessary skill;
- **Sustainability of expertise** – if all knowledge of the model resides in one person and this person leaves the organisation or changes role, then the model will gather dust.

Barriers to adoption included time and capacity; people were too busy ‘fire-fighting’ and saw modelling as a nice-to-have rather than a key priority which would help them prevent the fires in the first place. Lack of analyst capability was raised as an issue: even though Scenario Generator was designed to be easy to use, it was still a step too far for many PCT analysts, who had no concept of modelling or training in any aspect. This contrasts strongly with industry, where most people in similar positions would have engineering backgrounds. Moreover, the data required by the model exposed some data availability issues in some PCTs, although the respondents mainly said it was very useful to have these issues raised!

In terms of scaling up the adoption of modelling and simulation, a number of facilitating factors (i.e. carrots and sticks) emerged from the Scenario Generator evaluation project. Possibly the greatest ‘stick’ was the role of World Class Commissioning in raising senior management’s awareness of the value of modelling, but the ‘carrots’ included the establishment of a User Group, a network of NHS users who met periodically at the NHS Institute but were also able to contact each other informally and ask questions of each other (“How did you get Scenario Generator to do X or Y?”) that they might have felt embarrassed to ask Simul8. Another ‘carrot’ was the fact that Claire Cordeaux, the Simul8 employee who led the Scenario Generator project, was herself a former NHS employee and was regarded by users as ‘one of us’ and not an ‘IT geek’; rather like De Silva in the study reported by Brailsford and De Silva (2015). It seems that for modelling and simulation to be accepted within the NHS, it needs to be something that is done by NHS staff, and not to or for them.
3.6. The challenge

It is unarguable that a huge amount of modelling and simulation applied to healthcare systems is being undertaken. However, the story is one of fragmentation and reinventing of wheels. The literature shows that a plethora of small ad-hoc research/consultancy projects are carried out by academics in collaboration with a hospital or some other healthcare organisation, and often published in OR journals, but it is very rare for these to spread beyond the original client organisation. In the commercial world, the “holy grail” for consultancies is to be able to sell the same solution to client after client – yet even they report the same problem. It is almost as if each time, modeller(s) and client(s) have to go through the painful process of problem understanding, designing a conceptual model, dealing with the politics and overcoming the challenges of collecting data, before a model can be accepted as useful in a particular situation.

Our experience, after 25 years of working in this field, is that the “not invented here” syndrome plays a big part in this. In 2007, as part of the RIGHT project, we discovered 1,008 references to different simulation models of Emergency Departments. It is hard to believe that every single Emergency Department is completely unique and has nothing in common with any other one. We often hear comments like “Very interesting, but our casemix is totally different from hospital X” or “Ah, but of course our X-ray department is organised quite differently from theirs”. Data availability also plays a part: many hospitals have legacy IT systems that do not “talk to” each other. There is also a cultural issue, in that modelling can be perceived as treating individuals as inanimate objects (widgets) in the computer, whereas clinicians are trained to treat each patient as a unique human individual. It is interesting to note that public health specialists, who are trained to think at population level, do not have this problem; and some of the success stories with Scenario Generator came from users with a public health background.

Another interesting parallel is the role that modelling has played in health technology evaluation. The requirement of NICE for the economic evaluation of any new treatment in terms of cost per quality adjusted life years has undoubtedly led to the universal acceptance of modelling as a mainstream methodology for the assessment of cost-effectiveness.

Sadly, there is no equivalent statutory requirement on NHS management to perform a cost-effectiveness analysis of any proposed organisational change! The challenge facing the Cumberland Initiative, and all of us who wish to see the NHS realise the benefits of modelling that have been achieved in every other sector, is this: how can we convince people of these benefits? What does it take to persuade an NHS decision-maker to use modelling and simulation? What evidence is needed, and what form should this evidence take? Do different people need different types of evidence? These were among the questions discussed during the Festival of Evidence.
4. ANALYSIS OF EVIDENCE FROM THE FESTIVAL OF EVIDENCE

The highlights of the Festival of Evidence were the five keynote talks by well-known international experts: one from Canada, one from Australia, two from the USA and one from the UK. These ‘big name’ talks were clearly a major factor in attracting such a large number of participants. In addition, two commercial partners of The Cumberland Initiative (Claire Cordeaux from Simul8 Corporation, and Peter Lacey from the Whole System Partnership) led plenary discussion groups.

4.1. Mike Carter
University of Toronto, Canada

Mike’s talk was entitled *Major challenges in hospital modelling, and tips for meeting them.* Mike is an industrial engineer with decades of experience in healthcare modelling, and is Canada’s leading researcher in this field. He has worked with dozens of hospitals and healthcare organisations, and has developed one particular simulation model for patient flow and bed capacity planning which has been applied in several hospitals.

His talk consisted of a reflection on ‘lessons learned’ from his many previous projects. The first lesson related to the need for a ‘crisis’: if there is a real, critical problem, people will find the time, money and data. However, they do need to believe that your model will solve their problem. Mike described two similar models, the first of which is still waiting to be used, whereas the second happened to address an urgent business-critical issue for one particular hospital and led to Mike and his group being hired to develop it further. “Never waste a good crisis” was an expression used by one member of the audience, who works for a Clinical Commissioning Group. Mike’s second lesson related to the need for a strong champion: he said all his successful projects had involved an engaged clinical partner, although this brings risks if the champion should move jobs (the Scenario Generator project described in section 3.5 found the same thing).

The third lesson related to data: many Canadian hospitals lack qualified analyst capabilities. Things are improving, but many hospitals only carry summary level data, and the details are in dozens of subsystems that do not communicate, or only exist on paper.

Many systems are still scanning reports, requisitions, referrals, etc. into their ‘electronic’ system in which there is no search or summary capability. There is no reason to believe that Canada is significantly different from the UK. Mike’s fourth lesson was about culture: modellers often try to apply tools that worked in industry. The problems are the same, but the culture is very different so the solutions need to be different too. His final lesson was about managing expectations: the adage “all models are wrong, but some are useful” is relevant here. People often have totally unrealistic assumptions about the power of a model to predict the future. However, models are just scenarios: if the following assumptions are all true, then the modelling outcome appears highly likely. Mike’s view was that it is not the role of the modeller to predict the veracity of the assumptions, but to ensure that models are close to reality if the assumptions are correct.
4.2. Len Goldschmidt
Department of Veterans Affairs, USA

The topic addressed by Len was **Taking a systems approach to improving patient pathway: the VA Care-Coordination Tele-health Model.** The US healthcare system differs in many respects from the UK, but many of the challenges are the same.

