PenARC

Health Service Modelling Associates Programme

Introduction to Operational Research and Data Science

Dr Daniel Chalk

Senior Research Fellow

(NIHR Applied Research Collaboration South West Peninsula)

Introductions

What is Operational Research?

Let’s demonstrate using a simple exercise.

What is Operational Research?

Imagine you are in charge of an Emergency Department.

You need to cut costs, and quickly. You have come up with two options :

1) Get rid of a Senior Nurse
2) Get rid of a Treatment Cubicle

Let’s assume the costs of each are the same.

What is Operational Research?

In your ED, patients arrive, are registered at the registration desk, are triaged by a Senior Nurse, are treated in a Treatment Cubicle, and then are either admitted or discharged.

What is Operational Research?

I’m going to split you into small groups. You will have 3 minutes to discuss and come up with answers for these three questions :

1) How long (on average) do you think a patient spends in the ED now? (Imagine an average taken over 1 year)
2) What cost-saving decision are you going to make (get rid of a nurse or a cubicle?), so as to minimise the impact on the time patients spend in the ED?
3) What do you predict is going to be the impact of your decision on the average time patients spend in the ED? (ie how much longer do you think patients will spend in the ED as a result of your decision?)

1) How long (on average) do you think a patient spends in the ED now? (Imagine an average over 1 year)
2) What cost-saving decision are you going to make (get rid of a nurse or a cubicle?), so as to minimise the impact on the time patients spend in the ED?
3) What do you predict is going to be the impact of your decision on the average time patients spend in the ED? (ie how much longer do you think patients will spend in the ED as a result of your decision?)

- You currently have 2 Senior Nurses
- You currently have 4 Treatment Cubicles
- On **average** :
  - A patient arrives at the ED every 8 minutes
  - It takes them 2 minutes to be registered once they are seen at the desk
  - It takes them 5 minutes to be triaged once they are seen by a Senior Nurse
  - It takes them 30 minutes to be treated once a cubicle is free

What is Operational Research?

Let’s hear your answers.

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 cost-saving decision are you going to make?

“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) What cost-saving decision are you going to make?
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

* assuming the model has been built objectively!

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

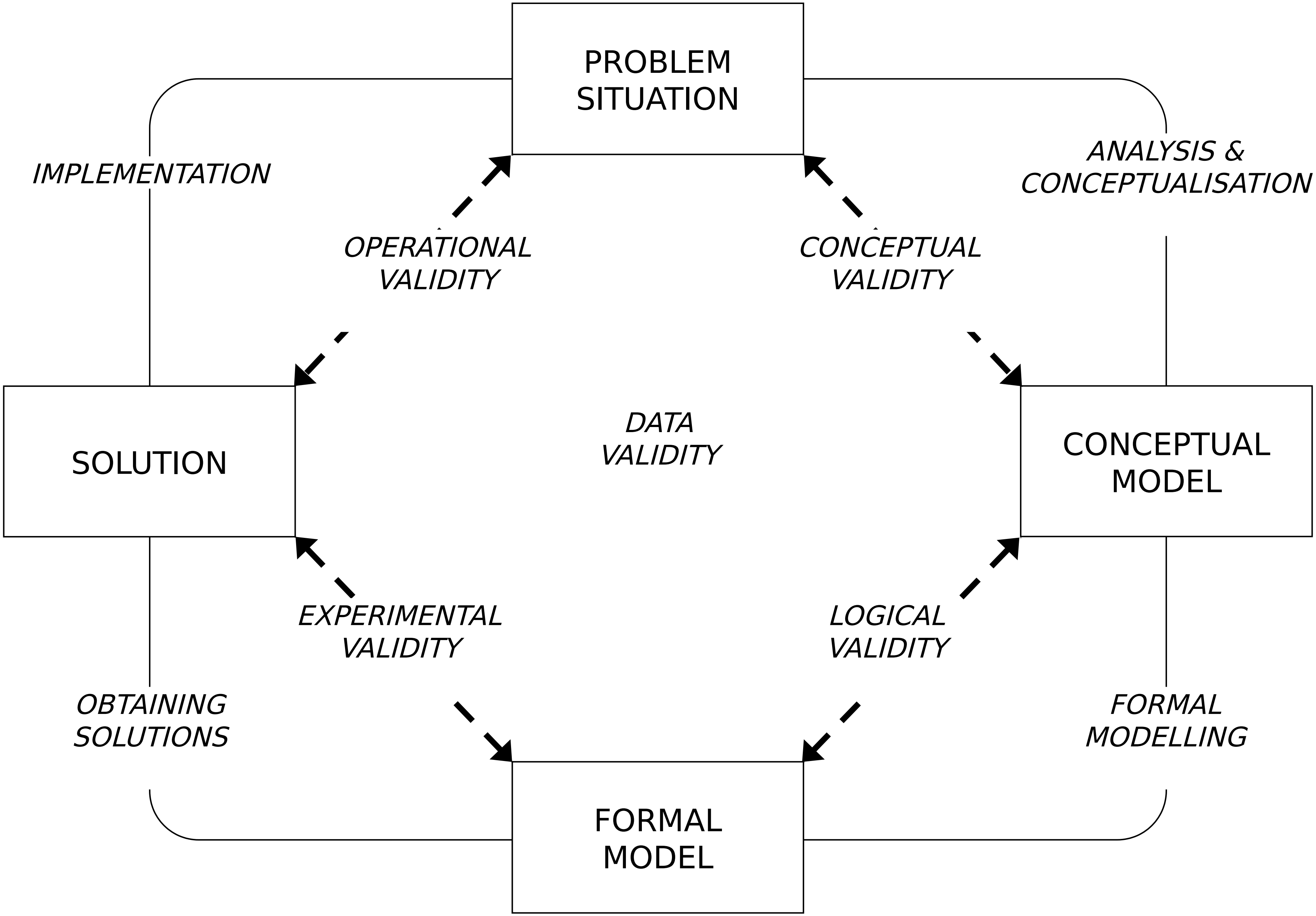
Modelling Techniques

Simulation
- Discrete Event Simulation
- Agent-Based Simulation
- System Dynamics (Whole Systems Modelling)
- Monte Carlo Simulation

Optimisation and Mathematical Approaches
- Mathematical Programming
- Forecasting
- Scheduling
- Location Analysis (Geographic Modelling)
- Markov Modelling
- Queuing Theory

Problem Structuring and Conceptual Modelling ("Soft OR")
- Validation
- Soft Systems Methodology
- SODA
- Strategic Choice Analysis
- Live Simulation

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?

