A&E Department and Hospital Bed Modelling
- Where are the bottlenecks?

Authors:
Dr. Jonny Pearson PhD  
NHS England
Dr. David Halsall PhD  
NHS England
Steven Paling  
NHS Improvement

www.england.nhs.uk/expo Join the conversation @ExpoNHS | #Expo18NHS
Slide to help structure draft – will turn into contents

Background
• Context
• System Modelling
• Simple Concept

Linear
• Example 1: Simple - Stock and Flow
• Example 2: Focussed – Regression
• Example 3: Focussed – Result and resilience of Impact on AE

Non-Linear Modelling
• Example 4: Simple – Backdoor
• Example 5: Simple – Front door
• Example 6: Focussed – GP Streaming
• Example 7: Focussed – Staff Utilisation
• Example 8: Complex – Full hospital model
• Example 9: Complex – “Big Shire”
• Example 10: Whole System model

Lessons Learnt

The way forwards

www.england.nhs.uk/expo Join the conversation @ExpoNHS | #Expo18NHS
The 4-hour A&E performance measure has been called a barometer for the overall running of a hospital. At a national level, the standard has not been met annually since 2013-14.

Many organisations have investigated the causes for crowding and missing the target. The most common issues have been recently summarised by the National Audit Office [1] and the Nuffield Trust [2]. Many of these directly relate to bed occupancy.

Is this issue related to strategy misdirection, operational inefficiency, or an inevitable crash?

Various attempts in the past have been made to define optimal occupancy rates. This figure is not just about allowing enough room for variations in supply/demand over time, but also takes into account that when the occupancy gets high the efficiency of the system drops dramatically.

This can be seen in motorway traffic jams.

Fredrick Winslow Taylor published “The Principles of Scientific Management” in 1909. He studied how work was performed, and looked at how this affected the worker productivity. He created Four Principles of Scientific Management:

1. Replace working by "rule of thumb," with the most efficient way to perform specific tasks.
2. Match workers to jobs based on capability and motivation, and train them to work efficiently.
4. Allocate managers their to planning and training, allowing the workers to perform efficiently.

The NHS doesn’t prescribe to the idea of “one right way”, but it is a useful starting point.

How can we apply this methodology to A&E departments and hospitals in general?

Create a virtual “time and motion” study by modelling each event that changes the state of the system using discrete event simulation.
Conceptually the patient flow through A&E is simple......

In-Flow

Bottlenecks

Out-Flow

Typically 1 – 7 days

4 hours
Example 1: Linear Simple

Model: Using hourly SUS data we can investigate variation in attendance, admission and discharge rates throughout the week.

Results: Simply highlighting these profiles gives a rich picture of the pressures on the system at different times of the day and week. Changes to the profiles, such as reducing electives on a Monday or discharging earlier in the day, can have a dramatic effect on the overall occupancy.
Example 1: Linear Simple

Stock and Flow

Model: Using hourly SUS data we can investigate variation in attendance, admission and discharge rates throughout the week.

Results: Simply highlighting these profiles gives a rich picture of the pressures on the system at different times of the day and week. Changes to the profiles, such as reducing electives on a Monday or discharging earlier in the day, can have a dramatic effect on the overall occupancy.

www.england.nhs.uk/expo Join the conversation @ExpoNHS | #Expo18NHS
Example 2: Linear Simple

Impact of additional demand on time in A&E and probability of breaching

An alternative way to use the hourly data is to investigate the relationship between different explanatory variables in order to see what is having the largest effect on our waiting times.

Model: Using a linear multiple-regression we can separate the effects of provider, hour of day, day of week, month of year, bank holidays, and investigate the additional demand measured by an additional arrival in the previous hour. This gives an indication of the impact of additional demand on both the waiting time and probability of breaching the 4-hour target.

Result: This modelling showed that an additional arrival in the previous hour caused just over 1 minute additional waiting time for non-admitted patients and 2.5 minutes for admitted patients. The chance of breaching increased by 0.4% and 1.4% for non-admitted and admitted patients, respectively.

www.england.nhs.uk/expo Join the conversation @ExpoNHS | #Expo18NHS
We have done detailed analysis of the factors that drive A&E performance

- We used econometric analysis (pooled OLS), which allows us to bring together all the factors and isolate the effect of each on A&E performance. We mean we can identify which factors are statistically significant and most influential on performance.
- Daily data from winters 2016/17-17/18, covering all 137 providers with a type 1 A&E department. This large dataset dilutes the effect of noisy data and makes the findings more robust.
- We have controlled for quality, size of A&E department and patient characteristics.