The problem Len talked about was the challenge of providing a uniformly high-quality service across a huge geography to all military veterans and their families in terms of access to care services. This was achieved through the use of a tele-health system in which consultations could take place remotely, and diagnoses be made through the use of technology such as remote digital imaging, enabling patients to have access to specialist services even though they were not physically in the same hospital (or even city). The results of a controlled trial for patients with complex chronic medical conditions, in which the tele-health cohort was compared with a cohort who received normal care, were impressive: pharmacy costs were 22% higher in the tele-health patients vs 15% for the control group, suggesting a greater emphasis on medication compliance among the former. However, Emergency Room and hospital admissions decreased in the tele-health group, although Medicare analysis showed that costs had not shifted. Most tellingly of all, the mortality rate was 9.8% in the tele-health group vs 16.58% in the control patients.

Len also described the use by VA of “Store and Forward Teleretinal Imaging” for patients with diabetic retinopathy. Again, this was found to be highly cost-effective. The observable trends identified in the screening population since the implementation of teledicine screening were:

- the number of known diabetic retinopathy cases has increased;
- the overall age of patients receiving screenings has decreased;
- the percentage of non-whites receiving screenings has increased;
- the average number of miles travelled by a patient to receive a screening has decreased;
- teleretinal screening participation is increasing.

Len concluded with a discussion of the key role to be played by healthcare information systems in a world where the volume of generated data is breathtaking.
4.3. **Carol Jagger**  
Newcastle University Institute for Ageing

Carol is AXA Professor of Epidemiology of Ageing at Newcastle, and the title of her talk was *Whole systems thinking for complexity and frailty: challenges and guidance*. Her research programme spans demography and epidemiology with a focus on mental and physical functioning in ageing. She is the leading UK researcher on healthy life expectancy.

Within Europe she is on the Scientific Advisory Board of the Joint Programming Initiative ‘More Years, Better Lives’, and she has advised the Office for National Statistics and the Scottish Public Health Observatory on Healthy Life Expectancy. Carol’s simulation model SIMPOP, linking health and disease scenarios to future projections of disability, provided evidence to the House of Lords report ‘Ready for Ageing’. Her talk at the Festival of Evidence concerned care for the frail elderly, and the links between the health and social care systems in terms of coping with disability. She argued that projections of future need for health and social care have generally assumed constant age-specific prevalence of disability, but disease is at the start of most conceptual models of the disablement process and, over time, substantial reductions in mortality from CHD and stroke have occurred. On the other hand, increases in obesity are projected to continue, impacting on CHD, stroke, arthritis, vascular dementia, and diabetes. There is a need for models incorporating multiple diseases as risk factors and treatments may affect more than one disease. She described the SIMPOP model and showed the results of some scenarios in which these trends either continue, or change.

SIMPOP is an excellent example of a model which has produced evidence that has had a major impact on policy. This accords with experience from the previous surveys and reviews, in which the fields of public health, epidemiology and health economics (through NICE) use evidence from modelling, and to a lesser extent simulation, routinely. The challenge seems to be, how to translate this across to the more practical aspects of healthcare delivery?
4.4. **Daryl Kor**  
**Mayo Clinic, USA**

Daryl gave a talk entitled *Taking a systems approach to improving healthcare delivery: challenges and ways to meet them*. Daryl is a critical care physician at the Mayo Clinic. The Mayo has an international reputation not only for excellent healthcare but also for its groundbreaking use of information technology solutions implemented via mobile devices: for example, it has created an iPhone/iPad app called Synthesis Mobile which integrates hundreds of its health IT systems.

Daryl discussed the role of electronic information systems in the Mayo to provide relevant, timely information to physicians. The USA and Europe have seen rapid advances in medical science and technology and the increasing complexity of healthcare. However (even in the USA) there is still a ‘cottage-industry’ structure and acute-care orientation of the delivery system, for a patient population that predominantly needs chronic rather than acute care. The structure of the US market for healthcare services supports innovation in procedures, drugs, devices and equipment, but remains indifferent to innovation directed at improving the quality and productivity of care delivery. This is due in part to persistent underinvestment by the healthcare delivery sector in information and communications technology, and the inability/unwillingness of the healthcare sector to take advantage of engineering-based systems-design, -analysis, and -management tools that have transformed other sectors of the American economy.

Daryl illustrated Mayo’s innovative approach to healthcare systems through an information system which has had a huge impact on transfusion practice. Recent evidence from the clinical literature suggests that conservative transfusion of red blood cells not only achieves better patient outcomes and is safer, but is also cheaper. However, in practice, Mayo doctors were continuing to use the more traditional protocols. This was not because they were unaware of the literature, but because the clinical algorithm for determining how much blood to transfuse was so complicated. It was a question of information overload: Daryl said doctors were "drowning in a sea of data". For example, over 748,000 data points per day were recorded in a 24-bed intensive care unit! The task of changing clinician behaviour was achieved through designing a system which provided relevant clinical information to the physician about each individual patient, in a comprehensible form, at the point of treatment. A “datamart” and a data visualisation tool were used in a new information system at the Mayo, resulting in savings of $25m and improved patient outcomes. Daryl gave several other examples of Mayo’s use of information technology, and argued that a multi-disciplinary approach is essential to optimise the value of these increasingly powerful and expensive electronic resources.
Mark Mackay
Flinders University, Australia

Mark’s talk was entitled Setting metrics and making service provision more effective and efficient: challenges and guidance. Mark is an operations research expert based in the Faculty of Medicine, Nursing and Health Sciences with many years’ experience of working with Australian hospitals.

He began by explaining that Australian hospitals face very familiar challenges as those in the UK: an ageing population, pressures to operate at 100% capacity (or more), a constant churn in senior management, worries over financial sustainability, and a vast number of often contradictory key performance indicators (KPIs) and metrics. There is also a desire to ‘stay off the front page’, but despite this rarely a day goes by without some media story about ambulance diversions or overcrowded Emergency Departments. The same issue of drowning in data arose in Mark’s talk: he described a plethora of management dashboards, differing from state to state (and hospital to hospital). However the problem with dashboards is that even though the information is regularly updated, the dashboard merely identifies the existence of a problem and doesn’t provide insight as to how problems may be fixed.

Mark discussed the way that KPIs can drive behaviour, and quoted Deming’s famous statement that “one of the seven deadly diseases of management is running a company on visible figures alone”. Often KPIs rely solely on averages, which any first-year statistics student knows are misleading.

More sophisticated data analysis can alleviate some of the problems but still does not enable testing of potential solutions. Mark then quoted Henry Mintzberg, a well-respected management researcher who has looked at healthcare organisations over many years, who recently stated that the hospital represents the most complex form of human organisation! Mark presented a compelling argument for ‘design thinking’ and the use of modelling, simulation and systems thinking to design smart KPIs which will transform healthcare delivery. He concluded with a contribution from Stephen Duckett on the role of stories and anecdotes in bringing about organisational change, a recurring theme during the Festival of Evidence. “Although anecdotes help to sell policies, they shouldn’t be the basis of policy development. If they are, they will almost certainly distort policymakers’ perceptions and start them down the wrong paths. Data should be used to... model the effects of new policies. Organisations need to invest in the mindset and skills to use data in policy, and have the mandate to do so.” (Duckett, S (2014). Forget the co-payment... Seven tips for an affordable, quality health system. The Conversation, 19 August 2014.)
4.6. Discussion sessions

Finally, there were two presentations from commercial members of the Cumberland Initiative, **Claire Cordeaux** (Simul8 Corporation, UK) and **Peter Lacey** (Whole Systems Partnership, UK). Both gave a short illustration of their work, and then led a group discussion.