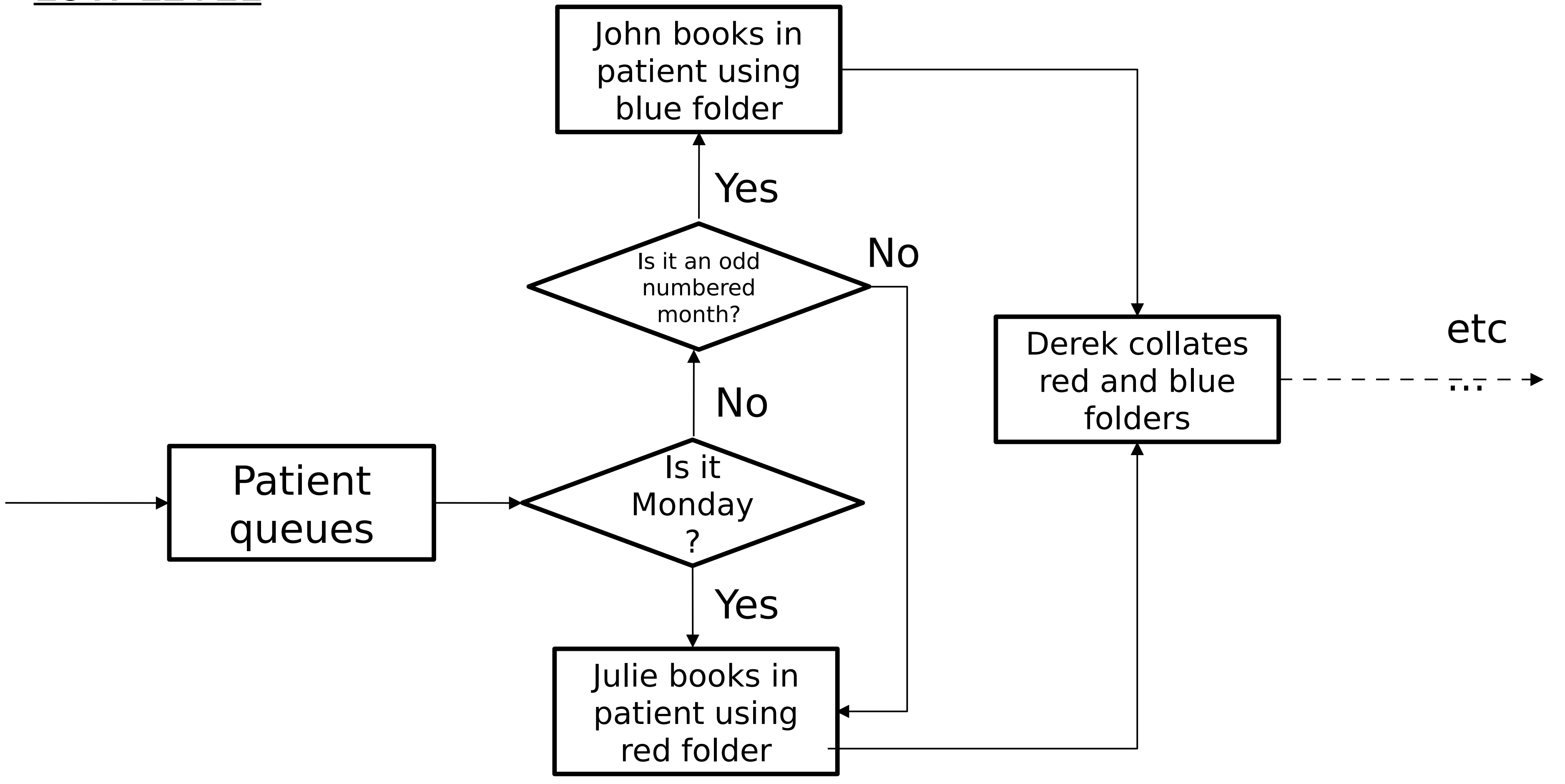
What’s the minimum you can model to answer your question?

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?

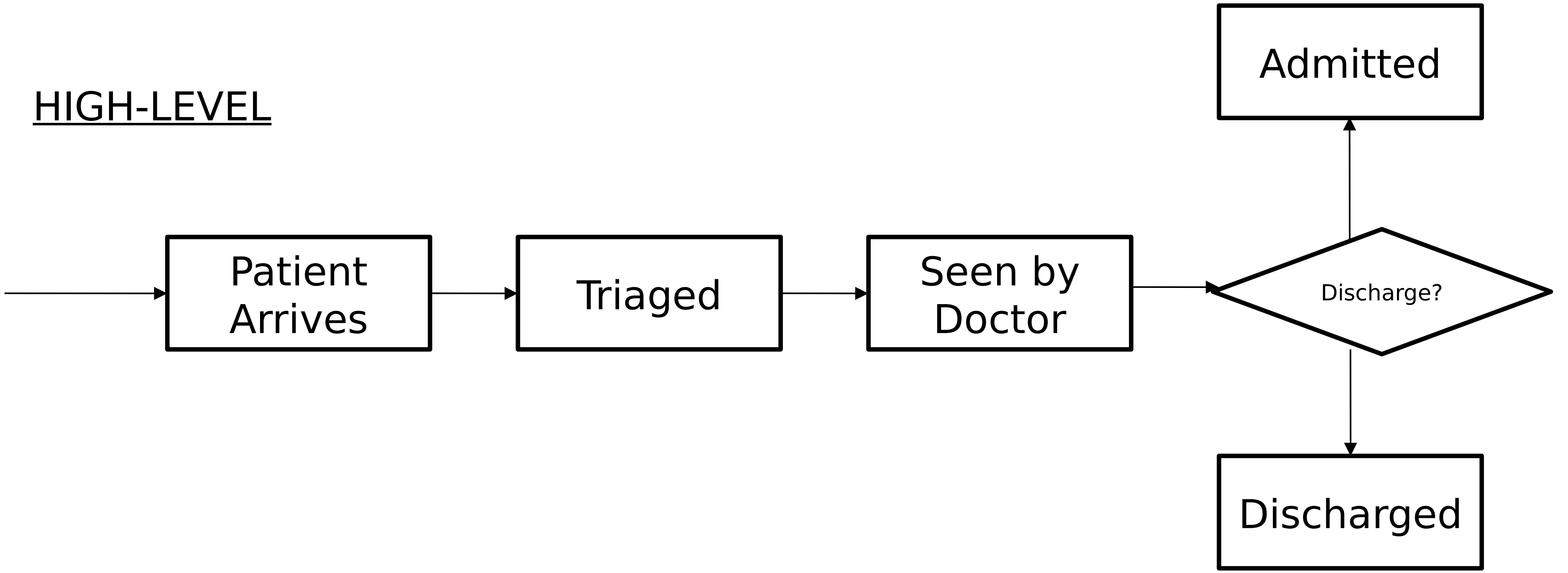
<u>LOW-LEVEL</u>



Level of Detail

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

HIGH-LEVEL



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….

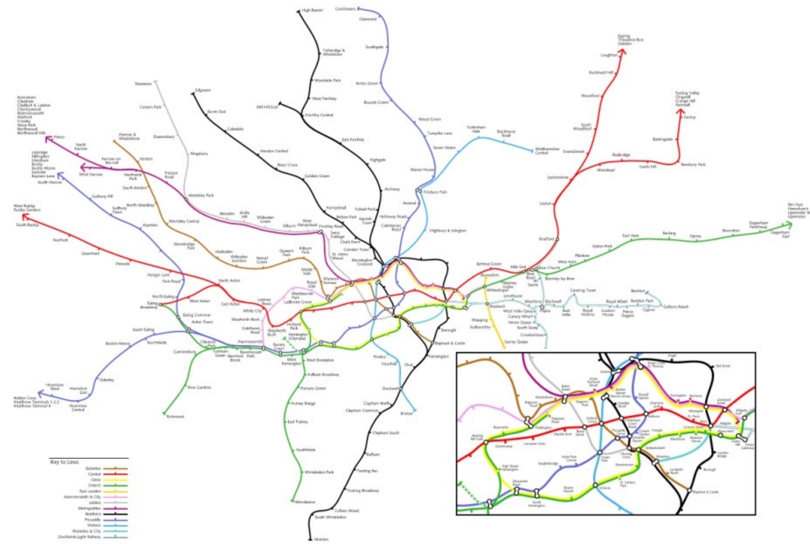
- … because they are a simplification of reality
  - … they miss out and ignore bits of the real world
- In the natural and physical sciences some models are extremely accurate predictors even though they are simplifications;
- In other areas such as health care operations a model may not be able to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

but some are still useful….

