NIHR | Applied Research Collaboration South West Peninsula

@penCHORD_UoE
@peninsula_ARC

Module 1 : Introduction to OR, Data Science and Programming
Session 1A : Introduction to Operational Research and Data Science

Dr Daniel Chalk

"Say what you see"

HSMA 5

#hsma5isalive

What is OR?

You run an Emergency Department which is struggling with lengthy waiting times. In your system :

- patients arrive on average every 8 minutes
- they wait to be registered at reception, which takes an average of 2 minutes once a receptionist is free. There is 1 receptionist.
- they then wait to be triaged by a nurse, which takes an average of 5 minutes once a nurse is free. There is 1 nurse.
- they are then treated in a cubicle, which takes an average of 30 minutes once a cubicle is free. There are 3 cubicles.

You have been given the option to increase one of your resources by 1 (an additional receptionist, nurse or treatment cubicle).

How long do you think patients spend in the ED now, on average (say over 1 year)? And which resource increase are you going to pick? You have <u>until the end of this song (approx 3 minutes)</u> to decide. Post your answers in the chat.

What is Operational Research?

Let’s look at your votes.

What is Operational Research?

We could have built a model to help us answer this question.

What is Operational Research?

Base Case Scenario
A model of the current system. How are things running now?

We can use this to see how well the current system works, validate the model and identify bottlenecks.

1) How long (on average) do you think a patient spends in the ED now?
2) What resource will you increase?

"What If" Analysis
Adapting the model to reflect potential future scenarios. How might things run if we were to change x, y and / or z?

We can use this to predict the impact of decisions, and help the decision maker to make an **informed evidence-based decision**.

2) Which resource will you increase?
3) What do you predict is going to be the impact of your decision on the average time patients spend in the ED?

What is Operational Research?

We could have built a model to help us answer this question.

And we have! Let’s show it to you.

What is Operational Research?

Applying modelling, simulation and analysis techniques to help **inform** decisions, and **improve decision making.**

The Benefits of Modelling

- **Emulation :** A model is a version of reality that can be altered without risk or consequence

- **Speed :** Typically, models can be designed and built much more quickly than real world changes can be effected.

- **Communication :** A model can help people to communicate about a problem using a shared language and point of reference

- **Systems Thinking :** The process of designing the model can help people to think about their systems

- **Objectivity :** A model can provide objective support for an argument

* assuming the model has been built objectively!

What is Data Science?

Data Science uses methods from machine learning, statistics, data mining and data analysis to generate insights from data.

In OR, we start with a system and / or process, and use data to parameterise a model to emulate the system / process in-silico.

In Data Science, we start with the data, and use techniques to explore hidden patterns and structures in the data to provide us with new information.

Data Science Questions

Example OR Question :

“We want to make these changes to our process for triaging patients. What do we predict the impact will be, and what resources will we need to ensure the process is efficient?”

Example Data Science Question :

“We have lots of data on readmissions. Can we teach a machine to automatically identify which patients are likely to be readmitted?”

Modelling Approaches

In the HSMA course, you'll learn about the following Operational Research methods :

- Discrete Event Simulation
- Agent Based Simulation
- System Dynamics
- Geographic Modelling
- Forecasting Methods

and the following areas of Data Science :

- Machine Learning
- Natural Language Processing
- Network Analysis
- Analytics using R

To start you thinking about how these methods might apply to problems in your own organisation, here's a very brief overview of each.

What is Discrete Event Simulation?

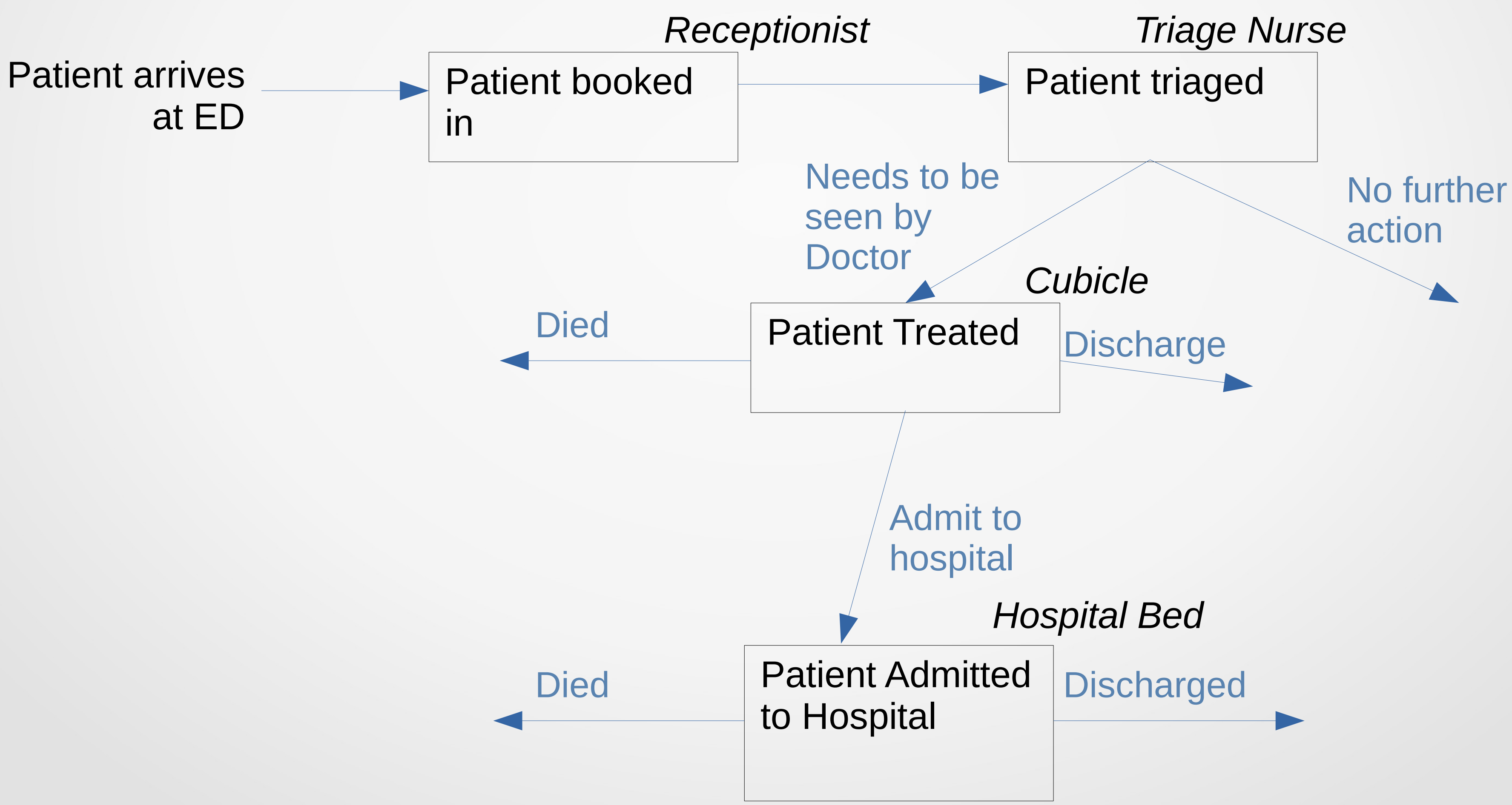
Discrete Event Simulation (DES) is a way of modelling *queuing problems*.

In a DES, *entities* flow through (and queue for) *discrete sequential processes* that use *resources*.

DES is typically used to model processes and pathways. For example, what happens to patients when they arrive at the Emergency Department.

Therefore, DES is useful for asking “what if?” questions about process / pathway changes.

An Example



Patient arrives at ED
Receptionist
Patient booked in
Triage Nurse
Patient triaged
Needs to be seen by Doctor
No further action
Cubicle
Died
Patient Treated
Discharge
Admit to hospital
Hospital Bed
Died
Patient Admitted to Hospital
Discharged

Bladder cancer = 7th most common cancer in UK (~10,000 new cases per annum)

20-25% invade muscle wall of bladder

Muscle-invasive bladder cancer (MIBC) five-year survival rate only around 50%

Definitive treatment = cystectomy + chemotherapy

Collaboration with Royal Cornwall Hospitals Trust (RCHT)

Mean time to TURBT
- reduced 3.5 weeks for fasttrack
- 9 days across all patients

Mean time to contacted by nurse specialist
- reduced 5 weeks for fastrack
- 11 days across all patients

What is Agent Based Simulation?

- A simulation method that focuses on the behaviours, interactions and motivations of individuals, and observes the population-level **emergent behaviour** that arises from these complex interactions.
- Commonly used in ecology, where often much is known about individual forager behaviours (typically known as Individual Based Modelling here)
- Extensively used in modelling disease spread / outbreaks in public health
- Different to ‘conventional’ mathematics

HARVEST (Harvesting Animals with Reinforced Values and ESTimates)

Field A

30% full
Estimate = 50%

Field B

100% full
Estimate = 70%

Field C

10% full
Estimate = 40%

GOAL :
To fill up with nectar (food) as quickly as possible

WHICH FIELD?
What do I think is the probability of me finding a full flower here / elsewhere?
What's the movement cost of moving elsewhere?

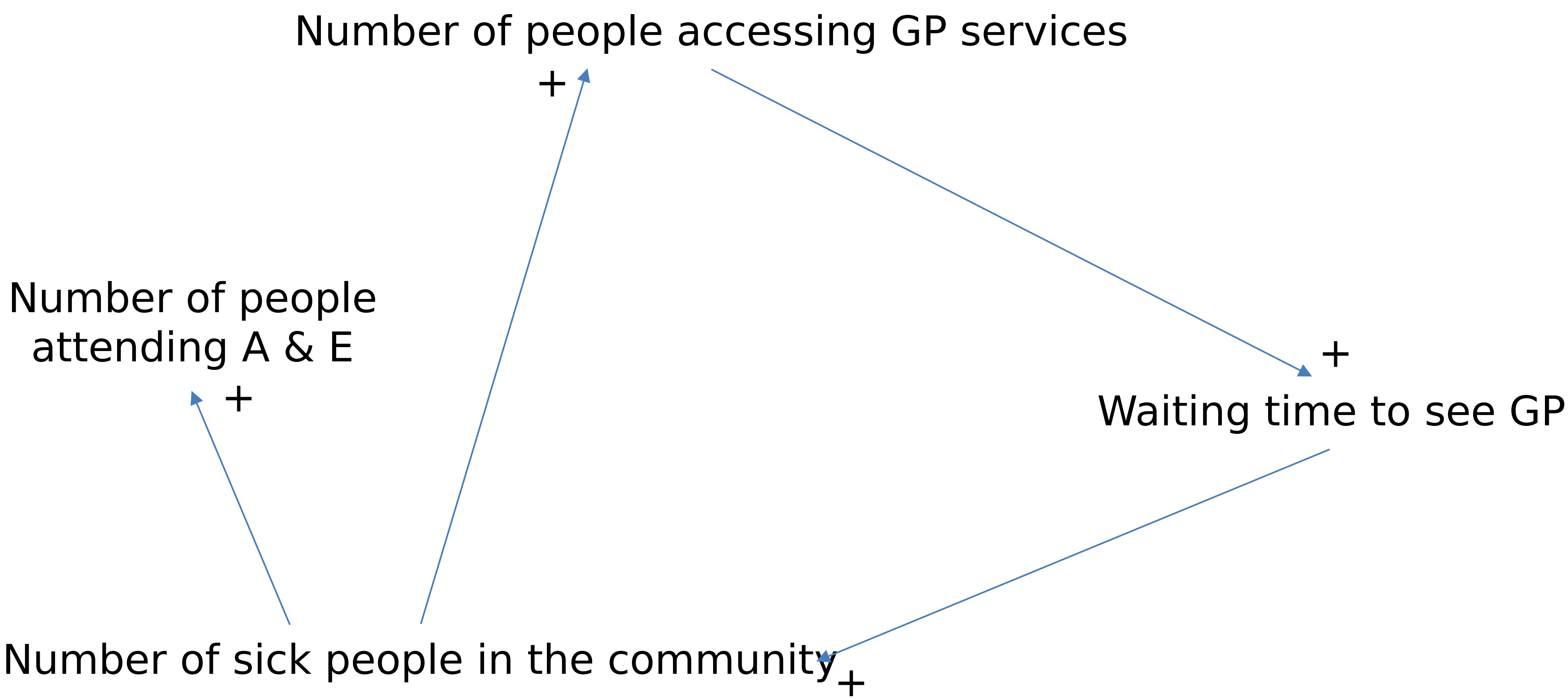
System Dynamics

- A modelling method in which system structures (components and the way in which they relate) are captured.