The Whole Systems Partnership (WSP) theme was understanding population health across the spectrum of needs and over time for the purpose of future workforce planning. Claire from Simul8, meanwhile, presented a case study on screening for hepatitis C and then led a practical hands-on workshop on using Simul8 models. Two of the examples of models submitted as evidence (see Appendix 7.3 for a complete list) were actually from local authorities working with the WSP, and the majority of the discrete-event simulation models in Appendix 7.3 used Simul8 as the modelling tool.

Both companies have many NHS and local authority clients and users of their software. They work both as ‘straight’ consultants and also in partnership to support clients in building their own models. In the case of Simul8, the Scenario Generator tool described in section 3.5 is an example of this approach. The Whole Systems Partnership has established a Workforce Collaborative consisting of about 15 Local Education and Training Boards (LETBs), in which the member partners are supported by WSP to develop their own system dynamics models to help with workforce planning. Both companies share the philosophy that the more people who use modelling and simulation, the better. They exemplify the involvement and role of the smaller commercial organisations within the Cumberland Initiative. However, the examples they described perfectly capture the difficulty of finding out what is going on, since this sort of work rarely gets published.
4.7. The submitted evidence

Each of the keynote talks, and the two discussion groups, were interleaved with short presentations of some of the submitted evidence in Appendix 7.3. Clearly, these 23 examples represent a minuscule proportion of all applications of simulation and modelling in health. However, interestingly, only seven of these examples involved academics. Two were entirely instigated by academics and merely used data provided by a real hospital for illustrative purposes (what the RIGHT project would have called suggested). These were:

- an outline research proposal from Pam Abbott of Brunel University for a patient-centred IT system ("Well-being care maps") which will be developed and piloted with a group of cystic fibrosis patients;
- a methodology developed by Jim Methven of the University of Manchester for automatically generating patient pathways in Emergency Departments, intended to form the input for a simulation model, using time-stamped patient data collected for the 12 months April 2012 - March 2013 for 98,235 ED attendances at Manchester Royal Infirmary. However, despite the automated time-stamping a significant proportion (56%) of patient pathways could not be fully mapped, due to data quality problems. Therefore qualitative clinical input is still required, meaning that the resulting DES model will be local to MRI and is less generalisable than originally desired. Clearly, the methodology is generally applicable and the dataset has great potential.

Two further case studies, which the RIGHT project would have called implemented, were led by an academic but conducted in close collaboration with a hospital partner and then actually used by the hospital. The first was part of an EPSRC-funded research project called MATCH, and the second was an MSc student project ("pseudo-consultancy"):  

- a model developed by Sally McClean from the University of Ulster to evaluate thrombolysis for acute stroke patients. The model was developed in close collaboration with Dr Ken Fullerton, stroke consultant at Belfast City Hospital. It benefits from the genuinely integrated health and social system in Northern Ireland. Statistical models were developed using data from BCH to cluster patients into groups and then model the times spent in different stages of the pathway. These were then coded in two models: a user-friendly and visual Simul8 model, and an analytic model in Matlab which was used to validate the Simul8 model. The model was used by BCH to redesign the service and see the impact of increased thrombolysis rates. It is a good example of the use of a more rigorous mathematical model to validate a more "acceptable" simulation which can be understood (more or less) by clinical stakeholders;

- a (fairly simple) Simul8 model developed by two MSc students from the University of Southampton for the Paediatric ICU of the Children’s Hospital Wisconsin. The model was used to inform strategic bed capacity planning in a new building. Despite excellent client engagement and a powerful project sponsor, data collection was still an issue. The model recommendations were implemented. However the main benefit was probably not the model itself, but its role in the process of getting various senior clinicians to talk to each other in what was essentially a power politics situation. CHW was so impressed that it decided to employ a full-time modeller, and has also taken placement students every summer since then. This study illustrates what can be achieved in three months, for a cost of about $5,000, by a couple of smart students and a small amount of academic supervisory input.

Three more examples arose from a very interesting initiative led by Professor Paul Harper at Cardiff University, whereby a group of five modellers hold joint appointments at the University of Cardiff and at Aneurin Bevan University Health Board (ABUHB). This initiative is similar in some ways to the Whole System Partnership’s Workforce Collaborative example, in that it shows the practical benefits of a strategic and ongoing relationship between academic modellers and NHS partners. There are many examples of implemented work from this group, and they made three contributions to the Festival of Evidence. Rather than describe one specific model, the first piece of submitted evidence described the process of integrating modellers into the multidisciplinary Patient Flow Team at ABUHB, which is tackling the challenges of increased demand for emergency and unscheduled care across the entire Welsh healthcare system. A whole series of different modelling approaches have been used for different
aspects of this programme. The programme has only been running for one year but some of the model recommendations (or decisions supported by modelling) have been implemented and are already showing quantifiable benefits. The submission contains the following quote from the Deputy COO of the Royal Gwent Hospital: “The outputs from the (simulation) project exceeded expectations, in that it provided a level of understanding which went on beyond what was anticipated...This has been an excellent way of making operational teams aware of the potential for using new improvement methods such as using simulation as a tool to design future services.”

The two case studies submitted by the ABUHB group were:

- a Simul8 model developed in 2013 for the Fracture Clinic at the Royal Gwent Hospital. It was a classical patient flow model, the aim of which was to suggest improved configurations of the clinic space. Originally the clinical directorate wanted to increase the number of rooms used, but the model showed that the number of treatment rooms made very little difference to flow in that clinic and moreover might have undesirable knock-on effects on other parallel clinics. As a result the directorate decided not to make the proposed changes, but to implement changes in other areas highlighted by the model as bottlenecks. Hospital staff were involved throughout the modelling process. This case study shows clear evidence of the benefit of using a model (‘try before you buy’), although it can sometimes be hard to obtain concrete evidence if an organisation is unwilling to admit to having intended to do something which was later shown to be undesirable;

- a caseload monitoring tool for community mental health teams (CMHts). The paper-based tool in use at the time (June 2013) was not fit for purpose and was adding to the pressure on an already stressed workforce. A user-friendly Excel-based tool was developed in close collaboration with future users. It has now been rolled out across other areas and specialties in ABUHB including occupational therapists, and there has been interest expressed by other Mental Health Boards. This tool has greatly enhanced the ability of managers to monitor and balance workload across teams. It has the benefit of being in Excel, which all staff have access to and can use (to some extent). This case study demonstrates that models don’t have to involve sophisticated mathematical modelling or simulation in order to be useful.