- … because they are a simplification of reality

Tube map

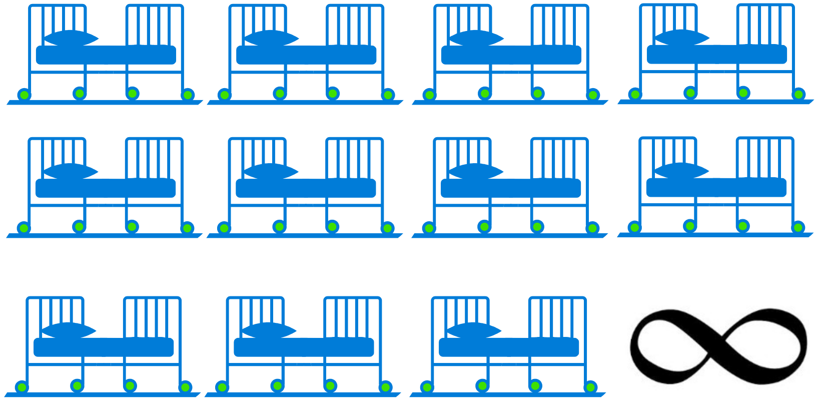
Actual layout of the underground



This model is wrong…

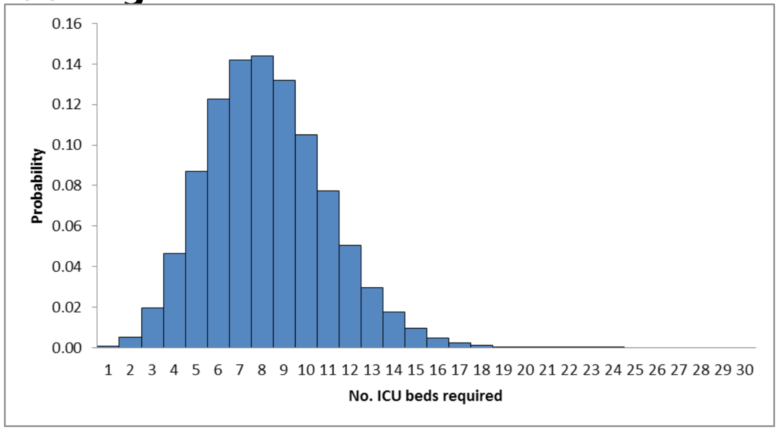
but very useful

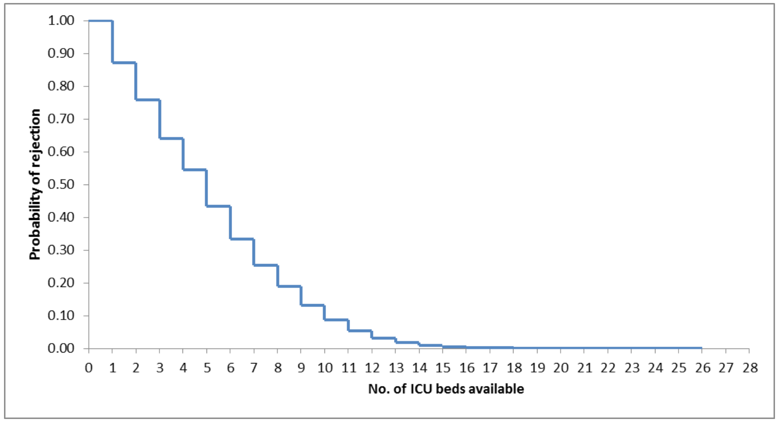
A model of a cardiac ICU with infinite capacity



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 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 vital tool for understanding how your system works.

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

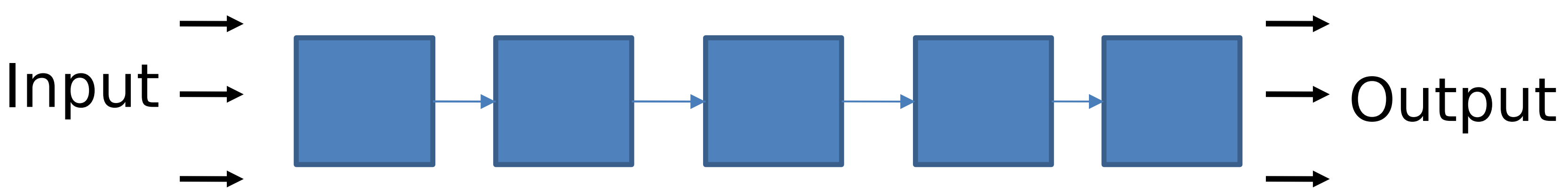
What is a Process?

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

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

Output = function (inputs)