Fundamental principle of SD :

Structure determines behaviour

Reinforcing Loops ("Vicious Circles")



Reinforcing Loops have an ***<u>even</u>*** *number of (****or zero****) minus signs.*
Therefore the ***same conclusion is reached each time you loop.***

Using Systems Dynamics modelling during COVID-19

Background

The first accurate model to help local hospital Trusts manage resources during the COVID-19 outbreak was developed using training from our Operational Research training programme, the Health Service Modelling Associates (HSMA) Programme. System Dynamics is a form of computer modelling that tries to understand the flows within a system and was used to predict peaks in COVID-19 infection rates.

A member of our annual HSMA programme received training in Systems Dynamics while taking part in the programme and went on to develop the model. The HSMA programme helps health service employees across the South West to use data modelling to tackle a work-based research project to solve a specific issue or question for their organisation.

Activity

During the early stages of the COVID-19 outbreak it could be difficult for NHS Trusts to access the necessary data to predict admissions, leading them to free up bed space in anticipation of the levels of infection rates seen in other European countries. The use of a System Dynamics approach enables Trusts to keep up with the pace of change by producing a model to accurately predict peaks of demand. The model has been widely adopted after being presented to MPs.

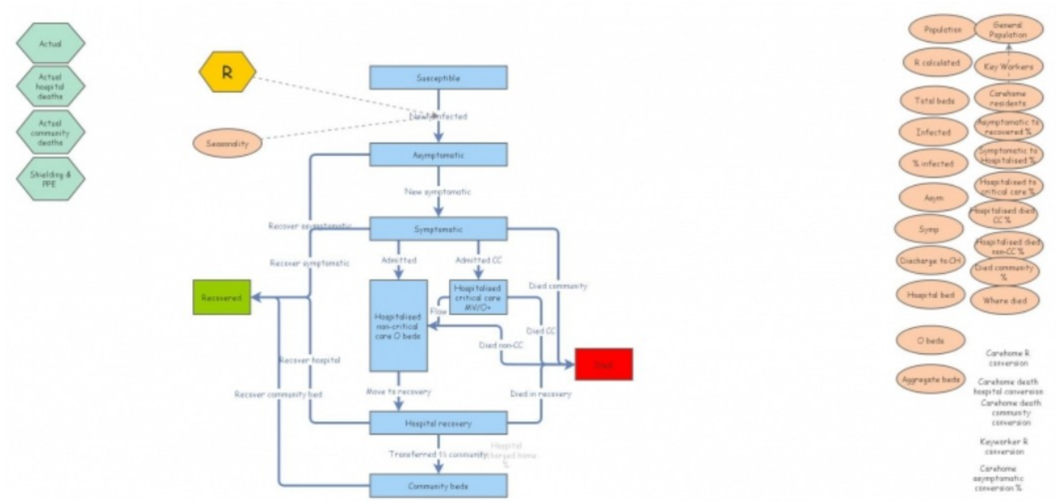
Speaking about the impact of the model a clinician commented: "One of the specific advantages of the System Dynamics approach is that clinical commissioners could really see how the model worked and why. The tool also offers a very visual way to engage with the data."

Outputs

Related Projects

The Health Service Modelling Associates Programme

The Police Services Modelling Associates Programme



News about this project

Health Services Modelling Associate uses new skills to manage Covid-19 vaccination programme

Lead collaborators

David Spencer

https://arc-swp.nihr.ac.uk/research/projects/systems-dynamics-modelling-covid19/

Geographic Modelling

Where?

Where to place services…

Where will patients go…

Where do we find people with this disease…

Predictions

…if our demand continued to grow?

…if we close these services?

…and how is it spreading?

Project Question

How do we optimise the geographic configuration of acute stroke services in:

1) South West (14 possible locations)
2) England (127 possible locations)

In order to:

1) minimise patient travel
2) maintain hospital admissions 600-2,000/year

Scenarios

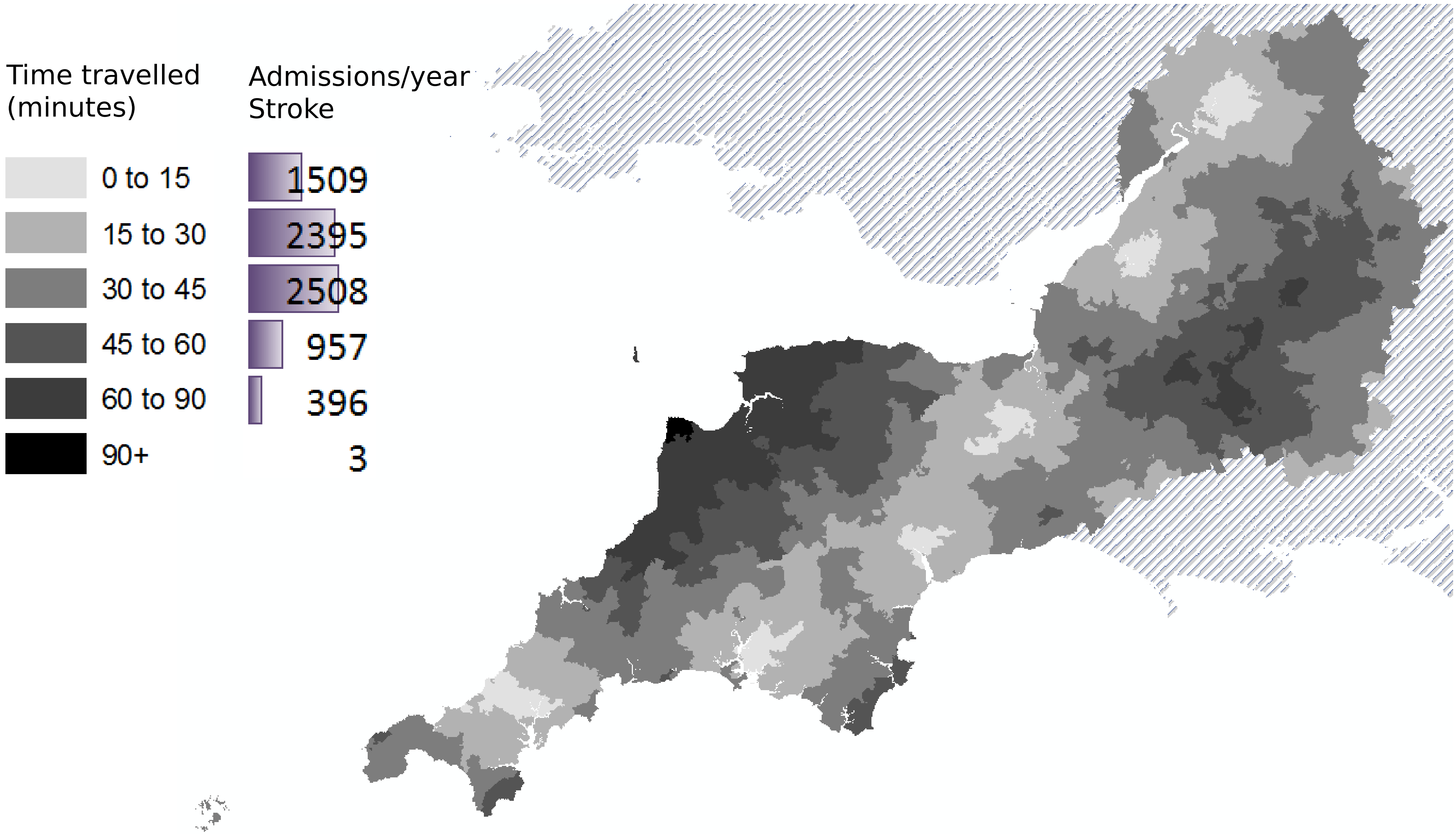
1) South West: All possible configurations (“brute force”)
2) England: Search for good configurations

Aim

Inform decision makers so they can better understand the system

SW: Map an example configuration

6 hospital solution: Vascular surgical centres



Forecasting

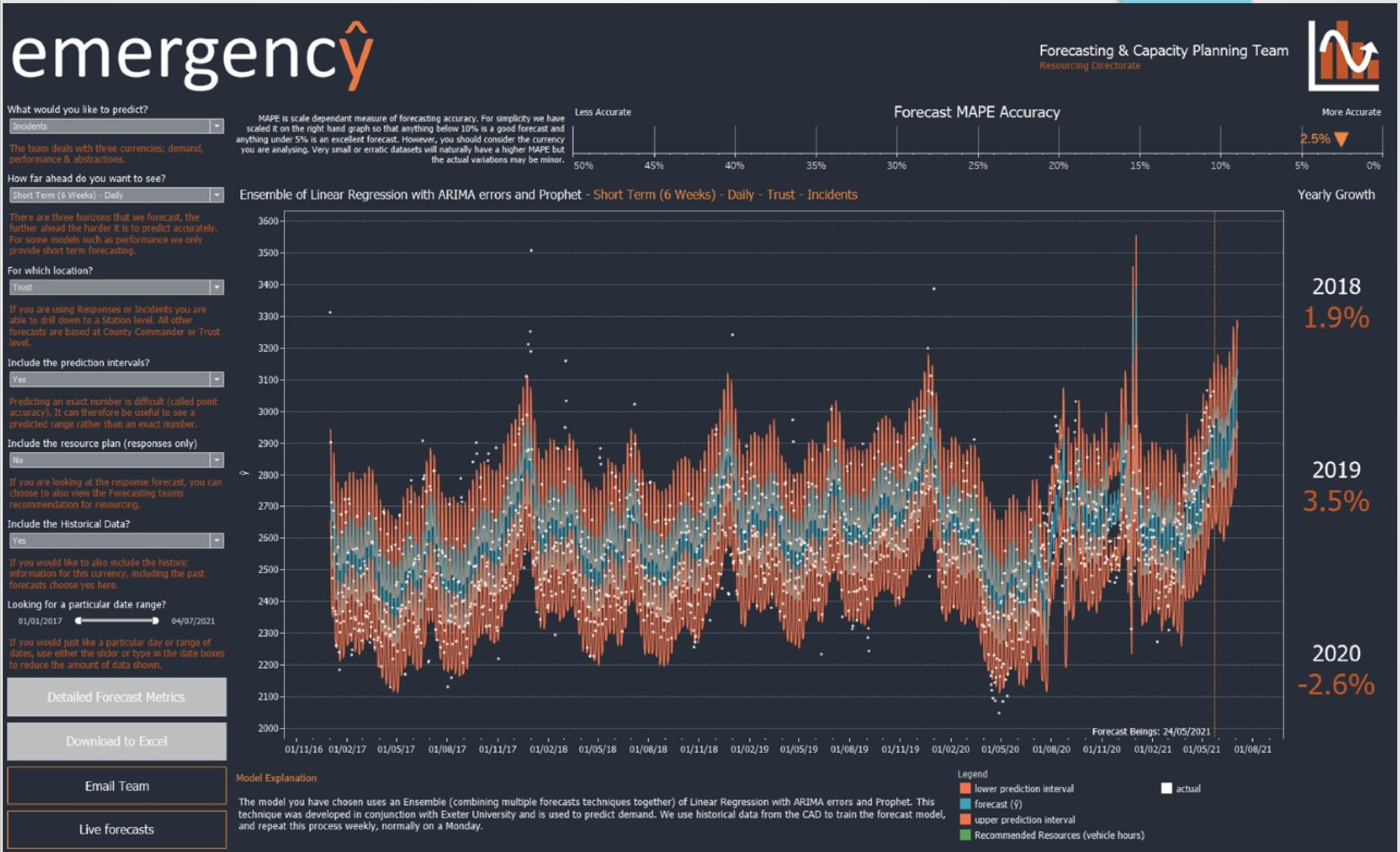
Forecasting methods try to make predictions about the future based on patterns observed in the past.