Our findings at a glance

The A&E department

Patient flow

Our model matches actual A&E performance well

www.england.nhs.uk/expo Join the conversation @ExpoNHS | #Expo18NHS

We’ll cover these in detail here
**Method**

We looked at how different levels of bed occupancy for general and acute beds affected A&E performance, holding all other factors constant.

**Findings**

Impact of bed occupancy on A&E performance, holding all other factors constant

---

Our analysis finds a significant A&E performance tipping point at 92% bed occupancy using Sitrep data, which updates the historic view of 85% based on KH03 data.

The 85% was based on simulation analysis of 2 hospitals in 1999. Since then:

- fundamental changes in the health system and UEC pathways
- newly available daily data on bed occupancy through Sitrep

Our research updates this using an operationally available and useful measure of occupancy.

---

www.england.nhs.uk/expo Join the conversation @ExpoNHS | #Expo18NHS
Who’s occupying the beds

Method

We looked at the proportion of stranded patients who have been in hospital for 21 days or more, and whether this had an effect over and above bed occupancy.

Findings

This ratio is closely correlated to A&E performance (see chart) and is a statistically significant driver in our model, over and above bed occupancy. As this effect is in addition to bed occupancy, it suggests long-stay patients affect the flexibility of beds.
Example 4: Non-Linear Simple

Impact of social care resource on DTOCs

Model: This model considers an average size hospital with patients being admitted non-uniformly each of which has a randomly assigned realistic length of stay. The patients enter the hospital and take up a bed until their assigned length of stay is complete. At this point 8% of the patients require a transfer to one of two social care settings. If these settings are full then the patient waits in the hospital; accumulating DTOC days. The model then allow the user to adjust the resources available in the social care settings and investigate the impact on DTOC if further social care settings were available.

Results: The model was able to demonstrate that a hospital like setting will always need a degree of “headroom” in order to maintain some flexibility. Without this, the non-linear nature of occupancy creates severe backlogs.

When approximate costings were applied this showed the cost/benefit of adding more of each type of social care setting versus releasing DTOC beds.

www.england.nhs.uk/expo Join the conversation @ExpoNHS | #Expo18NHS
Example 5: Non-Linear Simple

Scenario Testing

Model: Working with Lancaster University Masters Students from STOR-I, a simple A&E model was developed to investigate the impact that different scenarios may have.

The key was to highlight possible interventions which would have the greatest impact and where possible quick wins might be.

Results: The impact of multiple scenarios on trolley waits were investigated and compared.

Scenarios included changing the admission and discharge profiles, increasing available beds and reducing bed turnaround time.

Example: Reducing the bed turnaround time by 45% caused the average trolley waiting time to drop from 82 minutes to 16 minutes. Although, this may not be a realistic result due to the original assumptions applied, this does highlights the importance of getting a resource ready and back in the system as quickly as possible.
Example 5: Non-Linear Simple

Scenario Testing

**Model:** Working with Lancaster University Masters Students from STOR-I, a simple A&E model was developed to investigate the impact that different scenarios may have. The key was to highlight possible interventions which would have the greatest impact and where possible quick wins might be.

Results: The impact of multiple scenarios on trolley waits were investigated and compared. Scenarios included changing the admission and discharge profiles, increasing available beds and reducing bed turnaround time.

Example: Reducing the bed turnaround time by 45% caused the average trolley waiting time to drop from 82 minutes to 16 minutes. Although, this may not be a realistic result due to the original assumptions applied, this does highlights the importance of getting a resource ready and back in the system as quickly as possible.

Current ongoing work is further developing this modelling in order to make the scenarios easier to compare and to develop the base model further.

www.england.nhs.uk/expo Join the conversation @ExpoNHS | #Expo18NHS
Example 6: Non-Linear focussed

The cost of having a responsive system – staff need to be available as demand increases

We have shown that the average demand for A&E services and cyclic nature by day of the week and hour of the day is largely predicable. But there does remain random variations of who comes through the door, particularly in the case mix of the patients.