$$\mathbf{Y = f(x)}$$



The “Secret” of Process Mapping

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

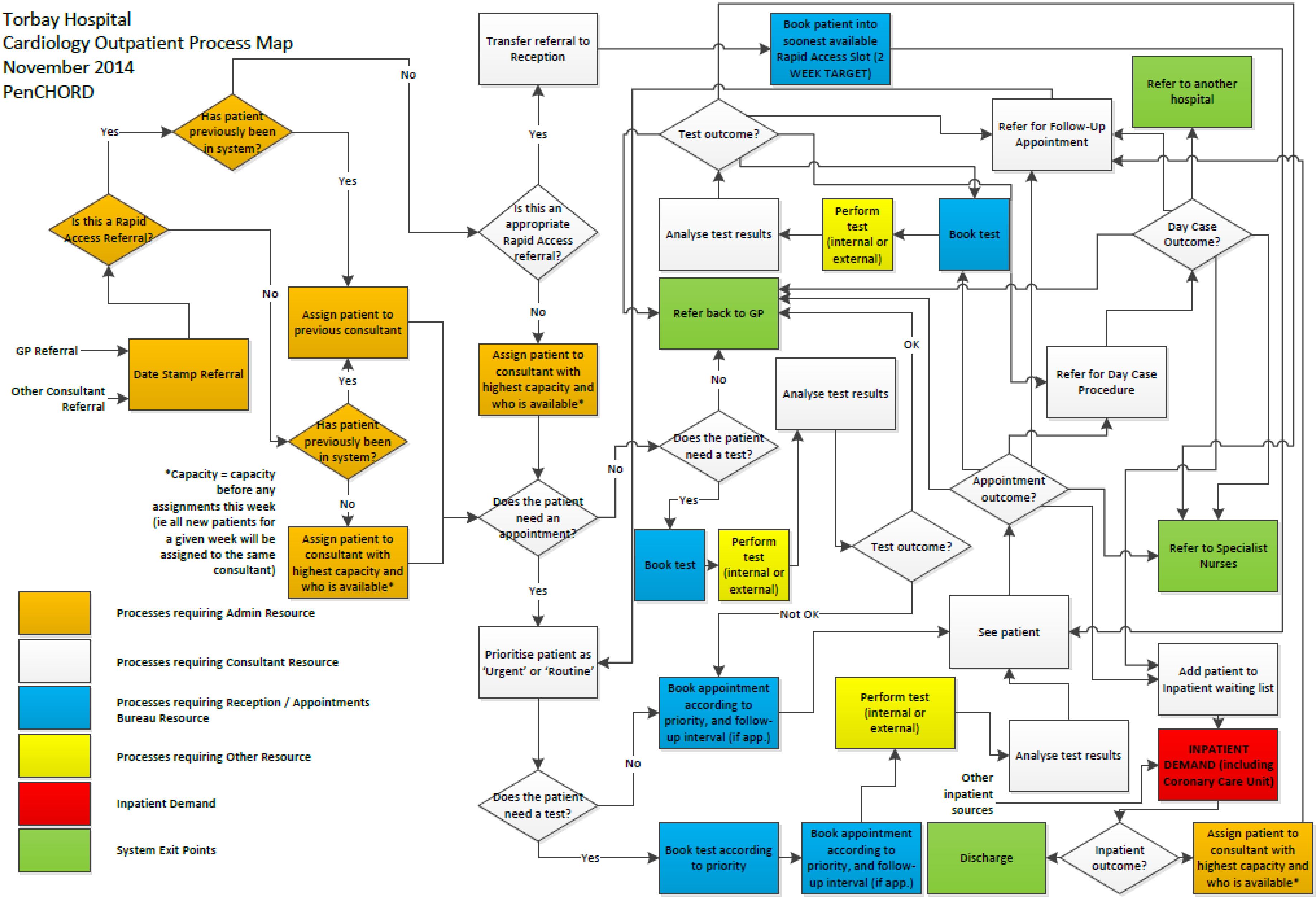
The “Secret” of Process Mapping

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

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 (detainees, officers etc) – how do they flow between processes in your system?
- Create a visual representation of your system

Example Process Maps

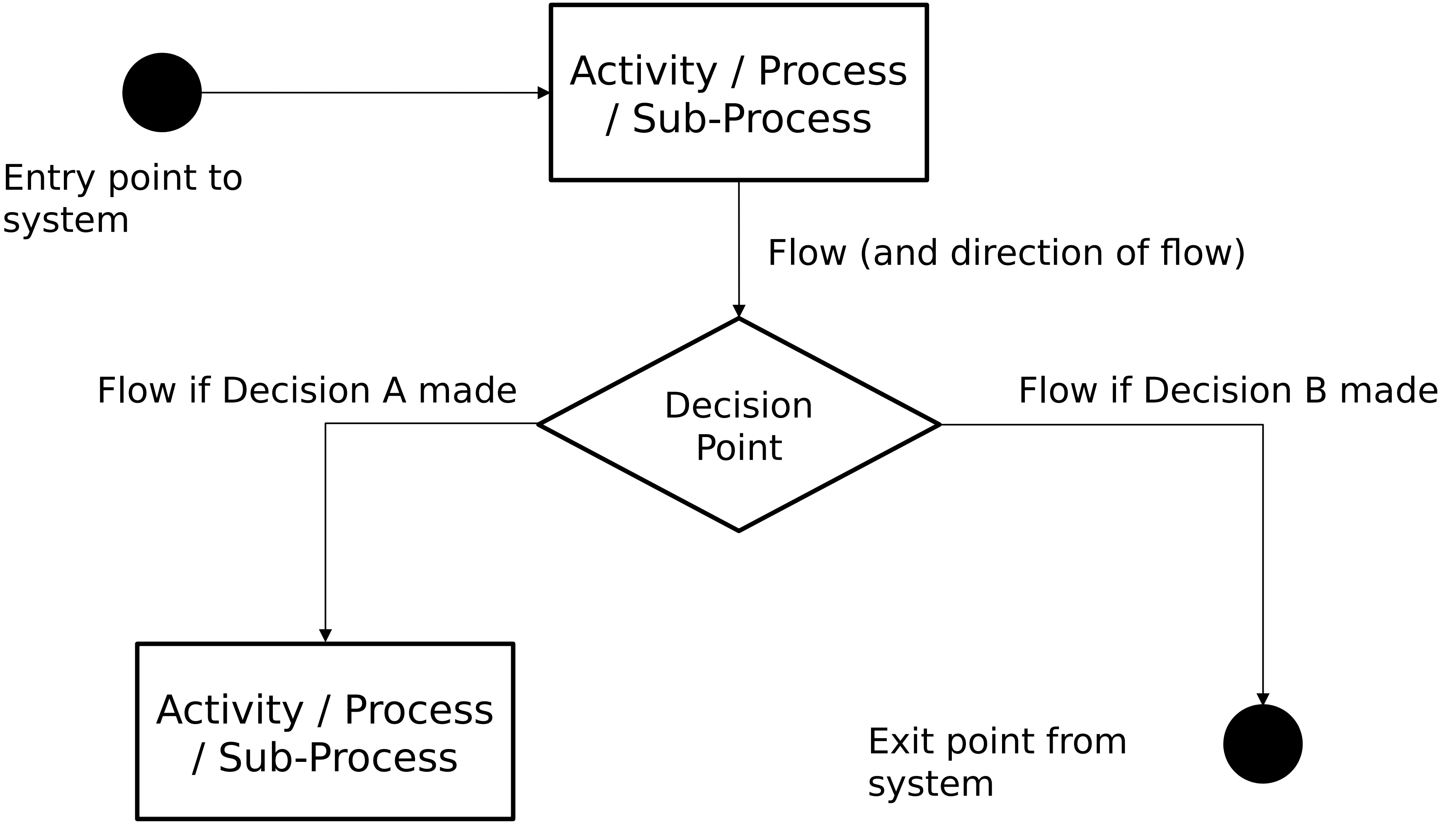


Format of Process Maps

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 :

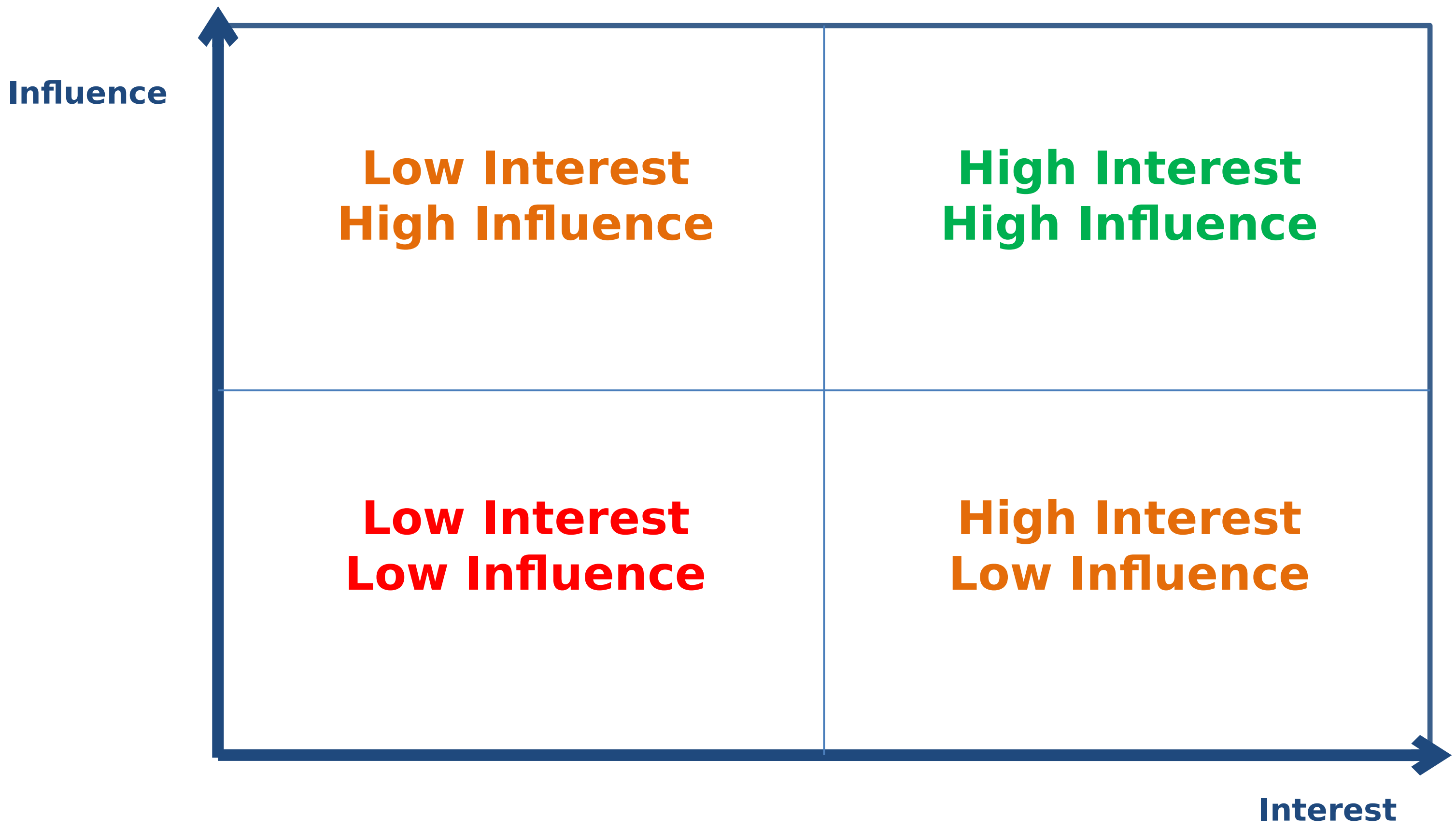
1) Who owns the problem?
2) Who might benefit from its solution?
3) Who might suffer from its solution?
4) Who has a legitimate interest in the problem and / or its solution?
5) Who has the power to influence the process of solving the problem and implementing the solution?