Forecasting typically involves trying to predict a future level of activity in a service to enable better planning of the resources need to meet that activity.

In the HSMA course, you'll learn about simple forecasting methods, as well as more cutting-edge methods, that use Neural Networks to try to make better predictions.

But they are not **crystal balls**.

Example : Forecasting Ambulance Demand



Machine Learning

In Machine Learning, our aim is to get a machine to *learn* to make predictions and *improve* from data. The machine is looking for hidden patterns in the data that can help determine “if x, y and z are like this, then I think the outcome would be this”

Supervised Learning is guided by feedback from examples, which tells the machine how well it is performing, and allows it to improve over time.

Unsupervised Learning doesn’t have “correct” examples to show the machine. Instead the machine looks for hidden structures in the data to try and organise the data.

Reinforcement Learning consists of *agents* interacting with an *environment* using trial and error. Each interaction provides a *signal* to the agent, which rewards the agent when the interaction is positive, and punishes the agent when it is negative. Agents gradually learn to undertake actions that lead to reward.

Current / Previous ML Projects

Predicting violent incidents on wards in inpatient mental health units **HSMA 4 Project**

Can we teach a machine to predict which patients are likely to not attend an appointment, and can we use this information to inform interventions to improve access? **HSMA 4 Project**

Predicting unplanned acute hospital admissions **HSMA 4 Project**

Can we teach a machine to determine when a stroke patient should receive clot-busting treatment, to help inform clinicians and improve the consistency of decision making? **SAMueL Project**
(https://arc-swp.nihr.ac.uk/research/projects/samuel-stroke-audit-machine-learning/)

What is Natural Language Processing?

Natural Language Processing (NLP) is focused on the way in which human language and computers interact.

There are two key sides to Natural Language Processing :

- Trying to get computers to understand natural language (speech recognition, building AI that can understand human interaction etc)
- Getting computers to learn how we form our language so that they can help us automate information extraction from free text

We focus on the second of these areas in the HSMA course.

Automating free text information extraction

There are a number of ways in which we can automate the process of extracting information from free text :

1. Named Entity Recognition – extracting named entities (such as people, places, organisations etc) from unstructured text, and categorising them.

2. Sentiment Analysis – automatically identifying whether unstructured text is positive, negative or neutral in tone. Can also work alongside Sarcasm Analysis.

3. Relation Extraction – identifying the relationship between entities in text, and automating the extraction of the entities and their relationships into Relational Tuples

Previous NLP Projects

- Can we find clues in GP and community patient notes that could be automatically extracted and fed into a ML algorithm to predict imminent admission? **HSMA 3 Project and HSMA 4 Project**

- Building a tool that automatically identifies the sentiment of tweets to every police force in the country, and identifies what people are talking about, to inform public engagement **HSMA 3 Project**

- Automatic classification of patient surveys as positive or negative, and extracting what people are talking about **HSMA 3 Project and HSMA 4 Project**

- Can we extract information from Modus Operandi (MO) text to identify the links between offenders and their victims? **HSMA 3 Project**

- Can we automatically extract information from ambulance crew notes about the procedures undertaken at scene? **PenCHORD Project**

Network Analysis

What is a network

Network graphs are used to:

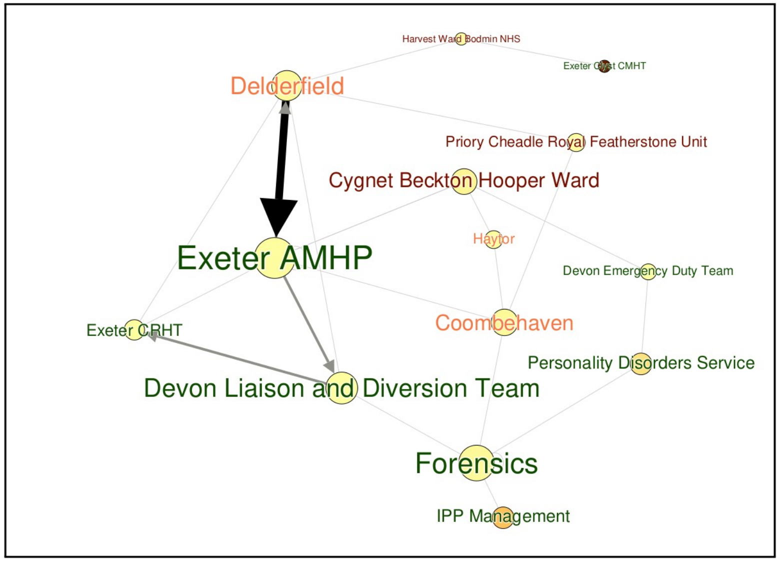
- Study the function of large and/or complex systems
- Visualise relationships and interactions in a intuitive way
- Represent abstract concepts in a concrete way

Network Analysis

Network-based Operational Modelling | THIS.Institute | NIHR | Applied Research Collaboration South West Peninsula | UNIVERSITY OF EXETER

Network-based operational modelling for healthcare

- Graph as service structure
- Node as service
- Edge as patient movement
- Edge weight as number of patients
- Node attributes as improvement measures



Analytics in R

R is a statistical programming language that is fast growing in popularity, and is currently being pushed in the NHS as part of the future of analytics.

R provides a powerful way to *automate* and more *efficiently* undertake routine data analysis tasks traditionally undertaken in Excel.

Excel is a spreadsheet software package. Unless you’re undertaking financial calculations, step away from the spreadsheet… :)

Exercise 1

Now you've been introduced to some of the methods / approaches we'll be covering on the course, I want you to take a little bit of time to start discussing some ideas for how some of this stuff might be used in your organisations.

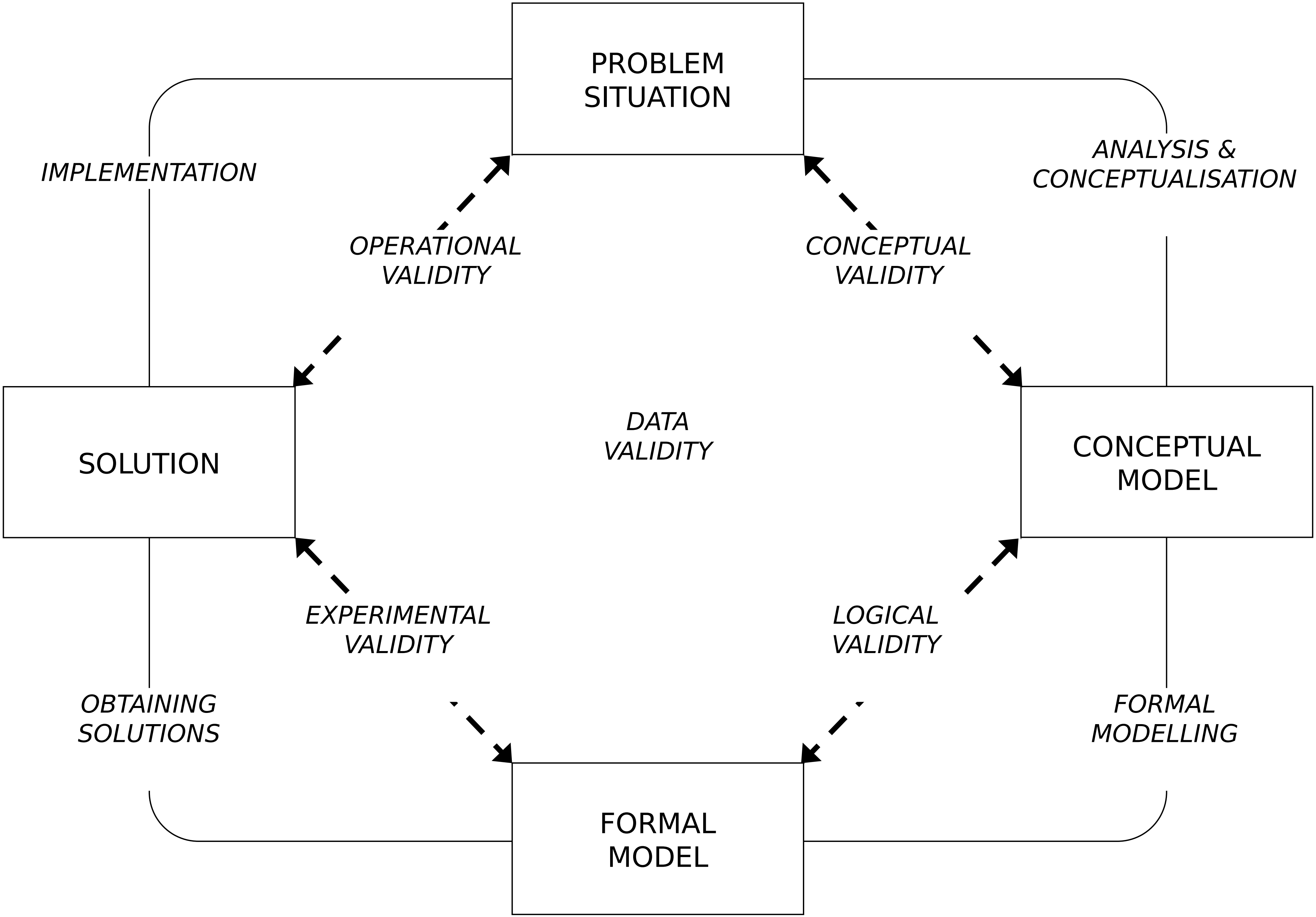
You'll work in your your **Peer Support Groups**. Each group will have a breakout room, and you join that breakout room.

Spend the first 5 - 10 minutes introducing yourselves and getting to know each other – you're going to be working together for the next 6 months! Then, come up with some ideas about how some of the methods we've just talked about could be used to tackle problems in your organisations. This will help you start thinking about potential project ideas (although we're not expecting you to come up with those quite yet!).

You have **25 minutes**, and then a **10 minute comfort break**, before we resume. When you come back, I'll ask a few of the groups to share some of the ideas you came up with, so nominate someone to be your spokesperson in case your group is picked.

Conceptual Modelling and Problem Structuring

The Modelling Cycle



Objectives of the Model

- What are you trying to achieve / why are you building the model?

Problem Statement	“What if?” Question(s)	Deliverables	Organisational Impact
There are significant delays referral to treatment	What if we reorganised the testing priority?	A report outlining the predicted results	More lives saved, targets met




Scope

- What are the boundaries of the system I need to model?

I need to model my ED

But many arrive by ambulance

So should I build a model of ambulance dispatch too?