The remaining 16 case studies were all either commercial (10) or from NHS/Local Authority groups (6). Given the challenges of searching the ‘grey’ literature, it was very pleasing to note that, for once, academic contributions were in the minority.

A few of the commercial examples simply described software tools or systems and had a ‘marketing’ flavour; they did not include any evidence of benefit but merely described the theoretical capability of the system. However, two software providers did report who was using their system and what benefits had been achieved:

- Clarity Informatics has developed a web-based Quality Improvement Service (QIS) toolkit, which analyses in-hospital mortality rates for 11 key conditions using pseudonymised SUS data. In developing the QIS tool, initial stakeholder engagement involved discussions with all staff involved in the quality improvement programme and then a set of clinical quality measures was developed through a modified Delphi process. Its website says “To date QIS has processed over 35 million patient records across more than 250 NHS Trust hospitals, mental health trusts and independent healthcare providers and is most extensively used throughout the North West, North East, East of England and South East coast regions of England. The findings from studies using the tool have been published in the NEJM and the journal Health Economics, both of which report statistical evidence of reduced mortality (and in the latter case, reduced LoS);

- GP Access is a commercial IT system which enables GP practices to monitor and manage demand for patient appointments and also manage other clinical data. It interfaces with all the main clinical information systems currently in use. The submitted evidence did not contain a description of how it works (presumably this is a commercial secret) or exactly what it does, but there are testimonials from happy patients and satisfied customers on the website.
The remaining commercial examples were submitted by modelling consultancies:

- Claire Cordeaux from Simul8 Corporation described a study commissioned by NHS England to model the implications of the Year of Care programme for patients with multiple long-term conditions. The Year of Care programme works across health and social care and considers the holistic needs of each person over a fixed period, rather than the episodic needs for each individual condition. The aim is to develop a total annual tariff for care provision for a given population classified by need, which can be used to commission services in a more integrated way. Simul8 modelled different risk groups (using commercial risk stratification tools to classify patients according to the number of LTC they have) and looked at clinical outcomes, costs and service utilisation. The model was applied for eight different ‘pathfinder’ sites and used data relevant for each site. The model provided a basis for discussion between commissioners and providers in these sites, and is about to be rolled out nationally. This case study provides clear evidence of influence on national policy;

- Justin Lyon from Simudyne Ltd developed a hybrid system dynamics and agent-based model for emergency evacuation of a hospital in Missouri in the event of major emergency or disaster scenarios, in collaboration with the security company Takouba. For obvious practical reasons, full-scale hospital evacuation scenarios cannot be rehearsed for real and so staff can only prepare for such events through computer simulation exercises. Participants/users learn by gaming with the model and seeing the effects of decisions they take under stressful conditions. The model has graphical features which add to the reality of the experience for participants. This is a classical use of simulation and one where there is a good evidence base for the success of such approaches;

- Peter Lacey from Whole Systems Partnership described a system dynamics model, developed in collaboration with the Cambridge Institute of Public Health, to explore the policy implications of new research which suggests there may be a cohort effect in dementia – in other words, later-born populations may have lower incidence than at present. The results were published in The Lancet, and enable stakeholders to explore the impact of changing incidence on service provision and policy-making. If true, this could have a profound effect on projected demand for dementia services. The impact of this model to date has really been to inform debate rather than change policy directly, but the work is clearly already influential;

- Peter also submitted a case study concerning one of the system dynamics models developed for the Workforce Modelling Collaborative (see section 4.6). This particular model relates to GP training and aims to incorporate both supply and demand aspects (population health need). The model has been applied at LETB level and the engagement process has involved local senior GP representatives in each location. The national-level models developed by the Centre for Workforce Intelligence do not take local variation, migration and cross-boundary flows into consideration in such a detailed way as these local-based models. The tool will be used to support discussions at the forthcoming National Committee of GP Education Directors. This study exemplifies excellent collaboration between modellers and NHS practitioners, and also shows the benefit of the NHS working with modellers who have a deep understanding of the domain;

- Douglas McKelvie from Symmetric SD, a modeller who used to be a social worker, described a system dynamics model he developed for ‘Winter Pressures’, which can be found in the NIHR Methods Review (http://www.sscr.nihr.ac.uk/PDF/MR14.pdf) (pages 29 – 32, section on variation). The model is essentially a game which enables planners to explore different strategies for coping with demand.

Finally, six examples were submitted by NHS organisations or Local Authorities. While a couple of these reported joint work with a commercial partner, others described work undertaken solely by themselves:

- Paul Schmidt from Portsmouth Hospitals Trust (PHT) is a consultant in acute medicine (i.e. not a modeller by trade). His case study described data analysis and modelling to develop an operational strategy for emergency and unscheduled care in the Emergency Department (ED) and Acute Medical Unit, which together form the Emergency Clinical Service
Centre (CSC) at PHT. Following an engagement and problem structuring process with CSC staff, a database was assembled of access, process and outcome data for each of the 115,000 patients in f/y 2011-12. This was used to segment and classify patient groups and identify flows and patterns. A family of increasingly complex Simul8 models was then developed to explore proposed improvements. Many of these require changes in staff rotas, work patterns and bed capacity, and the qualitative findings provided insight into how these might be achieved. The model has “substantially affected the planning for the Emergency CSC reconfiguration” but it is not yet clear whether the benefits predicted by the model will be realised in practice. One of the compelling things about this case study is that the model was developed by a doctor and not by professional modellers;

• James Friend, Director of Transformation at an anonymous acute hospital trust, presented a confidential case study. This study looks at the sources of variation in patient flow in emergency care, and the effects of this variation. The ED at this hospital is co-located with an Urgent Care Centre, and sees an average of 155 patients a day, of whom about 40 are admitted. The model is essentially an Excel-based analysis and presentation of routinely collected activity data. The model produces easily understandable graphical representations of the propensity for patients to be admitted at different times of day and days of week. The outcomes formed the basis for discussions with clinical staff in the ED, and highlight areas of concern. The model is maintained each month to identify trends and inform review process for individual clinical pathways.