Why Consider the Stakeholders?

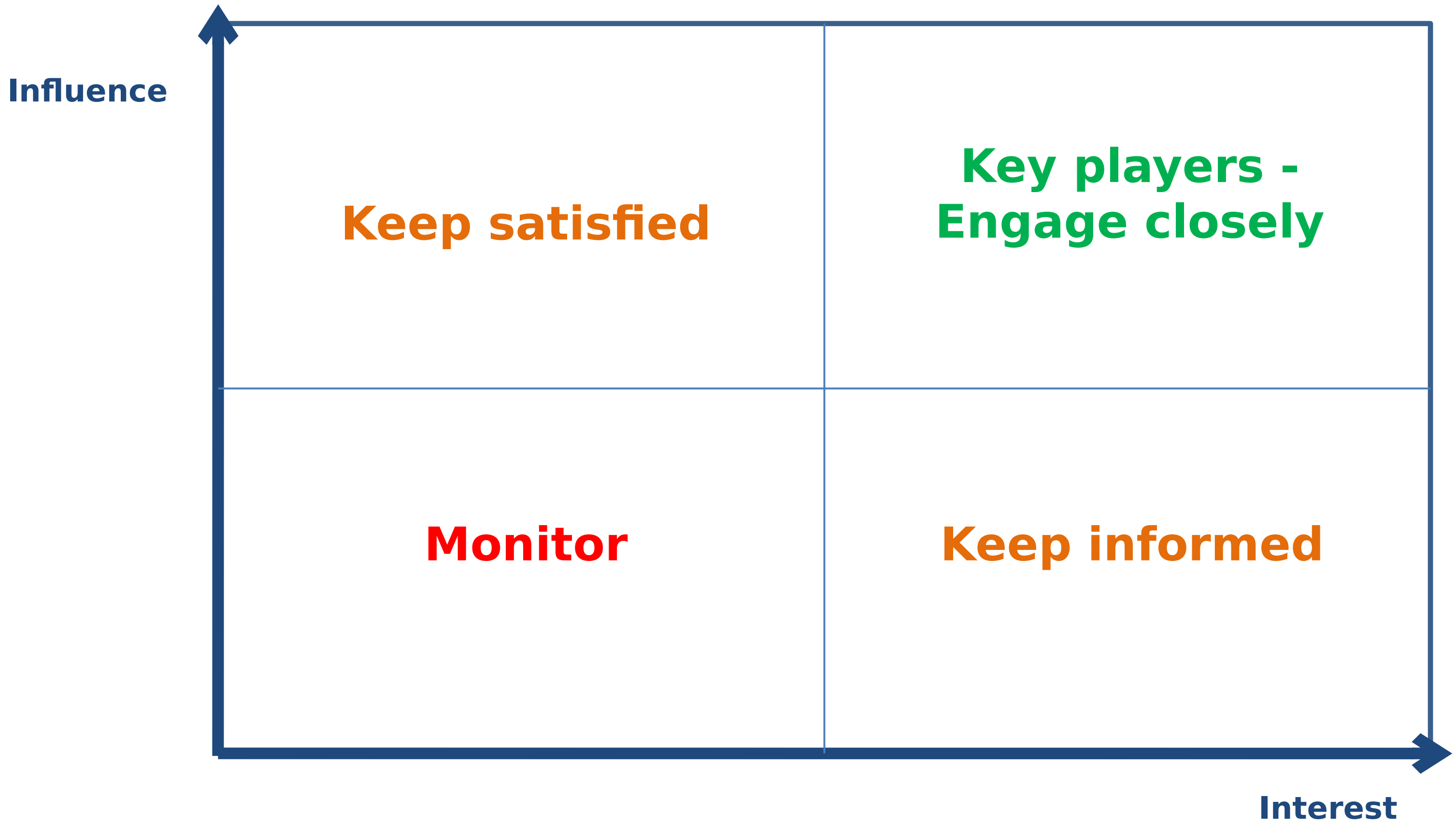
Why is it important to consider who are the stakeholders?