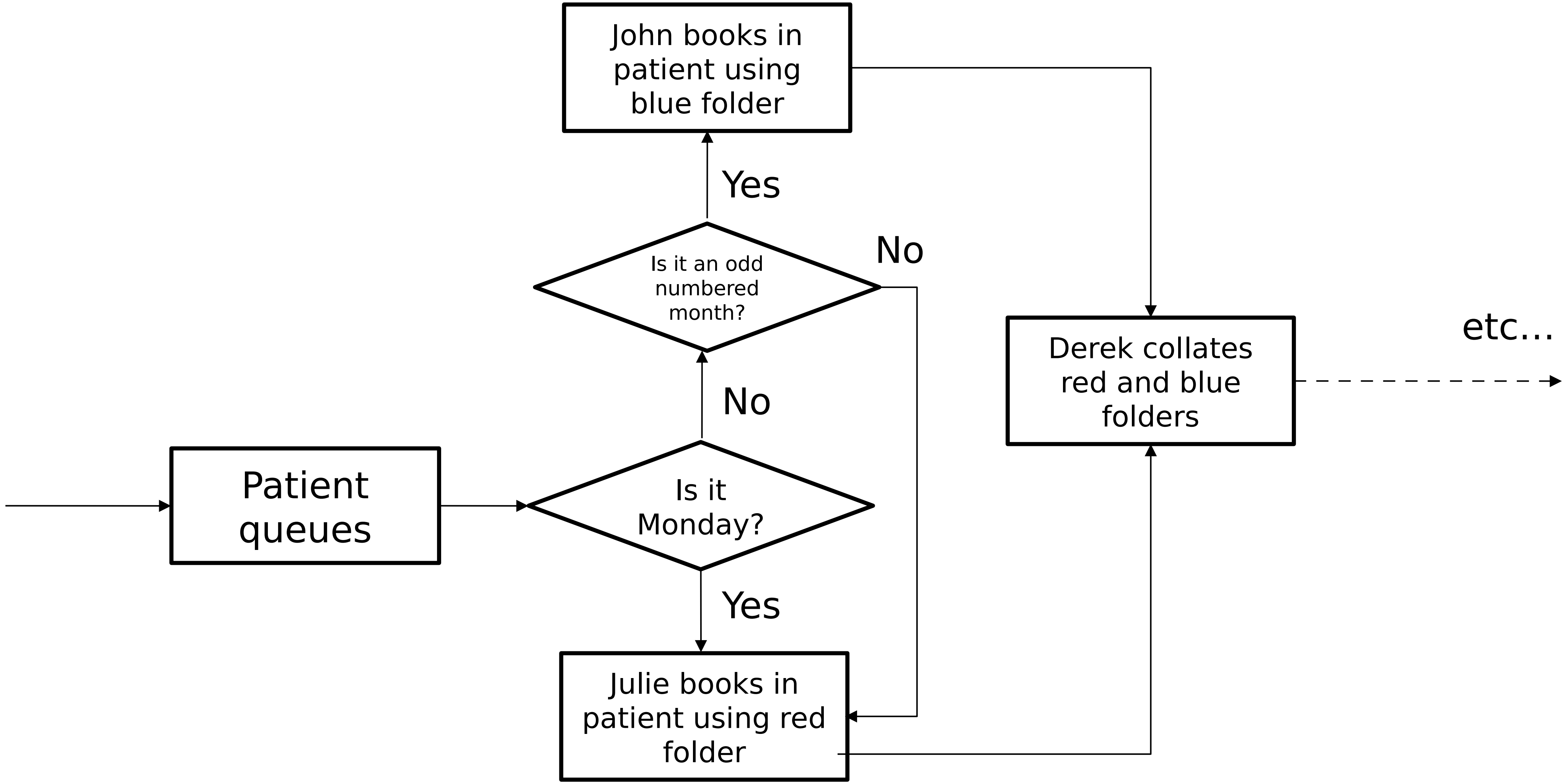
What’s the minimum you can model to answer your question?

If I need to model other systems, how can I simplify their representation?

Level of Detail

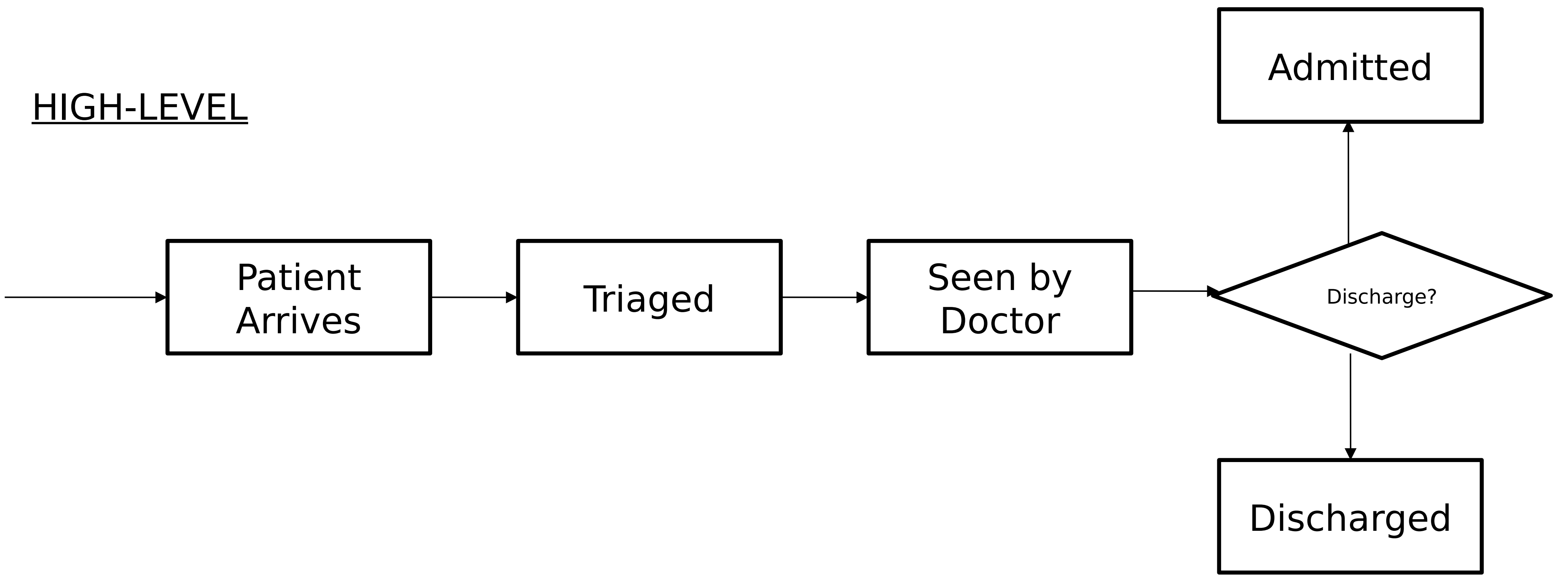
- How much detail do I need to put into the model?

LOW-LEVEL



Level of Detail

- How much detail do I need to put into the model?



Level of Detail

- How much detail do I need to put into the model?

What's the minimum you can model to answer your question?

If I need to model other systems, how can I simplify their representation?

Assumptions and Simplifications

- Assumptions are things that we must assume because we don’t / can’t know their real world properties
  - We assume that the data we’ve got is representative
  - We assume there are no travel times within the clinic for staff or patients (or that they’re trivial)
- Simplifications are things from the real world that we choose to distil down to simpler elements because we anticipate that added complexity does not provide benefit
  - We simplify the triage process into the patient spending an amount of time with the nurse
  - We simplify such that there are no limits to the queuing time for the MIU

All models are wrong….

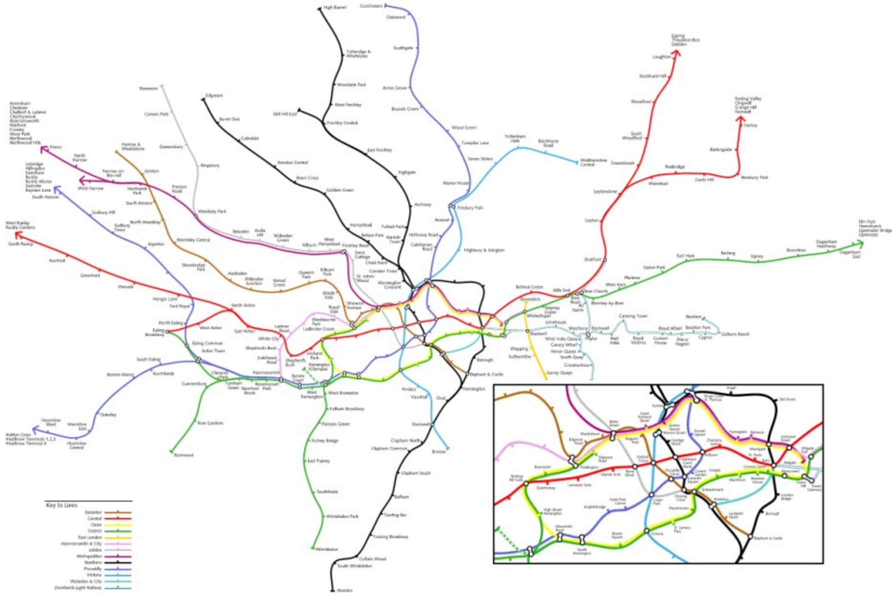
- … because they are a simplification of reality
  - … they miss out and ignore bits of the real world
- In the natural and physical sciences some models are extremely accurate predictors even though they are simplifications;
- In other areas such as health care operations a model may not be able to predict what will happen accurately…

but some are still useful….