• University Hospitals Birmingham Foundation Trust has developed a clinical decision support system which incorporates real-time e-prescribing, lab requests and reporting, imaging requests and clinical observations. It includes a kind of management dashboard, and has been used to reduce MRSA rates at University Hospitals Birmingham among many other things;

• Chad Oatley from the Isle of Wight Local Authority submitted a case study describing efforts to evaluate a programme on the Isle of Wight called ‘My Life: A Full Life’, which provides truly integrated patient-centred care to patients with disability. This is a very ambitious research project which aims to answer some challenging questions about evidence: how do you demonstrate the benefit (or even effect) of such a complex intervention, what constitutes evidence of improvement over and above what might have happened anyway, how can you identify causal links, etc. While there was no obvious modelling or simulation in this example, it raises some profound issues as we try to pin down what we mean by evidence. If we could develop a methodology to evaluate such programmes, we might be further down the road to establishing the evidence base for modelling and simulation;

• Del Herridge from Kent & Medway Public Health Observatory and Kent County Council described an integrated care pathway model for older mental health patients, the purpose of which is to assist in bed capacity planning (the overall strategy being to reduce acute inpatient bed numbers) and inform commissioning by assessing the need for other services such as a high dependency unit or crisis resolution service. A system dynamics model for patient pathways was developed using pseudonymised SQU data obtained from the local Mental Health Trust and Informatics services. This was undertaken in partnership with the Whole Systems Partnership, which provided training in the use of the modelling software iThink, but Del and his colleagues have done most of the actual modelling themselves. This seems to be rather similar to the Scenario Generator business model, where the intention is to ensure that the skills to do the modelling are embedded in CCG/LA staff, and although this is more sustainable than employing consultants clearly the risks identified in the Scenario Generator study remain (vulnerability to staff turnover, etc). It links to the Xavier case study on community bed modelling (below) and is still work in progress;

• Su Xavier from Dartford, Gravesham and Swanley Clinical Commissioning Group described a case study looking at the future demand for community beds. Otherwise, the underlying model seems to be the same as the model presented in e) above and it is underpinned by the same data.
4.8. Reflections and analysis of this evidence

We conclude with an attempt to categorise some of this evidence in terms of our empiricist-historicist-rationalist triangle of evidence, as presented in section 2.8. We recall that empiricism considers empirically-based data to be the "highest form" of evidence, while rationalism emphasises the importance of theoretical understanding of what is going on (causality), and historicism underlines the importance of the social context of knowledge (appropriateness and feasibility). Randomised control trials therefore lie at the empirical apex, stories and anecdotes lie at the historicist apex, and quantitative modelling and simulation (arguably) lie at the rationalist apex.

The decision about where to place each case study within the triangle is inevitably subjective, based on our evaluation of the extent to which a) a theoretical model played a key role in the example (Rationalist); b) the example was based on observed data (Empiricist); and c) the example took account of the context and human behavioural aspects of the problem. The point we wish to convey is that, ideally, evidence should lie at the centre of this triangle: the implication is that the conclusions of many of these studies would be "centred" (i.e. strengthened) by evidence of complementary types to that already identified.

Figure 4.1. The E-H-R triangle revisited
<table>
<thead>
<tr>
<th>Affiliation</th>
<th>Topic &amp; method</th>
<th>Position on the E-H-R triangle</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brunel University</td>
<td>Patient centred care IT system for cystic fibrosis</td>
<td>Currently firmly rationalist</td>
</tr>
<tr>
<td>University of Southampton &amp; Children’s Hospital Wisconsin</td>
<td>DES model for PICU capacity planning</td>
<td>Close to the centre</td>
</tr>
<tr>
<td>Simul8 Corp</td>
<td>DES model of Year of Care</td>
<td>Empiricist/rationalist</td>
</tr>
<tr>
<td>Cardiff University &amp; ABUHB</td>
<td>DES model of fracture clinic</td>
<td>Close to the centre</td>
</tr>
<tr>
<td>Anonymous Hospital Trust</td>
<td>Variation in ED flow using Excel</td>
<td>Empiricist/rationalist</td>
</tr>
<tr>
<td>Kent &amp; Medway PHO</td>
<td>Mental health bed SD model</td>
<td>Rationalist</td>
</tr>
<tr>
<td>Birmingham University Hospital</td>
<td>Decision support system and management dashboard</td>
<td>Empiricist</td>
</tr>
<tr>
<td>Cardiff University &amp; ABUHB</td>
<td>Caseload management Excel tool</td>
<td>Close to the centre</td>
</tr>
<tr>
<td>Whole Systems Partnership (WSP) Cambridge</td>
<td>SD model for dementia service planning</td>
<td>Rationalist</td>
</tr>
<tr>
<td>WSP workforce</td>
<td>SD model for GP workforce planning</td>
<td>Rationalist</td>
</tr>
<tr>
<td>GPAccess</td>
<td>Online tool for GP data collection</td>
<td>Empiricist</td>
</tr>
<tr>
<td>Simudyne</td>
<td>Hospital evacuation model (US)</td>
<td>Close to the centre</td>
</tr>
<tr>
<td>University of Ulster</td>
<td>DES model for acute stroke</td>
<td>Close to the centre</td>
</tr>
<tr>
<td>Symmetric SD</td>
<td>NIHR methods review and Winter Pressures SD model</td>
<td>Empiricist/rationalist</td>
</tr>
<tr>
<td>University of Manchester</td>
<td>DES model of an ED</td>
<td>Empiricist apex</td>
</tr>
<tr>
<td>Clarity Informatics</td>
<td>Tool for analysing mortality data</td>
<td>Empiricist apex</td>
</tr>
<tr>
<td>IoW Local Authority</td>
<td>Mixed methods story-telling for integrated care</td>
<td>Historicist apex</td>
</tr>
<tr>
<td>Portsmouth Hospitals Trust</td>
<td>Strategy development (Excel &amp; Simul8 models)</td>
<td>Close to the centre</td>
</tr>
<tr>
<td>Dartford, Gravesham and Swanley CCG</td>
<td>In collaboration with WSP: SD model for mental health workforce</td>
<td>Rationalist</td>
</tr>
</tbody>
</table>

Table 4.1. Suggested position of the submitted evidence on the Empiricist-Historicist-Rationalist triangle
5. OVERALL CONCLUSIONS AND RECOMMENDATIONS

5.1. Encouraging signs

On Thursday 23 October 2014, Day Four of the Festival of Evidence, Simon Stevens, Chief Executive of NHS England, published his NHS Five Year Forward View (available at http://www.england.nhs.uk/ourwork/futurenhs/). Page 34 of this document contains the following statement: “Working with NIHR and the Department of Health we will expand NHS operational research, RCT capability and other methods to promote more rigorous ways of answering high impact questions in health services redesign.” This explicit mention of Operational Research made hugely encouraging reading for participants at the Festival.

There are other signs of an increasing recognition that modelling and simulation are useful tools. Modelling and simulation groups exist within several NIHR Collaborations for Learning in Applied Health Research & Care (CLAHRCs). The PenCHORD group (http://clahr-peninsula.nihr.ac.uk/penchord.php), the Peninsula Collaboration for Health Operational Research & Development, was one of the first of such groups, and is part of the NIHR CLAHRC for the South West Peninsula. PenCHORD now employs six modellers and a research manager. PenCLAHRC was one of the five original pilot CLAHRCs (2008-13) and has recently received another five years’ funding under the new tranche of NIHR funding which now supports 13 CLAHRCs across England. The PenCHORD group, under the leadership of Professor Ken Stein and Dr Martin Pitt, has undertaken a large number of highly collaborative modelling projects with local providers and commissioners across Somerset, Devon and Cornwall. Another CLAHRC funded in the first tranche which directly employed operational research modellers was Cambridgeshire and Peterborough CLAHRC, which has now become the NIHR CLAHRC East of England (http://www.clahrc-eoe.nihr.ac.uk/). Two new CLAHRCs, funded from 2014-18, also employ dedicated staff to undertake modelling and simulation with their NHS partners. The NIHR CLAHRC Wessex has a Methodological Hub based jointly at University Hospital Southampton and the University of Southampton, consisting of two OR modellers, a statistician, a health economist and a database expert. The NIHR CLAHRC North Thames (http://www.clahrc-norththames.nihr.ac.uk/) also has a methodological theme led by Professor Martin Utley, director of the Clinical Operational Research Unit (CORU) at University College London. CORU also undertakes a wide range of modelling work on behalf of the Department of Health.