- 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

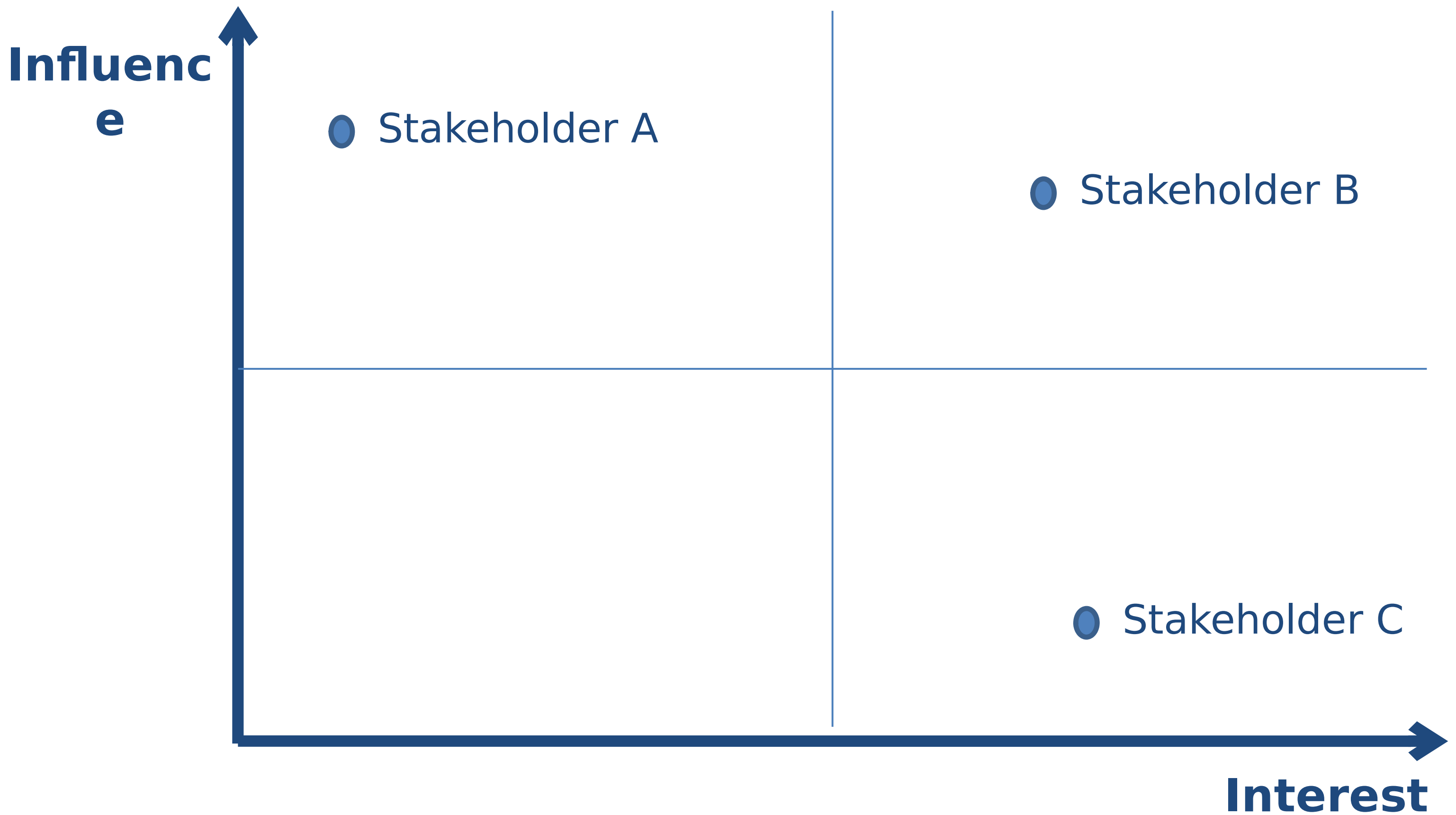
Interest / Influence Diagram



Interest / Influence Diagram



Interest / Influence Diagram



Exercise 1

You have been provided with a fictional problem. In your small groups, you will now spend the next 45 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.

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.