- … because they are a simplification of reality

Tube map

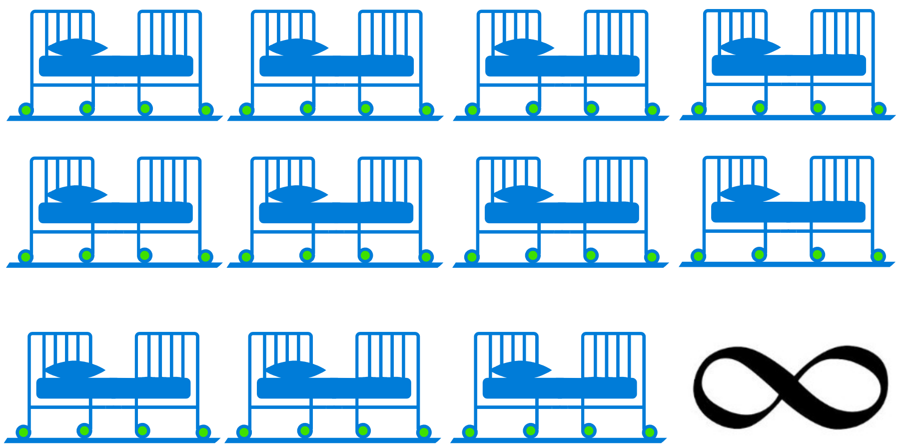
Actual layout of the underground



This model is wrong…

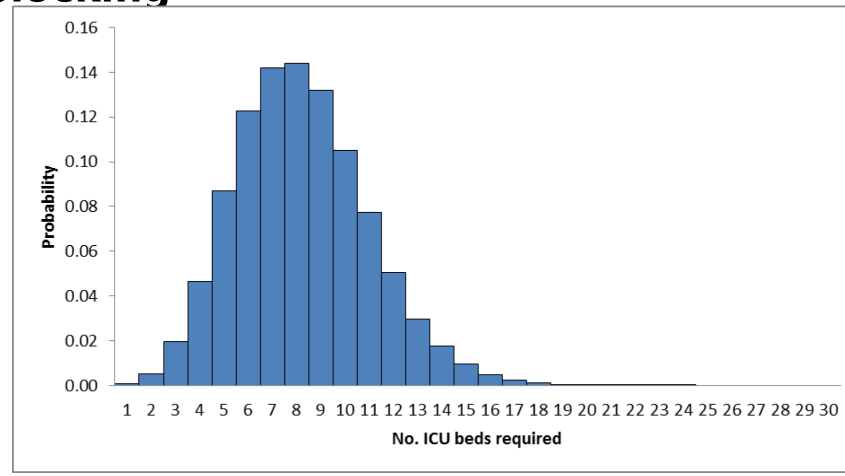
but very useful

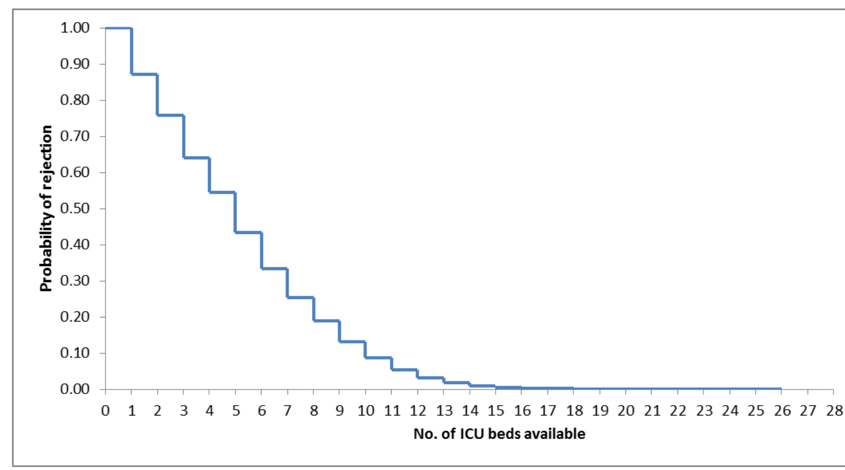
A model of a cardiac ICU with infinite capacity



This model is wrong…

Output: trade-off between beds and blocking





but useful

Models often come in two flavours of simplification

Simple models, that give you a feel for the direction of travel

High fidelity models that estimate where you are, where you are going & what time you will get there.

As a user / creator of models it is your job to distinguish between the two and make sure results are not misrepresented!

What is Process Mapping?

A means of capturing the discrete processes within a system, and the potential inputs and outputs to these processes, in order to better understand how a system works.

Process Mapping is a <u>vital</u> tool for understanding how your system works.

How can you improve a system if you don’t know what your system looks like?

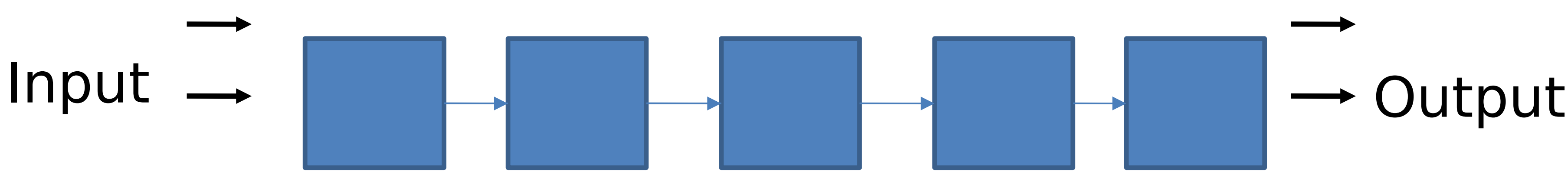
What is a Process?

Processes are activities that have a number of **steps** and which transform **inputs** into **outputs**.

In this way, processes can be thought of as functions.

Output = function (inputs)

Y = f(x)



The “Secret” of Process Mapping

Process mapping is extremely easy if you just bear in mind one simple rule :

The "Secret" of Process Mapping

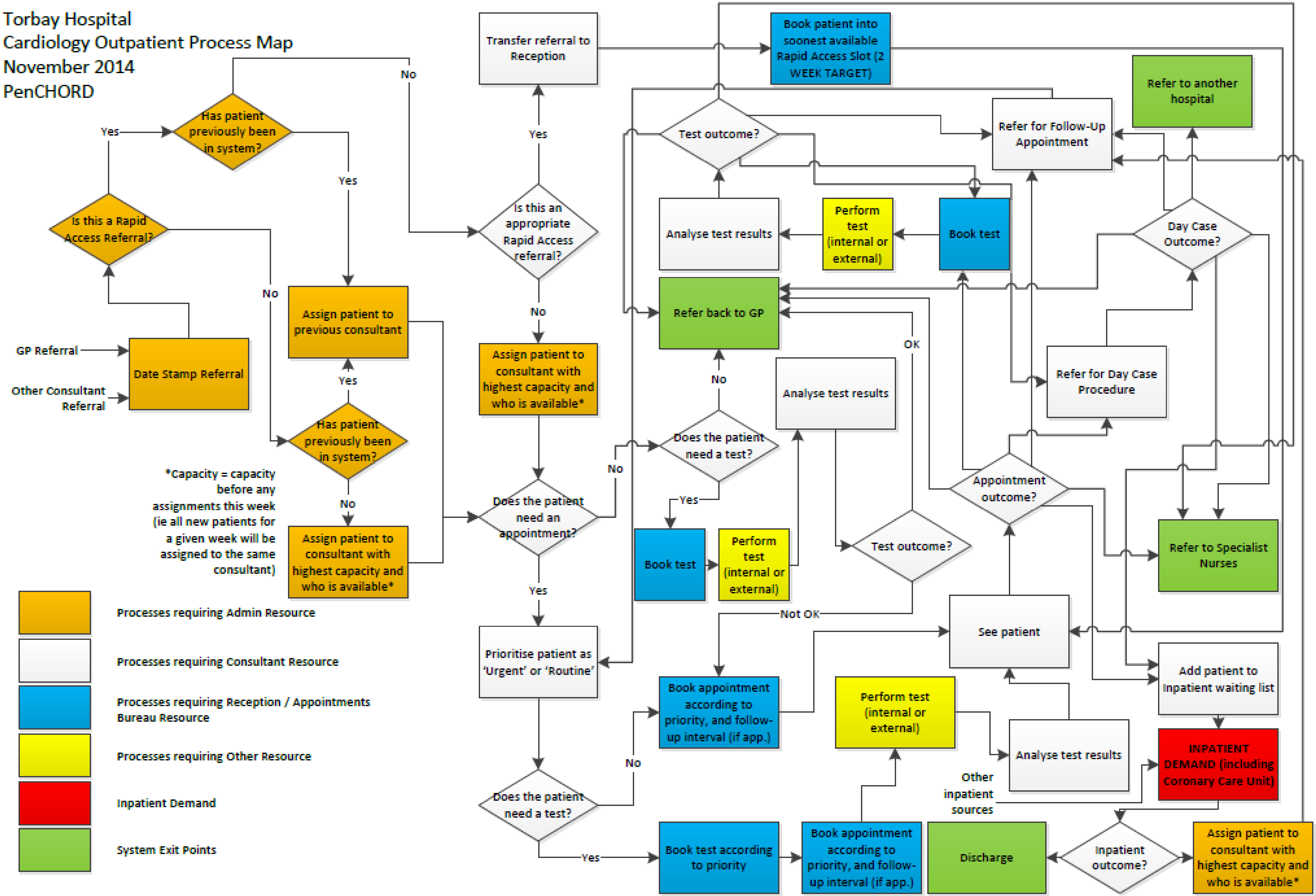
Process mapping is extremely easy if you just bear in mind one simple rule :

Say what you see!

Top Process Mapping Tips

- Have clear start and end points (there may be multiple of each – that’s fine as long as they’re well defined)
- Describe what <u>really</u> happens, not what should happen
- Don’t worry about one-off exceptions to the process unless they’re important for your question
- Process map as a team – **everyone** has a different perception of the “true” system and **everyone** should contribute
- Think about the journey of the entities (patients, test results etc) – how do they flow between processes in your system?
- Create a visual representation of your system – **a Process Map**

Example Process Maps

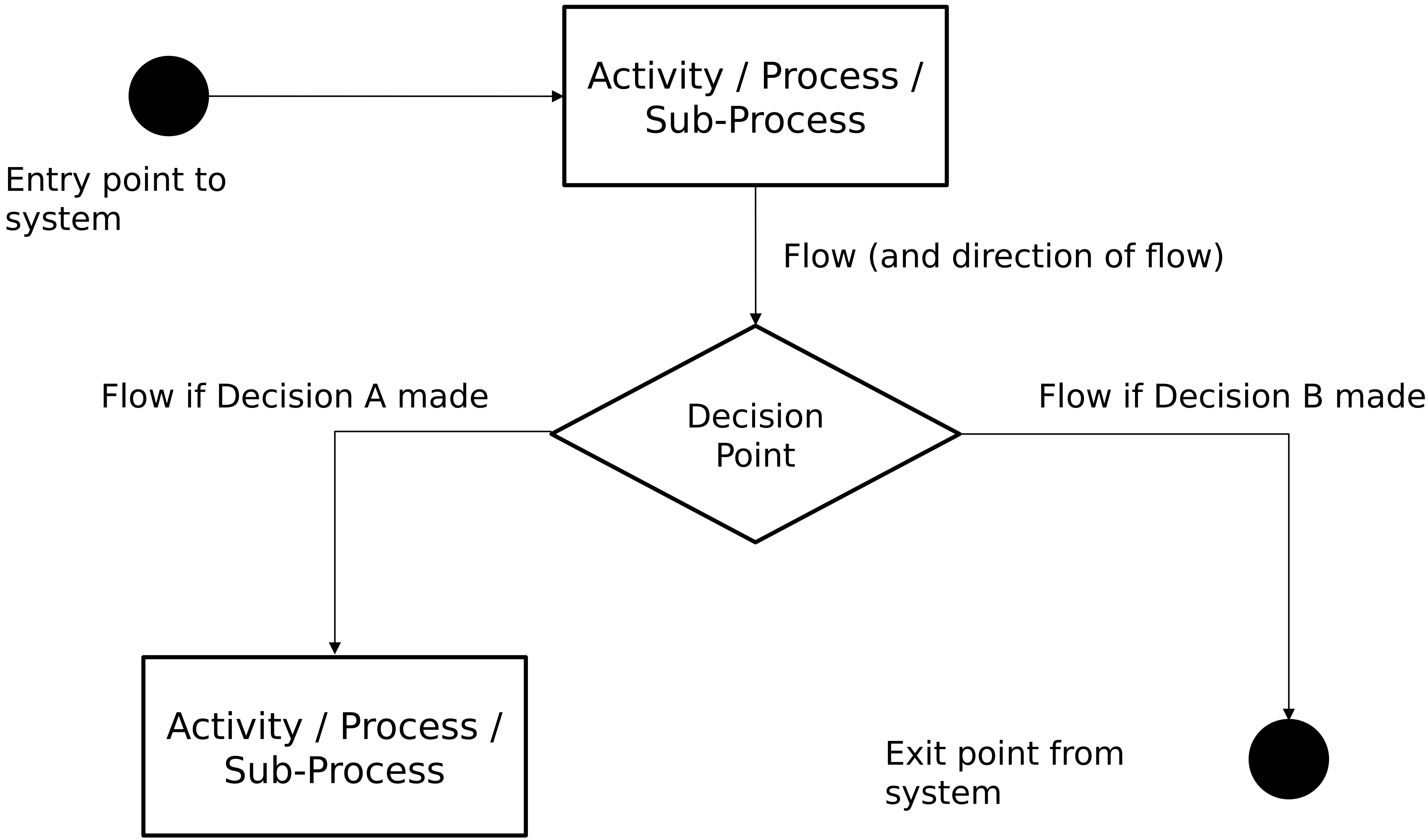


Format of Process Maps

There are different ways to visualise process maps, and there is no single “right” way of doing things.