In general there are seven classes of patient specialities which will admit patients in a typical acute hospital. Ensuring prompt decisions to admit for all patients requires spare capacity across a range clinical specialities as well the appropriate bed capacity.

Model: The A&E model has been used to show how much staff utilisation “headroom” is required to provide a responsive service.

Results: An average staff utilisation of 70% will provide a 98% 4-hour performance. That is staff will be ready but not activity involved in patient care for 30% of their shift (point 1)

If there is a short term increase in demand of +15% performance will drop to 95% and staff utilisation will go up to 80% (point 2)

If there is a short term increase in demand of 40% performance drops to 75% seen in 4 hours and staff utilisation goes up to 90% (point 3)
Example 7: Non-Linear Focussed

Using GP Streaming to divert non-urgent demand

Less than 10% of patients arriving in a typical A&E department are initially judged as needing imminent or urgent assessment by a doctor. A significant minority of patients attending an A&E department are said to be using the service when they would be better treated in a primary care setting.

Co-location of nurse led walk-in clinics and more recently portable GP surgeries next to A&E departments have been promoted to help reduce demand for major A&E services.

**Model:** The A&E model was modified to permit the testing of active diversion of non-urgent patients away from the A&E department at initial assessment. This reduced the demand for the minor stream of patients. The impact on the 4 hour metric at three levels of demand was tested.

**Results:** This showed that active diversion has minimal impact at low average demand when the staff utilisation was sufficient to provide a responsive service. However it did make significant improvements short term demand increased and the system started to run hot.

<table>
<thead>
<tr>
<th>demand</th>
<th>4 hour performance</th>
<th>A&amp;E staff utilisation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Baseline</td>
<td>+20%</td>
</tr>
<tr>
<td>with non-urgent patients</td>
<td>98%</td>
<td>91%</td>
</tr>
<tr>
<td>without non-urgent patients</td>
<td>98%</td>
<td>93%</td>
</tr>
</tbody>
</table>

www.england.nhs.uk/expo Join the conversation @ExpoNHS | #Expo18NHS
Example 8: Complex

Full hospital model

The obvious progression of this modelling would be to bring together the A&E and admitted parts of the hospital into a single model.

Model: This worked asked what’s the simplest model we can build that will incorporate effects of both the A&E and admitted settings within a hospital. This is useful as it allows a direct comparison between the front and back doors as well as the process.

Results: The model is able to recreate realistic outputs for A&E waiting times, bed occupancy and flow rates through a generic hospital.

Currently, being QA’d and tested before scenarios applied.

Needs further development in the discharge interaction and the Bed approval stages.
Example 9: Complex – is a hospital a rational system?

Bigshire modelling the whole pathway from arrival to discharge

We have shown that a generic A&E simulation model can provide insights into the working of a complex and non-linear system. Is the model still valid and producing good results when further detail is added?

Model: The Bigshire model represents a simplified patient journey for non-elective patients through medical and surgical specialties. A number of simplifications are made. It permits investigations of how increasing resources can translate into improved performance. In particular it was used to show how redistributing staff resources into weekend working could balance the load across the week better.

Results: Hospitals performed better than the model in most cases because of the informal processes to allocate resources on a patient by patient basis - as described by Eric Wolstenholme (2005) in “Coping but not Coping in Health and Social Care”

[Diagram of a hospital pathway from arrival to discharge]

www.england.nhs.uk/expo Join the conversation @ExpoNHS | #Expo18NHS

---

1 For example no route through intensive care, no specific surgical interventions and a generic diagnostic pathway is used.
Example: system factors not in the data

**Method**
Measuring resilience helps us identify the differences in performance between providers, which are often driven by static, local factors that are difficult to quantify. We define resilience in two ways.