A further source of encouragement is the current and potential role of Academic Health Science Networks (AHSNs). The Kent, Surrey and Sussex AHSN has been a long-standing partner in the Cumberland Initiative and has supported it financially by sponsoring an overnight meeting. The Wessex AHSN has funded a PhD project on dementia modelling, and a research project on simulation modelling for colorectal cancer screening in Wessex.

Finally, other innovative collaborations are beginning to show great benefits. A Welsh Health Board (ABUHB) now employs five modellers under a joint initiative with Cardiff University, and several commercial organisations are supporting the use of modelling software by NHS users through membership groups such as the Whole Systems Partnership Workforce Collaborative and Simu8’s Scenario Generator User Groups.
5.2. Overall conclusions

- Modelling and simulation are extremely useful techniques to improve decision-making, but (unlike in other domains) have not yet achieved their potential in healthcare.
- There is an abundance of evidence, both national and international and in many forms, for the benefit of simulation and modelling in healthcare, but it is anecdotal, fragmented and diverse.
- Globally, there is little evidence of widespread successful ‘rollout’ of modelling interventions beyond the original client organisation.
- We believe that the best evidence is that which is composed of a mix of empiricist, rationalist and historicist sources. To assemble such evidence takes time, and we would generally expect the journey of evidential support to start from an empiricist, rationalist or historicist extreme, and move towards a more balanced profile.
- Evidence means different things to different people; and different people are convinced by different types of evidence.
- There are encouraging signs in the UK that modelling and simulation are becoming more embedded within NHS organisations.

5.3. Recommendations: scaling up the use of modelling and simulation

- For modelling and simulation to become part of the NHS management ‘toolkit’ it is essential that these skills are embedded within NHS organisations, and are not seen either as complicated techniques requiring a PhD, or as something that can only be done by specialist management consultants.
- It is also essential that senior NHS management has an appreciation of the value of modelling and simulation, and there is evidence to suggest that a statutory requirement to demonstrate use of modern management approaches is helpful in bringing this about.
- The cultural divide between the clinical and management points of view is a factor inhibiting the appropriate use of evidence in healthcare. Efforts should be made not to eliminate the divide, but rather to create a pluralist ‘meta-culture’ in which each side appreciates the values and strengths of the other, and acknowledges its own limitations.
- Researchers (in the areas of both medicine and management) should take care to present their findings in ways which facilitate their adoption where appropriate by potential users.
- Evidence from modelling organisational innovation/interventions needs to be presented to clinicians and managers in a way that is different from how evidence for new treatments (RCTs or meta-analyses) is presented. This requires a culture change in the NHS.
- The modelling community needs stronger links with the implementation science community; why do some innovations spread and become common practice, while others never take off?

5.4. A research agenda

- We endorse the view which suggests that evidence-based practice should allow for more diverse evidence generation, and recognises the strengths of such generation methods.
- There is a need for new ways of categorising and evaluating different types of evidence, especially for organisational interventions.
- There is a need for more academic research into the ‘grey’ literature, and new, rigorous tools for analysing and interpreting its contents.
- We need to think about the nature of evidence for organisational innovation/interventions, and develop an equivalent to evidence-based medicine that has the same rigour and universal acceptability as the RCT.
6. REFERENCES


von Bertalanffy L., 1969. General System Theory. (George Braziller)


7. APPENDICES

7.1. List of keynote speakers

**Mike Carter**
(University of Toronto, Canada):
Major challenges in hospital modelling, and tips for meeting them

**Len Goldschmidt**
(Department of Veterans Affairs, USA):
Taking a systems approach to improving patient pathway: the VA Care-Coordination Tele-health Model

**Carol Jagger**
(University of Newcastle, UK):
Whole systems thinking for complexity and frailty: challenges and guidance

**Daryl Kor**
(Mayo Clinic, USA):
Taking a systems approach to improving healthcare delivery: challenges and ways to meet them

**Mark Mackay**
(Flinders University, Australia):
Setting metrics and making service provision more effective and efficient: challenges and guidance

**Claire Cordeaux**
(Simu8 Corporation, UK):
Cross-boundary, cross-sector – using simulation to understand the impact of integration

**Peter Lacey**
(Whole Systems Partnership, UK):
Moderating discussion panel: Understanding population health across the spectrum of needs and over time
7.2. Pro-forma for collecting evidence from participants

FESTIVAL OF EVIDENCE | 20-24 OCTOBER 2014

PRO-FORMA FOR PROVIDING EVIDENCE

- Personal details
- Type of contribution
- Evidence details
  - Title
  - Where it is from
  - Who owns it
  - When was it collected (month, year, or period)
  - Media in which it is available (electronic, paper, etc.)
  - Confidentiality of information
  - Relevant category/categories based on outcomes (strategic, operational, etc)

- Description
- Methods and tools
- Process
- Outcomes
- The implications
## 7.3. Submitted evidence