However, it is common to use elements of Flow Chart nomenclature as a framework for developing a Process Map, because it allows you to capture the flows, processes, decisions, start and end points that are common in a process map.

Format of Process Maps



Who are Stakeholders?

Those people who have an interest in and / or influence on a problem and / or its solution.

These questions may help you identify them :

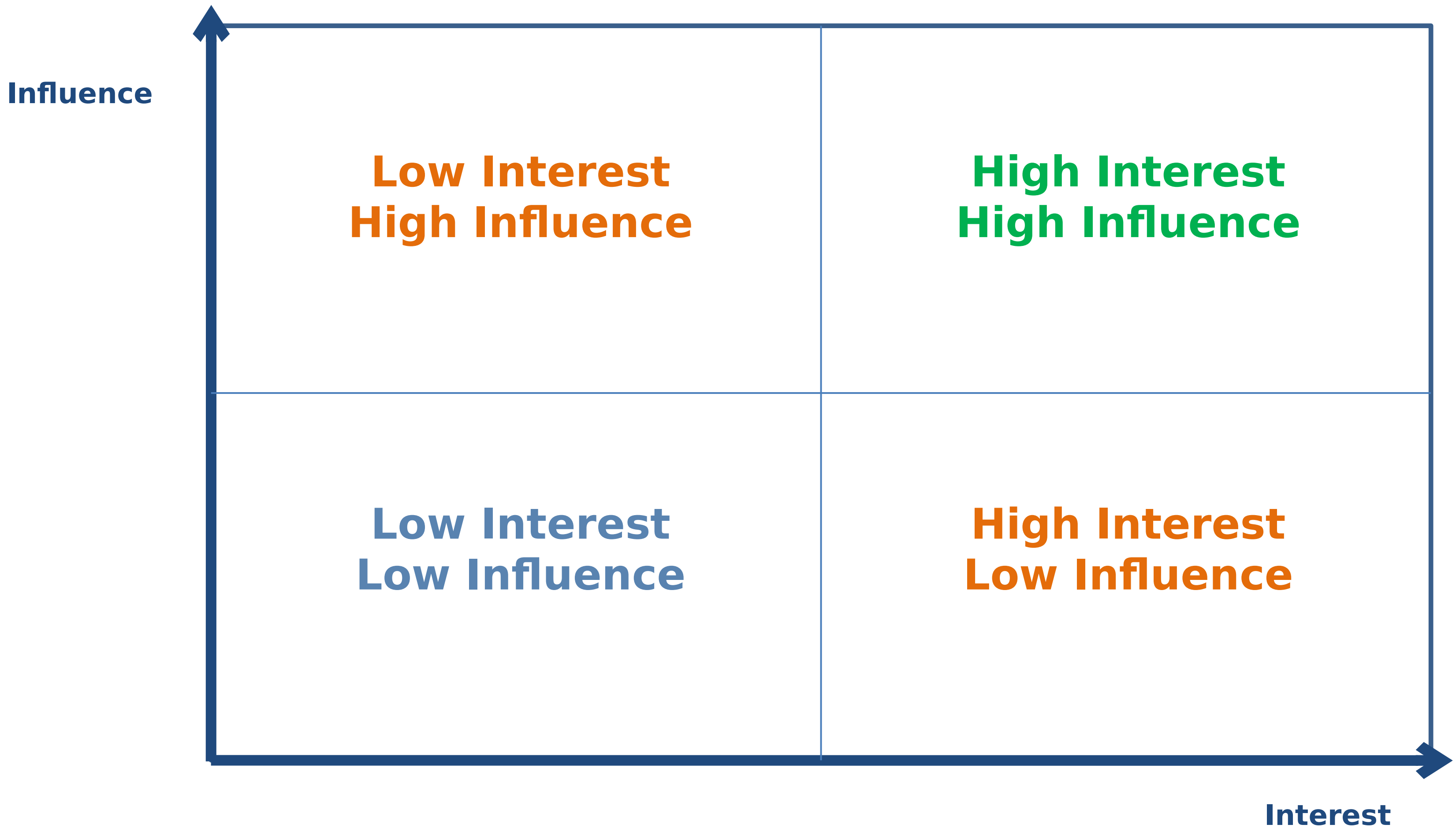
1) Who owns the problem?
2) Who might benefit from its solution?
3) Who might suffer from its solution?
4) Who has a legitimate interest in the problem and / or its solution?
5) Who has the power to influence the process of solving the problem and implementing the solution?

Why Consider the Stakeholders?

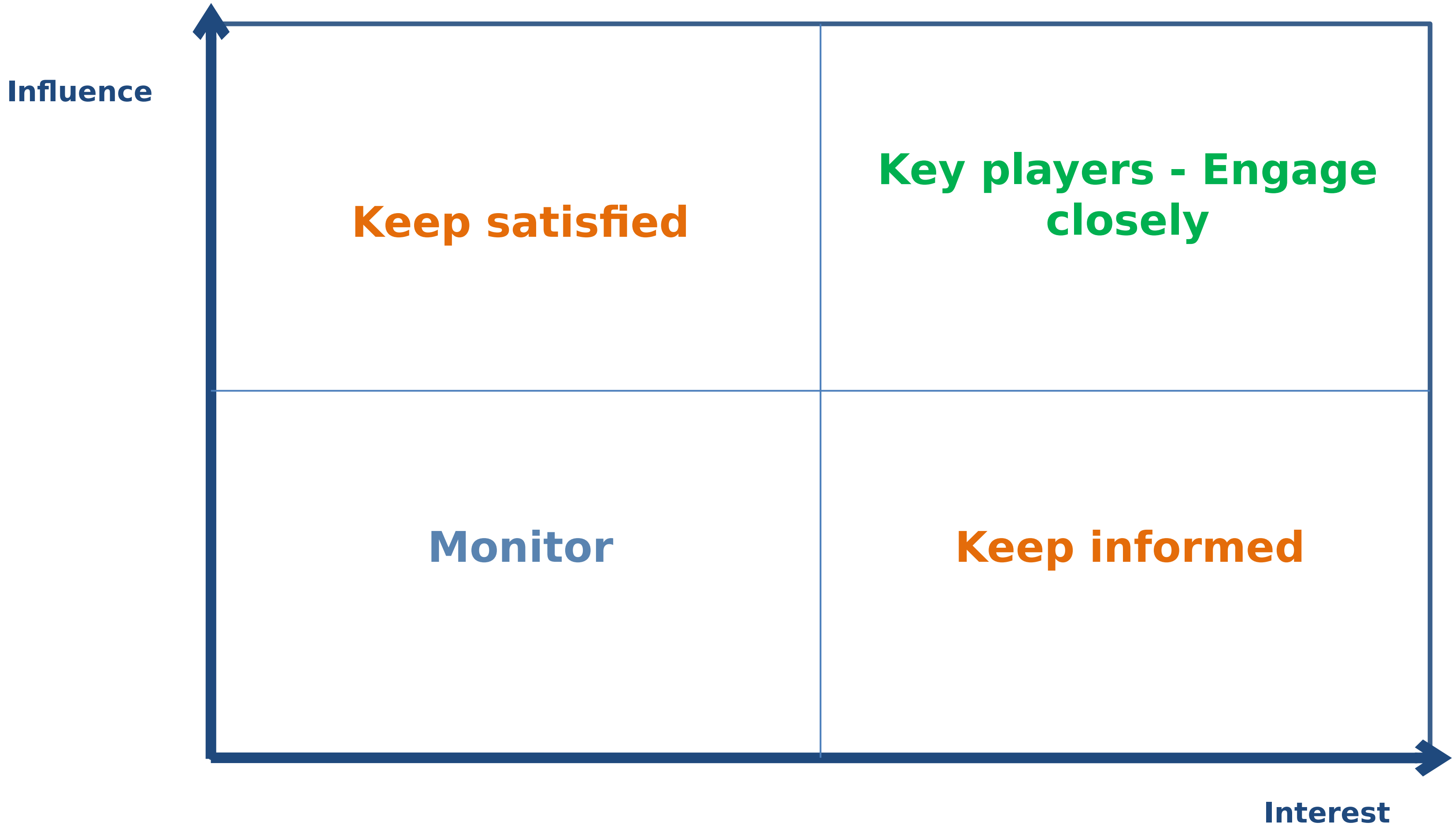
Why is it important to consider who are the stakeholders?

- You need to know who holds information and / or data you’ll need
- You need to know who might help or hinder solving the problem or implementing the solution
- You need to understand the nature of any conflicts (or potential conflicts), how that might impact the work, and how they might be mitigated

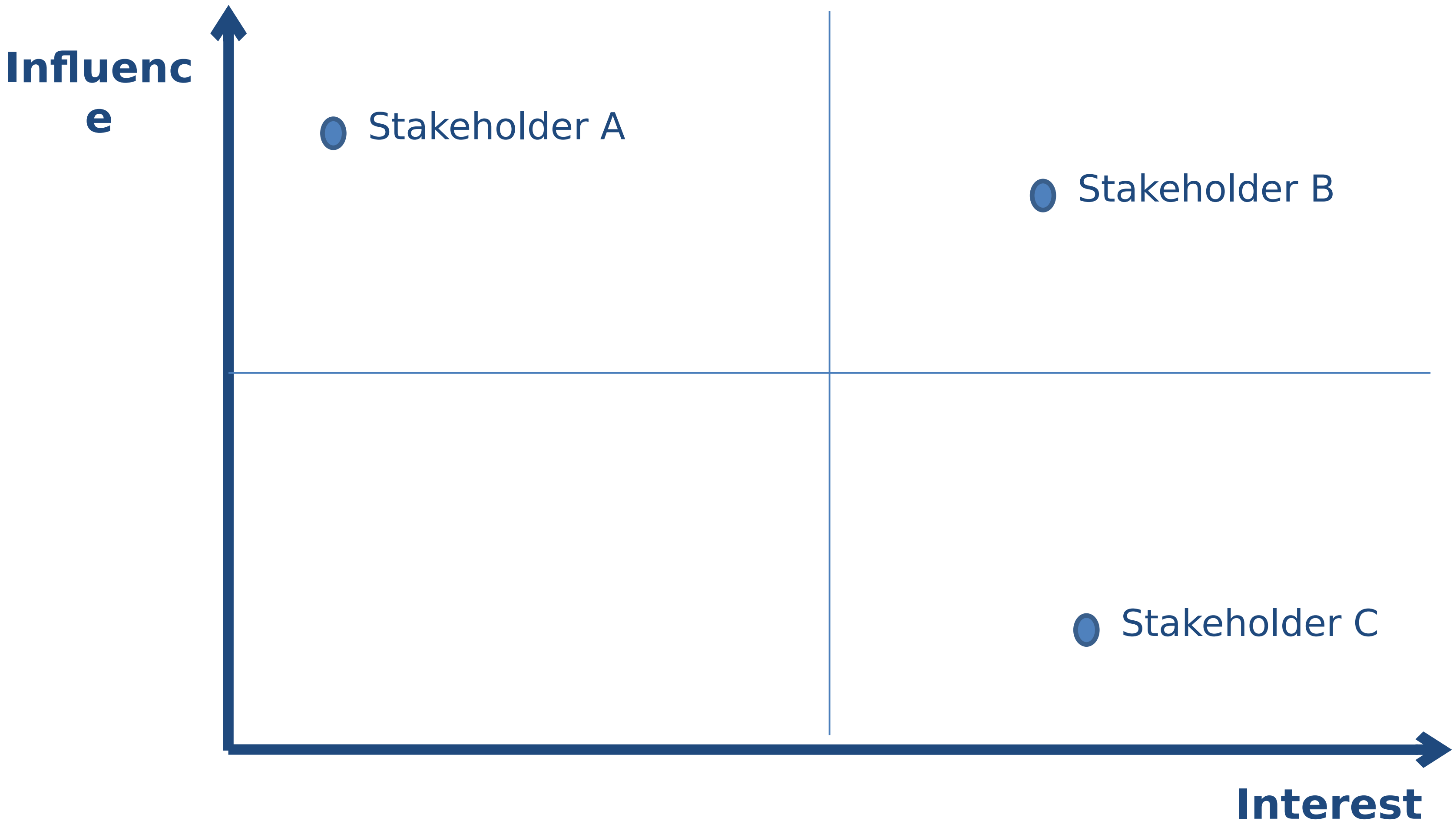
Interest / Influence Diagram



Interest / Influence Diagram



Interest / Influence Diagram



Exercise 2

You have been provided with an example problem (exercise_2.pdf). In your groups, you will now spend the next 30 minutes :

1) identifying the “what if?” question(s) in the problem

2) building a process map of the “base case” system that’s being described

3) building an interest-influence diagram based on stakeholders you think you would need to consider if undertaking a project on this fictional problem.