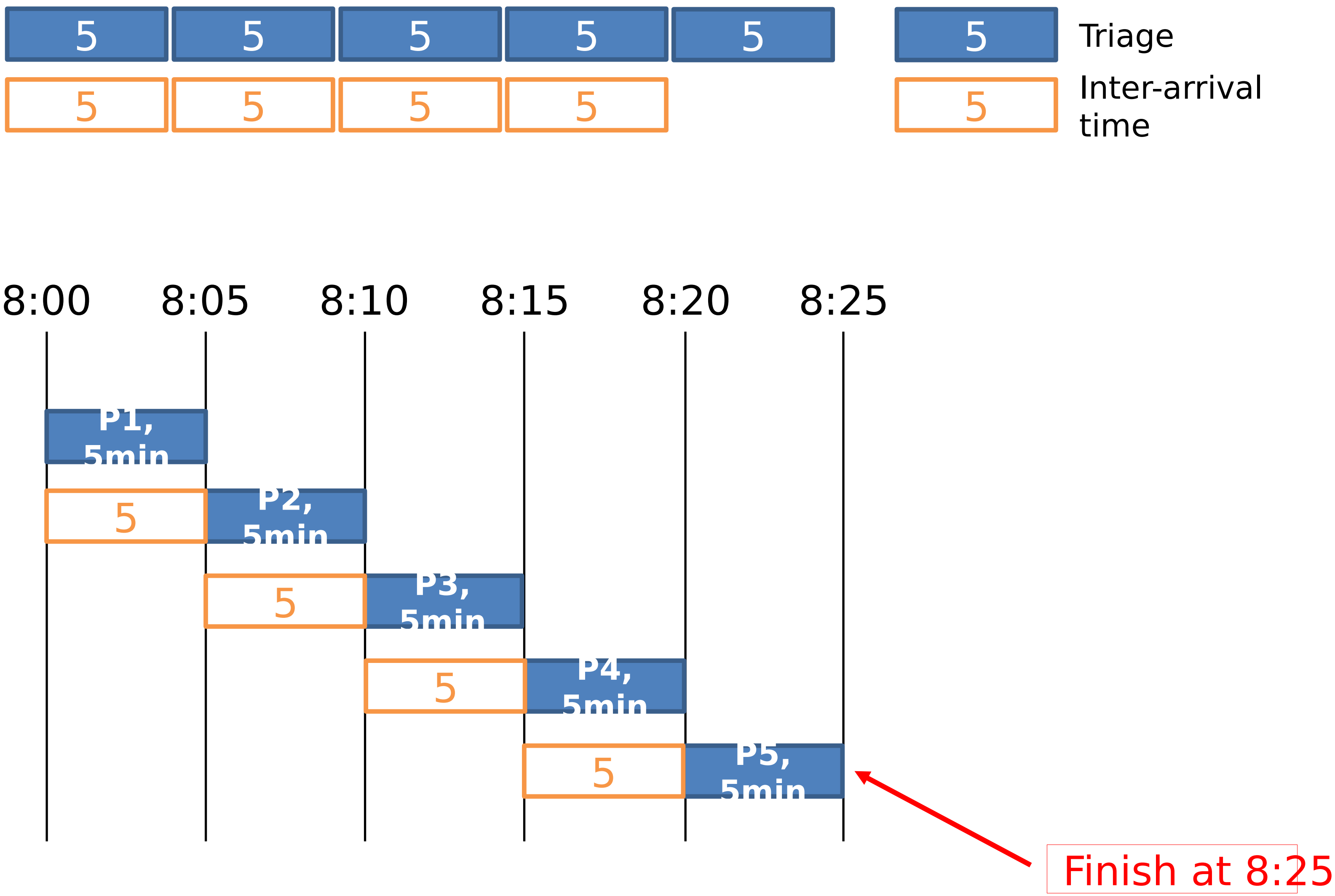
Scenario A

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

Scenario B

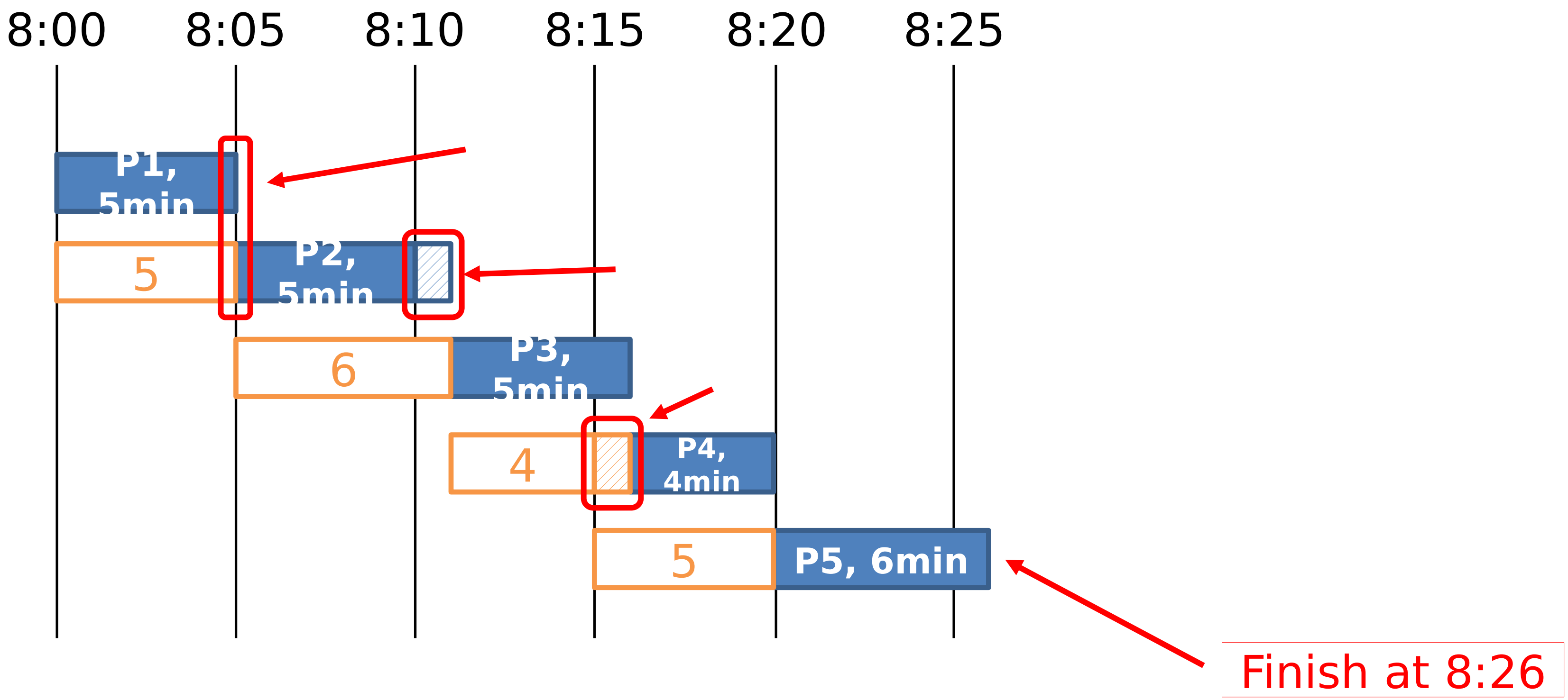
						Mean
Triage	5	5	5	4	6	5
Inter-arrival time	5	6	4	5		5

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



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

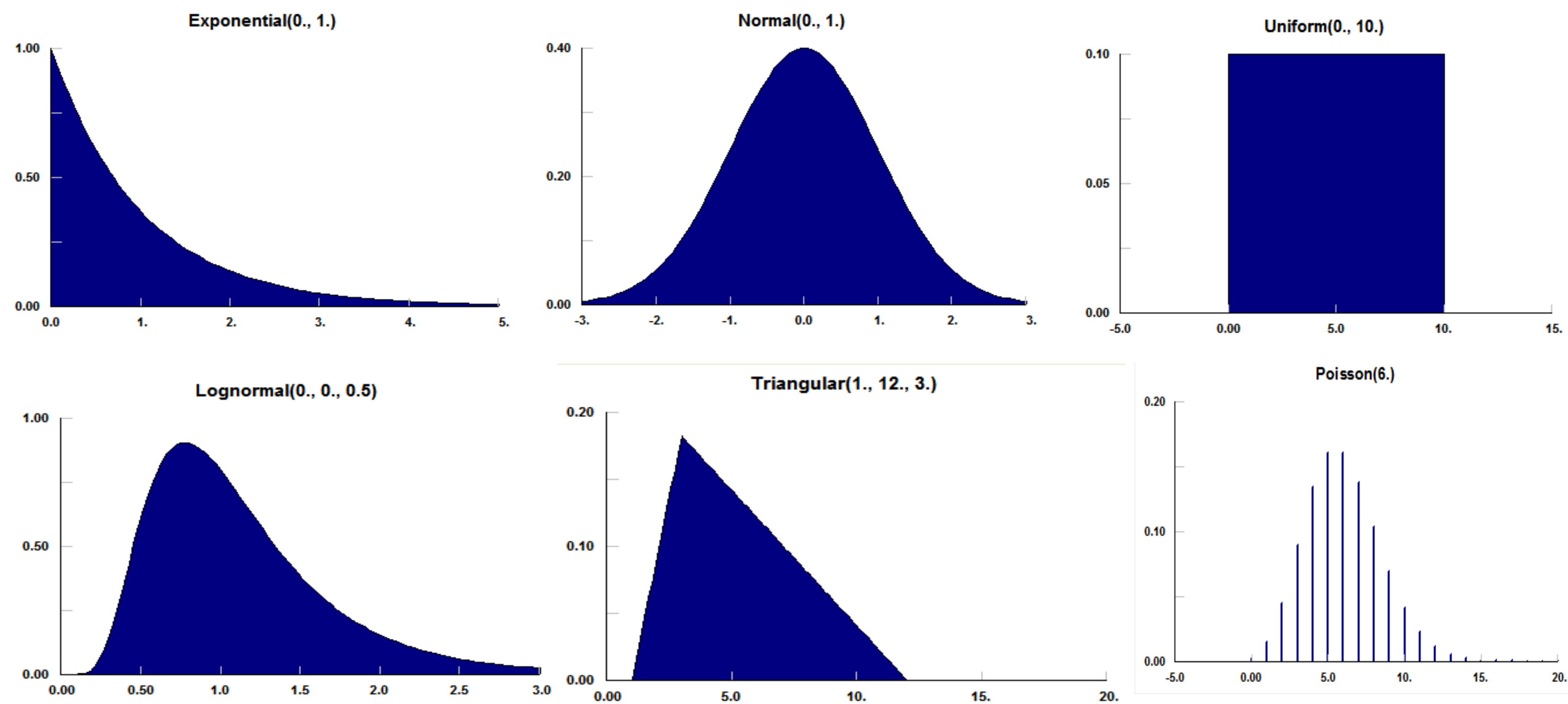
A distribution is a way of representing the variability within data. It provides us with an estimate of the probability of a value occurring in the future.

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



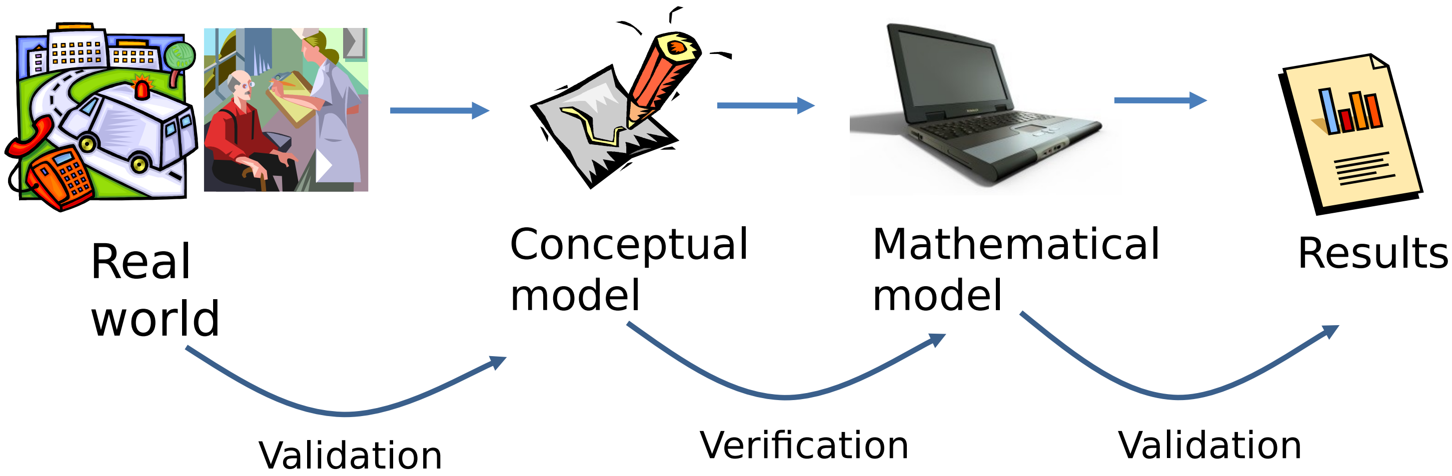
Exponential – common for inter-arrival times