<table>
<thead>
<tr>
<th>Name</th>
<th>Affiliation</th>
<th>Main role</th>
<th>Model used?</th>
<th>Topic &amp; method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pam Abbott</td>
<td>Brunel University</td>
<td>Academic</td>
<td>No (a research proposal)</td>
<td>Patient centred care IT system for cystic fibrosis</td>
</tr>
<tr>
<td>Steve Box</td>
<td>X-lab</td>
<td>Commercial</td>
<td>Hard to tell</td>
<td>IS tool for sepsis</td>
</tr>
<tr>
<td>Sally Brailsford</td>
<td>University of Southampton</td>
<td>Academic</td>
<td>Yes</td>
<td>DES model for PICU capacity planning</td>
</tr>
<tr>
<td>Claire Cordeaux</td>
<td>Simul8 Corp</td>
<td>Commercial</td>
<td>Yes</td>
<td>DES model of Year of Care</td>
</tr>
<tr>
<td>Tracey England</td>
<td>Cardiff University and ABUHB*</td>
<td>Joint academic &amp; NHS</td>
<td>Yes</td>
<td>DES model of fracture clinic</td>
</tr>
<tr>
<td>Brian Fisher</td>
<td>PAERS Ltd</td>
<td>Commercial</td>
<td>Unclear</td>
<td>Personal story</td>
</tr>
<tr>
<td>James Friend</td>
<td>Anonymous Hospital Trust</td>
<td>NHS</td>
<td>Yes</td>
<td>Variation in ED flow using Excel</td>
</tr>
<tr>
<td>Del Herridge</td>
<td>Kent &amp; Medway PHO</td>
<td>NHS</td>
<td>Work in progress</td>
<td>Mental health bed model (System dynamics)</td>
</tr>
<tr>
<td>Tim Jones</td>
<td>Birmingham University Hospital</td>
<td>NHS</td>
<td>Apparently</td>
<td>Decision support system and management dashboard</td>
</tr>
<tr>
<td>Izabela Komenda</td>
<td>Cardiff University and ABUHB*</td>
<td>Joint academic &amp; NHS</td>
<td>Work in progress</td>
<td>Caseload management tool</td>
</tr>
<tr>
<td>Peter Lacey</td>
<td>Whole Systems Partnership (WSP) Cambridge</td>
<td>Commercial</td>
<td>Yes (to aid discussion)</td>
<td>SD model for dementia service planning</td>
</tr>
<tr>
<td>Peter Lacey</td>
<td>WSP workforce</td>
<td>Commercial</td>
<td>Yes (to aid discussion)</td>
<td>SD model for GP workforce planning</td>
</tr>
<tr>
<td>Harry Longman</td>
<td>GPAccess</td>
<td>Commercial</td>
<td>Yes</td>
<td>Online tool for GP data collection</td>
</tr>
<tr>
<td>Justin Lyon</td>
<td>Simudyne</td>
<td>Commercial</td>
<td>Yes</td>
<td>Hospital evacuation model (US)</td>
</tr>
<tr>
<td>Sally McClean</td>
<td>University of Ulster</td>
<td>Academic</td>
<td>Yes</td>
<td>DES model for acute stroke</td>
</tr>
<tr>
<td>Douglas McKelvie</td>
<td>Symmetric SD</td>
<td>Commercial</td>
<td>N/A</td>
<td>NIHR methods review findings</td>
</tr>
<tr>
<td>Jim Methven</td>
<td>University of Manchester</td>
<td>Academic</td>
<td>Unclear</td>
<td>DES model of an ED</td>
</tr>
<tr>
<td>Gerry Morrow</td>
<td>Clarity Informatics</td>
<td>Commercial</td>
<td>Yes</td>
<td>Tool for analysing mortality data</td>
</tr>
</tbody>
</table>
7.3. Submitted evidence (continued)

<table>
<thead>
<tr>
<th>Name</th>
<th>Affiliation</th>
<th>Main role</th>
<th>Model used?</th>
<th>Topic &amp; method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chad Oatley</td>
<td>IoW Local Authority</td>
<td>NHS / LA (Public Health)</td>
<td>Unclear: ongoing work</td>
<td>Mixed methods story-telling for integrated care</td>
</tr>
<tr>
<td>Paul Schmidt</td>
<td>Portsmouth Hospitals Trust</td>
<td>NHS</td>
<td>Yes</td>
<td>Strategy development (Excel &amp; DES models)</td>
</tr>
<tr>
<td>David Southern</td>
<td>Pathway Communications</td>
<td>Commercial</td>
<td>Unclear</td>
<td>DES tool for Parkinson’s</td>
</tr>
<tr>
<td>Julie Vile</td>
<td>Cardiff University and ABUHB*</td>
<td>Joint academic &amp; NHS</td>
<td>Largely</td>
<td></td>
</tr>
<tr>
<td>Su Xavier</td>
<td>Dartford, Gravesham and Swanley CCG</td>
<td>NHS</td>
<td>Ongoing but has already influenced strategy</td>
<td>In collaboration with WSP: SD model for mental health workforce</td>
</tr>
</tbody>
</table>
7.4. Flipchart and Post-it feedback

Over the course of the whole week, participants were asked to add their responses, using Post-it notes, to the following questions, and add them to a set of flipchart sheets affixed to the walls of the main room. People were asked to state their background (although not all did) as we were interested in the opinions of different stakeholder groups. The questions were:

“What gives people confidence that models are right/useful? What would convince you?”

1. Systematic review (clinician)
2. RCT (for interventions, not processes)
3. Cochrane collaboration report
4. Peer-reviewed publications (academic)
5. Recognised publisher (academic)
6. Repeatability of successful model predictions
7. Transparency; examples demonstrating how the model is used and what the outcomes are (academic: mathematics/computer science)
8. Selling – delivering – being paid (business consultant)
9. Documented case study (NHS commissioner)
10. Statistical analysis to confirm similar performance measures and similar numbers of patients entering different points of the system in the model and the real world system (academic: mathematics)
11. Clinical outcome accuracy and final diagnosis (clinician)
12. It depends! It works, it was useful, it generated discussion and/or learning. It was validated, publications (NHS manager)
13. What was the process, who was involved, what validation was done (IT management consultant)
14. For proof, a test of a previous prediction against actual performance. For trust: comparison to a reasonable simplified calculation; credible results in a sensitivity analysis (software engineer)
15. Patient feedback
16. Business planning/general management: understand assumptions and know how these were derived; extent to which assumptions are estimates of estimates; confidence in model developers (not just technicians)
17. Valid hypotheses
18. Is framing of the problem valid?
19. Are you able to model the existing system, and how well?
20. How many inputs are real, valid samples, and how many are assumed or ?valid
21. Understand the language: no additional workload; is it based on; where did the data come from; who was involved in the model and do I trust them; will it improve my practice; does it generate more paperwork; where else is it used; did it work; how long is it in use; do the model developers understand my work; can I be involved; who can help me and what is their input; does it cost in terms of time, money, effort (nurse)
7.5. Themes arising during the week

- Stories v statistics
- Different stakeholders need different evidence
- Importance of networking and sharing experience
- What counts is what works – case studies
- Taxonomy of types of evidence
- Model transparency: data validity, to see utility; to trust
- Appropriate validation: testing assumptions, understanding approximation
- Transformational: moving from modelling vs anecdotes and heuristics to modelling and anecdotes and heuristics (verifying them)
- Training/exposure at med school; quality improvement ‘curriculum’, skill development and support
- Critically appraising models; levels of robustness
- How do we achieve buy-in?
- How do we provide evidence of the benefits of simulation and modelling?
- Demonstrating benefits: use simulation to solve a small/simple problem well, with a champion in an organisation (probably clinical)
- Working with stakeholders to successfully embed simulation in an organisation
- Stakeholder engagement: best practice in modelling
  - involve public, patients
  - set patient outcomes against cost/efficiency
- Sharing the models: sharing the evidence
### Attendees