Once we return, I’ll ask a couple groups to present what they’ve done.

Distributions

Let’s say we want to build a model of an Emergency Department Triage process. We know that patients are arriving on average every 5 minutes.

We could tell the model to put a new patient into the system every 5 minutes.

Now let’s imagine that the time a patient takes to be triaged is also, on average, 5 minutes.

We could tell the model that each patient spends 5 minutes with the triage nurse.

Sounds reasonable, right? Let’s see how this model would pan out.

Distributions

Patient 1 arrives at 0700. They are seen by the triage nurse. They finish with the triage nurse at 0705.

Patient 2 arrives at 0705. They are immediately seen by the triage nurse, who has just finished with Patient 1. They finish with the triage nurse at 0710.

… (some time later) …

Patient 73 arrives at 1300. The patient is delighted to discover that there is no queue for the triage nurse (who incidentally is looking quite tired…) and they see them straight away. They finish with the triage nurse at 1305. The triage nurse lets out at a sigh as they spy Patient 74 coming through the door…

Distributions

What a wonderfully efficient system this is (although the triage nurse may not be quite so happy about it!).

This is clearly a very accurate model of how real world triage processes work.....

Except, that’s not how things work.

Why is the model not capturing that?

Because we haven’t accounted for variability.

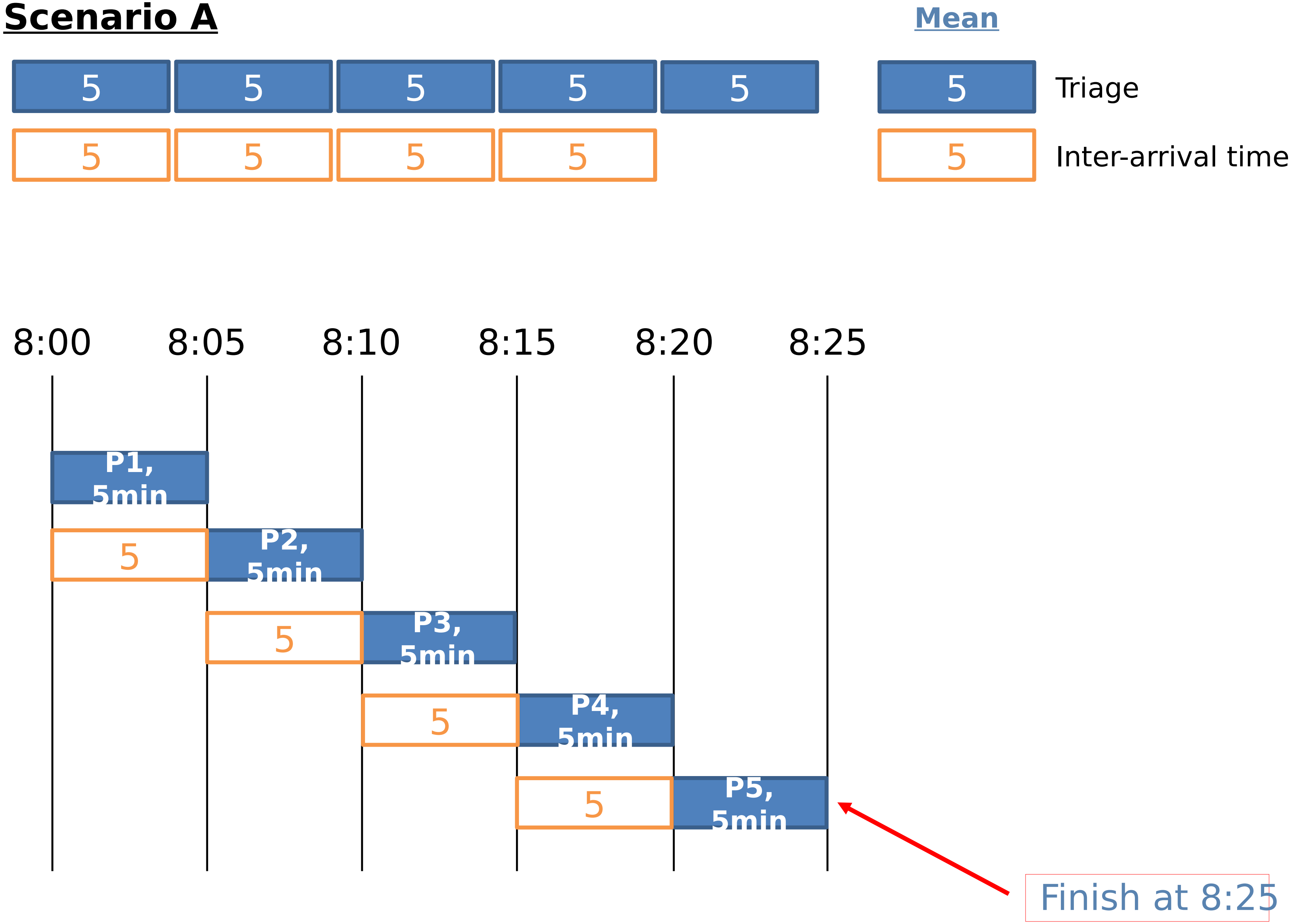
Scenario A

						Mean
Triage	5	5	5	5	5	5
Inter-arrival time	5	5	5	5		5

Scenario B

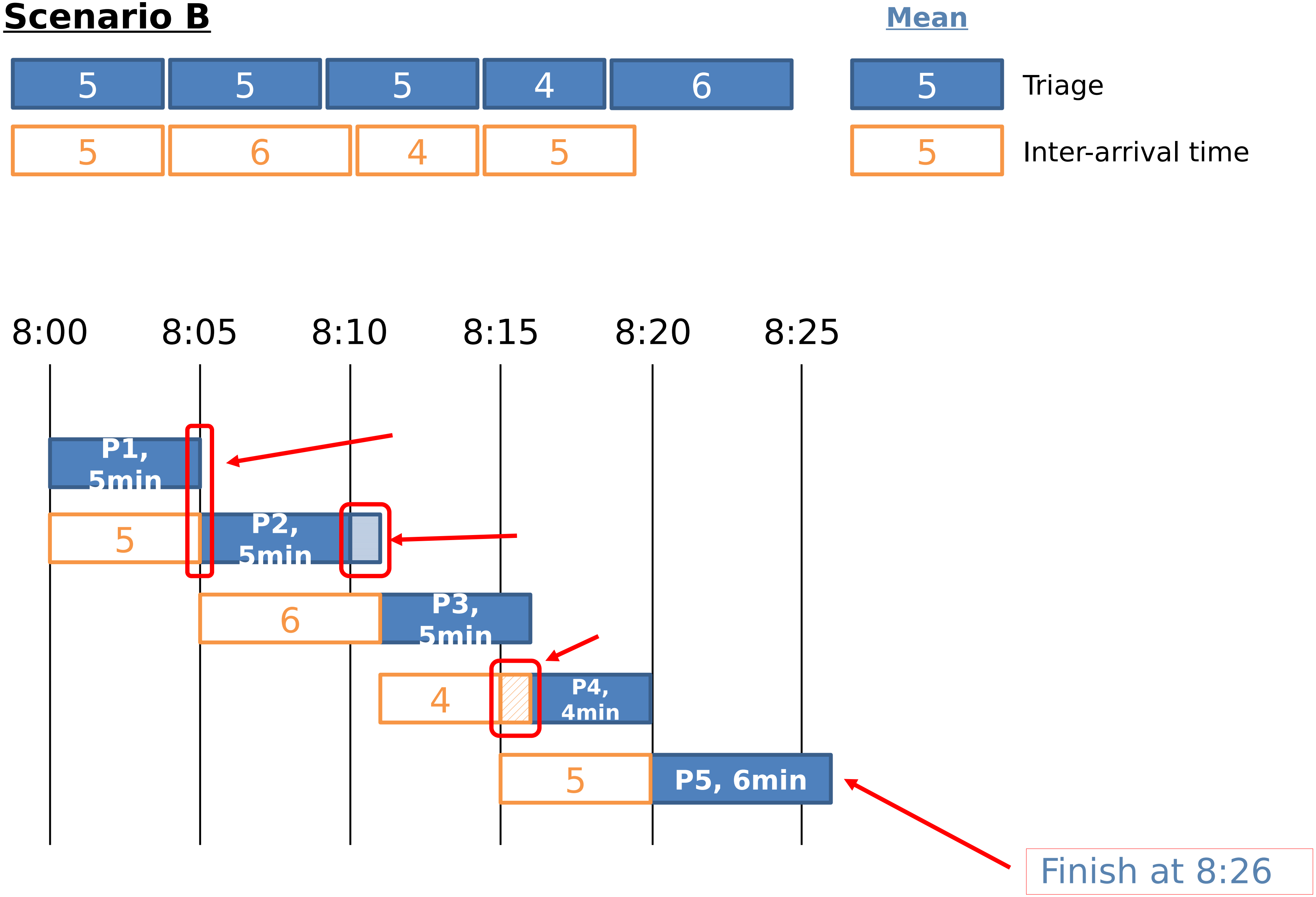
						Mean
Triage	5	5	5	4	6	5
Inter-arrival time	5	6	4	5		5

Scenario A



Mean
5
5
5
5
5
5
Triage
5
5
5
5
5
Inter-arrival time
8:00
8:05
8:10
8:15
8:20
8:25
P1, 5min
5
P2, 5min
5
P3, 5min
5
P4, 5min
5
P5, 5min
Finish at 8:25

Scenario B


Mean
5
5
5
4
6
5
Triage
5
6
4
5
5
Inter-arrival time
8:00
8:05
8:10
8:15
8:20
8:25
P1, 5min
5
P2, 5min
6
P3, 5min
4
P4, 4min
5
P5, 6min
Finish at 8:26

Distributions

Patients may well arrive, on average, every 5 minutes. But this doesn't mean that every patient turns up every 5 minutes on the dot.

Similarly, patients may well spend an average of 5 minutes with the triage nurse. But that doesn't mean the nurse has a stopwatch and stops triaging them in once the five minutes are up, or drags it out if it's done quicker.

Real life is full of variability. And if our models are to be as accurate as possible, they should account for this variability as much as they can.

To do this, we can use distributions.