1. How much an A&E departments’ performance falls after a **surge in admissions**

2. How many days it takes a providers to **bounce back** from poor A&E performance

**Findings**
- The trusts that best dealt with **surges in admissions** experienced half the dip in A&E performance.
- This resilience factor isn’t related to the operational factors – more resilient providers do not seem to have lower bed occupancy or more senior workforce for example.
- This could therefore be picking up factors that we can’t measure, such as managerial capacity, culture or leadership.

www.england.nhs.uk/expo Join the conversation @ExpoNHS | #Expo18NHS
Example 10: Whole System

Whole system simulation of UEC

What about other healthcare settings and the impact of care being provided in an inappropriate setting? Can we show the whole healthcare system in a single model?

**Model:** A model was built in excel which included hourly arrival times across 32 points of delivery and 80 different patient types. Each region would of then needed to input figures for activity and resources for over 60,000 inputs.

The model was then a stock and flow model without feedback and so essentially linear. The aim was to show the fluctuations in all settings across a whole system.

**Result:** Model was unusable as the input requirements were unrealistic. Additional as no connections existed between buckets and pathways assumed to be linear, it’s difficult/impossible to predict how changes would have affected the system or to gain new insight into the workings of the system.

**Current:** Demand and Utilisation models continue to be developed.
Lessons Learnt

Relating to the system dynamics
• A&E performance is directly dependent on the number and flexibility of available admitted beds.
• Hourly flows are important.
• The impact of peaks in demands takes time to dissipate.
• Head room is essential for efficient running (in both beds and staff time). Need a flexible short term setting to maintain flexibility.
• A closed back door ripples back through the hospital to the A&E performance.
• The quickest win can be found in turning beds around faster and getting them back in the system.
• Diversion settings work when things get hot but little impact on average demand.
• There is a lot more to hospital system models than flows of numbers.

Relating to the modelling method
• Bigger isn’t always better
• A single focus on a question in hand is easier than a general look but can miss the additional dependencies outside of the focussed issue.
Lessons Learnt

Why hasn’t simulation had a bigger impact on healthcare?

From the early days of computer simulation we learnt it is not simply a technical process to improve and optimise a process. There are very successful examples of were simulation of steel works, ports and car factories have had an important role to play in productivity improvement.

But even in these examples simulation is as much a method for learning for those involved in the processes as a back room optimisation tool.

We have shown that simulation does have a role to play in healthcare. But its impact is possibly less than would have been anticipated.

The biggest successes have been in simpler models which have gained wide acceptance. Visual interactive modelling has always been promoted as a way to engage users – but this strength may have been underplayed in health care applications.

(i) It can provide the vocabulary for those who know what the situation is but don’t know how to describe or quantify the effect.
(ii) But it can be also used to challenge the views of those who are responsible in delivering improvements.
The way forwards

Are we attempting to get blood out of a stone!

The 20th century was the age of the rational system designers

The work of management scientists like Fredrick Winslow Taylor feed mass production experts such as Henry Ford who stopped at nothing to control everything.

The tension between rational designers were, workers are just part of the system and the socio-technical view where there should be a constructive interaction between people and technology in workplaces is possibly as strong today as when the debate started nearly a century ago.

When you also bring into the mix that patients and the public are being encouraged to give voice to how they want the health service to provide the care they want we find the simulation has many audiences

This gives a pointer that in the world of big data we might find it easier to build more complex models but harder to get those model to change the way we provide services.

We may want to get more blood out of the stone. But have we asked the stone how it feels?

www.england.nhs.uk/expo Join the conversation @ExpoNHS | #Expo18NHS
The way forwards

Are new avenues available using big data flows?

If we are being hindered by the assumptions required for non-linear discrete event simulation then are there new and upcoming techniques that we could link to our modelling in order to allow us to look at the exceptional cases and the nuances of the non-linear system.

Rather than build a model to represent what the modeller thinks is reality, and then trying to populate it with data, why don’t we create a model based solely on available individual record level data. To do this each case is run through an algorithm which maps out each time the record has a change of “location”. When the next case is mapped, if it’s route is not already available then a new route is created. When a big data set is run through, thousands of possible routes will have been mapped and populated with some general statistics.

We can then test this model by clustering the location nodes against an outcome (e.g. long waiters) to investigate bottlenecks (i.e. locations with strongest correlation to outcome).

In non-linear systems with high degrees of flexibility in route choice, granular discrete event mapping may be more informative.
Annex: Useful Literature

**General Hospital Modelling**


Annex: Useful Literature

Specific Hospital Modelling