<table>
<thead>
<tr>
<th>Name</th>
<th>Affiliation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pamela Abbott</td>
<td>Brunel University</td>
</tr>
<tr>
<td>Mark Attah</td>
<td>NHS Cambridgeshire &amp; Peterborough CCG</td>
</tr>
<tr>
<td>Adrian Baker</td>
<td>Royal College of Nursing</td>
</tr>
<tr>
<td>Paul Benson</td>
<td>Southampton City Clinical Commissioning Group</td>
</tr>
<tr>
<td>Sally Brailsford</td>
<td>University of Southampton</td>
</tr>
<tr>
<td>Harald Braun</td>
<td>i5 Health</td>
</tr>
<tr>
<td>Mike Carter</td>
<td>University of Toronto</td>
</tr>
<tr>
<td>Salma Chahed</td>
<td>University of Westminster</td>
</tr>
<tr>
<td>Thierry Chausalet</td>
<td>University of Westminster</td>
</tr>
<tr>
<td>Carol Cochrane</td>
<td>Whole System Partnership</td>
</tr>
<tr>
<td>Claire Cordeaux</td>
<td>Simu8 Corp</td>
</tr>
<tr>
<td>Steve Counsell</td>
<td>Brunel University</td>
</tr>
<tr>
<td>John Crawford</td>
<td>IBM</td>
</tr>
<tr>
<td>Bridget Dibb</td>
<td>Brunel University</td>
</tr>
<tr>
<td>Rob Duncan</td>
<td>Welsh Clinical Leadership</td>
</tr>
<tr>
<td>Tracey England</td>
<td>Cardiff University and Aneurin Bevan University Health Board</td>
</tr>
<tr>
<td>Michaela Finegan</td>
<td>NHS Improving Quality</td>
</tr>
<tr>
<td>Brian Fisher</td>
<td>PAERS Ltd</td>
</tr>
<tr>
<td>Richard Fitton</td>
<td>Self</td>
</tr>
<tr>
<td>Andrew Fordyte</td>
<td>South Devon Healthcare NHS Foundation Trust</td>
</tr>
<tr>
<td>James Friend</td>
<td>Anonymous Hospital Trust</td>
</tr>
<tr>
<td>Abraham George</td>
<td>Public Health, Kent County Council</td>
</tr>
<tr>
<td>Leonard Goldschmidt</td>
<td>Department of Veterans Affairs, USA</td>
</tr>
<tr>
<td>Stuart Grossman</td>
<td>Clarity Informatics</td>
</tr>
<tr>
<td>Richard Guerrero</td>
<td>University of Southampton</td>
</tr>
<tr>
<td>Paul Harper</td>
<td>Cardiff University</td>
</tr>
<tr>
<td>Romana Hoossein</td>
<td>Brunel University</td>
</tr>
<tr>
<td>Carol Jagger</td>
<td>University of Newcastle</td>
</tr>
<tr>
<td>Jo James</td>
<td>NHS Improving Quality</td>
</tr>
<tr>
<td>Jonathan Klein</td>
<td>University of Southampton</td>
</tr>
<tr>
<td>Vince Knight</td>
<td>Cardiff University</td>
</tr>
<tr>
<td>Alex Komashe</td>
<td>University of Cambridge</td>
</tr>
<tr>
<td>Izabela Komenda</td>
<td>Cardiff University and Aneurin Bevan University Health Board</td>
</tr>
<tr>
<td>Daryl Kor</td>
<td>Mayo Clinic, USA</td>
</tr>
<tr>
<td>Jasna Kullis</td>
<td>Brunel University</td>
</tr>
<tr>
<td>Peter Lacey</td>
<td>Whole System Partnership</td>
</tr>
<tr>
<td>Kerry Little</td>
<td>Health Informatics Services, Dorset County Hospital NHS FT</td>
</tr>
<tr>
<td>Harry Longman</td>
<td>GPAccess</td>
</tr>
<tr>
<td>Linda Luxon</td>
<td>Royal College of Physicians</td>
</tr>
<tr>
<td>Justin Lyon</td>
<td>Simudyne Ltd</td>
</tr>
<tr>
<td>Mark Mackay</td>
<td>Flinders University, Adelaide</td>
</tr>
<tr>
<td>Hugo Mathias</td>
<td>KPMG</td>
</tr>
<tr>
<td>Elaine Maxwell</td>
<td>Royal College of Nursing</td>
</tr>
<tr>
<td>Sally McClean</td>
<td>University of Ulster</td>
</tr>
<tr>
<td>Fiona McDermott</td>
<td>KPMG</td>
</tr>
<tr>
<td>William McGeehin</td>
<td>Frazer-Nash Consultancy</td>
</tr>
<tr>
<td>Douglas McKelvie</td>
<td>Symmetric SD</td>
</tr>
<tr>
<td>Ann McMahon</td>
<td>Royal College of Nursing</td>
</tr>
<tr>
<td>Jim Methven</td>
<td>University of Manchester</td>
</tr>
<tr>
<td>Paul Mizen</td>
<td>Aneurin Bevan University Health Board</td>
</tr>
<tr>
<td>Tom Monks</td>
<td>University of Southampton</td>
</tr>
<tr>
<td>Name</td>
<td>Organization/Role</td>
</tr>
<tr>
<td>--------------------------</td>
<td>----------------------------------------------------</td>
</tr>
<tr>
<td>Felix Mukoro</td>
<td>NHS Improving Quality</td>
</tr>
<tr>
<td>Chad Oatley</td>
<td>IoW Local Authority</td>
</tr>
<tr>
<td>Sue Oliver</td>
<td>Royal College of Nursing</td>
</tr>
<tr>
<td>David Paynton</td>
<td>Southampton City Clinical Commissioning Group</td>
</tr>
<tr>
<td>David Peregrine-Jones</td>
<td>Torus Business Web</td>
</tr>
<tr>
<td>Caroline Powell</td>
<td>University of Southampton</td>
</tr>
<tr>
<td>Rhonda Reilly</td>
<td>Royal College of Nursing</td>
</tr>
<tr>
<td>Paul Schmidt</td>
<td>Portsmouth Hospitals Trust</td>
</tr>
<tr>
<td>James Sheffield</td>
<td>Data Analytics Lab</td>
</tr>
<tr>
<td>Steven Sherrat</td>
<td>Pollard Systems</td>
</tr>
<tr>
<td>Mike Sinclair</td>
<td>Health Informatics Services, Dorset County Hospital NHS FT</td>
</tr>
<tr>
<td>Rob Smith</td>
<td>NHS Improving Quality</td>
</tr>
<tr>
<td>David Southern</td>
<td>Pathway Communications</td>
</tr>
<tr>
<td>Lampros Stergioulas</td>
<td>University of Surrey</td>
</tr>
<tr>
<td>Joe Viana</td>
<td>University of Southampton</td>
</tr>
<tr>
<td>Julie Vile</td>
<td>Cardiff University and Aneurin Bevan University Health Board</td>
</tr>
<tr>
<td>Carrie Ward</td>
<td>Health Informatics Services, Dorset County Hospital NHS FT</td>
</tr>
<tr>
<td>Suzanne Wixey</td>
<td>IoW Healthcare &amp; Clinical Commissioning Group</td>
</tr>
<tr>
<td>Terry Young</td>
<td>Brunel University</td>
</tr>
</tbody>
</table>
To transform the quality and cost of NHS care delivery through simulation, modelling and systems thinking.