Distributions

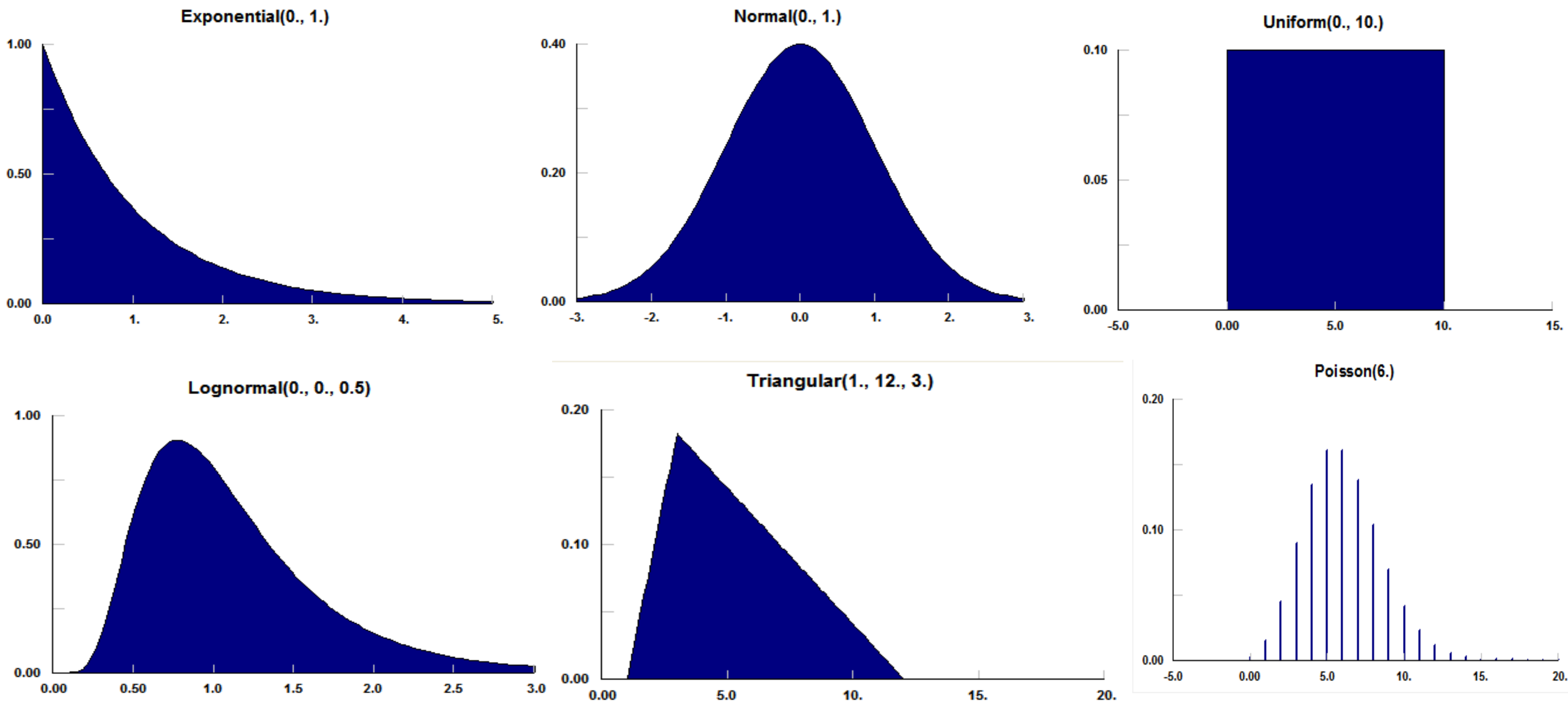
A distribution is a way of representing the variability within data. It provides us with an estimate of the probability of a value occurring in the future.

For example, a distribution might tell us that 30% of patients in the past have spent 6 minutes with the triage nurse. We can then say that, for each new patient coming in, there is a 30% chance that the time they spend with the triage nurse will be 6 minutes.

There are lots of “named distributions” available – distributions that have certain shapes and characteristics. The one you choose will depend on the shape of your real world data – you want to find one that best “fits” the shape of your data.

Each distribution is ‘defined’ by zero to many parameters. These parameters can specify the skew, range etc of the distribution.

Some Distributions



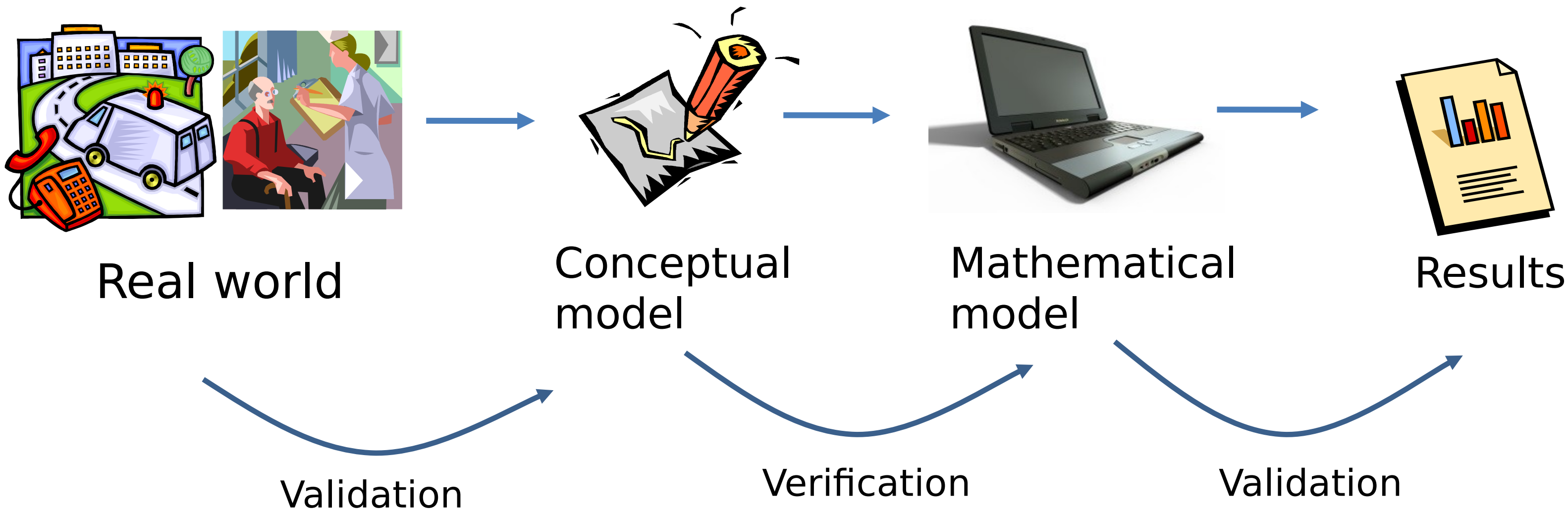
Exponential – common for inter-arrival times. Also good to use for everything initially as it's easy to change the parameters. Then move on to others.

Log Normal – common for process times. But start with exponential for first draft model.

Poisson – describes the number of arrivals in any given period if arrival is random

Triangular – useful when data is limited

Model Validation and Verification



- **Validation**: the process of determining whether a model is *sufficiently accurate for the particular objectives of the study*

- **Verification:** the process of determining if the conceptual model has been correctly translated into a mathematical or computer model (testing).

Simple 'Black Box' Validation

- A simple approach to validating models is to compare the model's predicted outputs with data from the real system
  - *E.g. a comparison of modelled and historical average queuing times*

- Issues?
  - Ignores the internal workings of the model
  - How do you validate a what-if scenario (no real world to compare against)?
  - Often real world data are unavailable or inaccurate

- Black box validation is a good start, but not enough! We need to:
  - Validate smaller parts of the model
  - Validate input data
  - Work closely with subject matter experts
  - Compare results to similar (sometimes simpler) models
  - Assess uncertainty

Confidence not validity

- It is not possible to prove that a model is valid
    - But you can disprove that it is valid
- View validation as a process of building confidence that the model and its results are sufficiently accurate for the purpose it was built to address
- Confidence thresholds vary by user and modeller.
    - This means that a valid model may not be viewed as credible by users
    - This means that an invalid model may be viewed as credible by users!
- Do not fall into the trap of thinking that a valid model is the most ‘realistic one’

Parameter Uncertainty

- Let’s say we wanted to use a normal distribution to represent a process time in our model (it’s very unlikely you’d use this distribution in reality).
- The normal distribution is defined by two parameters – the mean and the standard deviation.
- The mean and standard deviation that we use to define the distribution is derived from a <u>sample</u> of data from the real system.

Problem: estimates from samples contain uncertainty

- Our sample mean might differ from the true population mean because of chance;
- Or because the sampling process is unreliable/biased

(eg if we repeatedly resampled treatment time data from the real world the mean would differ slightly each time)

How do modellers handle parameter uncertainty?

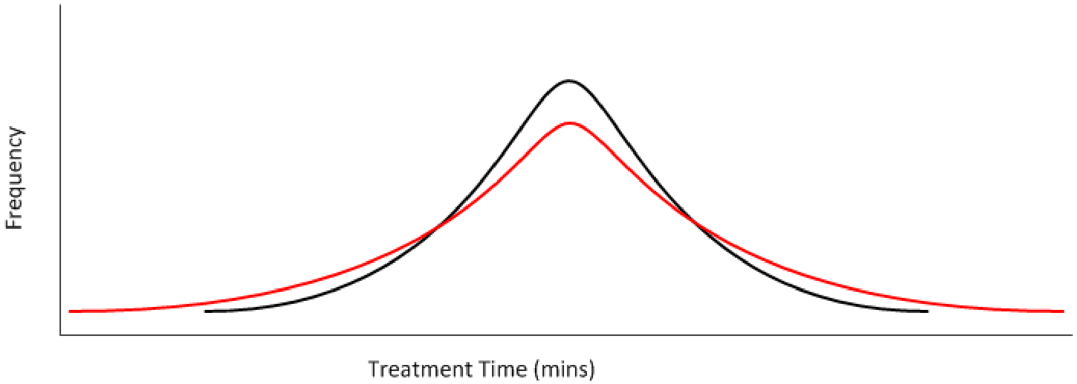
- Take a bigger sample?
  - Not always possible!
  - It helps, but doesn’t eliminate all uncertainty
  - If sampling process is biased then this will not help
  - Longitudinal issues (the process may have recently changed)

- Estimate the uncertainty in sample statistics
  - There are some statistical approaches that can help

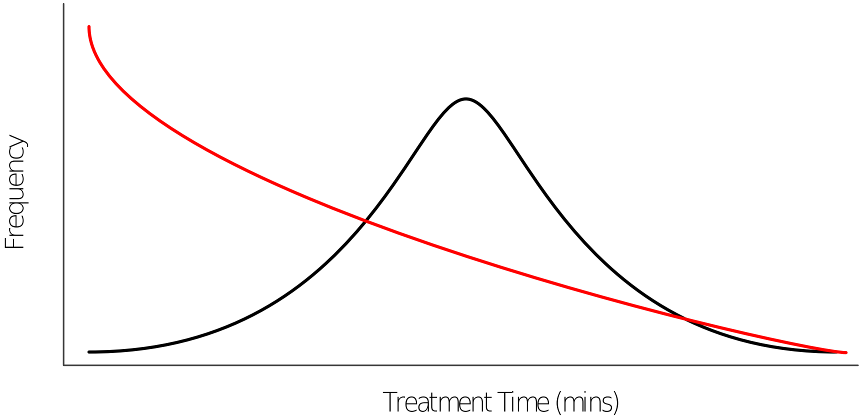
- Analyse if the results are sensitive to a parameter
  - Make small to medium size changes in a parameter
  - Look for large differences in results
  - This is known as **Sensitivity Analysis**

Let’s take a look at an example using the ED model we saw at the very start of the session…

Other types of parameter uncertainty



The normal distribution takes two parameters: mean and std dev. In this case the sample std. dev is an underestimate of the population std. dev (lower spread of process times)



The type of distribution used is also a parameter. Here the two distributions have the same mean, but the process times follow a different distribution