Log Normal – common for process times

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)
  - Assess uncertainty

Confidence not validity

- It is not possible to prove that a model is valid
  - But you can <u>disprove</u> that it is valid

- View validation as a process of <u>building confidence </u>that the model and its results are sufficiently accurate for the <u>purpose it was built to address</u>

- 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 <u>invalid model may be viewed as credible</u> by users!

- Do not fall into the trap of thinking that a valid model is the most ‘realistic one’

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

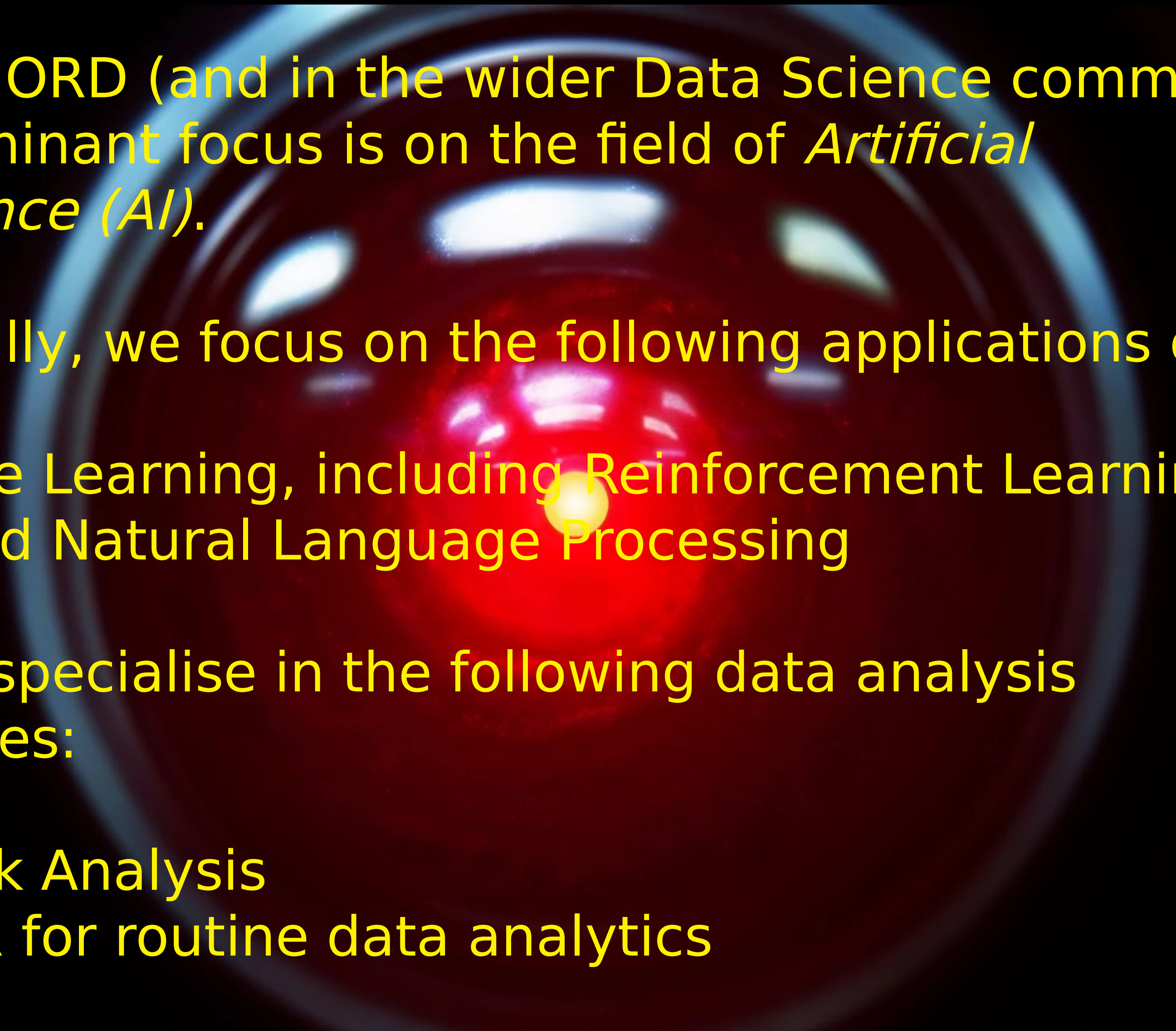
Example OR Question :

“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?”

Just what do you think you're doing Dave?



In PenCHORD (and in the wider Data Science community), a predominant focus is on the field of *Artificial Intelligence (AI)*.

Specifically, we focus on the following applications of AI:

- Machine Learning, including Reinforcement Learning
- AI-based Natural Language Processing

We also specialise in the following data analysis techniques:

- Network Analysis
- Using R for routine data analytics

We will teach you about all of these areas in this course, but here is a brief introduction.

Machine Learning

In Machine Learning, our aim is to get a machine to *learn* and *improve* from data.

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 patterns and 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.

Natural Language Processing

In Natural Language Processing (NLP), we are trying to extract information from *free text data* that contains natural language.

Commonly, we are trying to :
- Identify the *sentiment* of written text
- Extract *named entities* from text (people, places, groups etc)
- Identify relationships between entities in text data
- Identify the common themes / topics in text data

Lorem ipsum dignis eium fugit aspelluptat eati odis net exeri rectem lia venihil icipsapid qui dolupient quam aceatemque repedi tem lantiae provid quia sitatia temquibuste voluptatquo comnit fugiat invenit fugia quo ditias ipitatat erspel id utesecuptur solorio. Hari veria dis nis et millic tota comnimet inctor sum aut laboressedis deribus illore non et fugitiosam, soluptat in eatur? Endignatem el et ex endicim re occullu ptatem laut audipita num fugit adis delenti cum sus aut iduntur arumqui blam eos molorum quissimint, nis ut aut adisci odic te as everspe ditati accum alit rempel iumque nobis repudamus.

Epro que pra nis atempor ad quis am sim explitem sum et, illaute mposape lestiores rehenis ante senducidus quatet doluptatur sumquia ntiore, et idus volutem voloreria volorem fugit ento blabore pro et, od quam sam re volori aut etur seque velignia voloreic tem es nosam volo magnam que ad que voles et voloribeatem rati ipsae dolentiatur rempore sequunt, quis ea volupti am secuptaquate cus etum eliqui dolorep erferrum diore quibus re omnisci aditibusda qui aut alit, ide et possequias nonsed est int eos il earciis sit aliquidis eum quiam, quam doluptatis vel eum nos pos dolorat emoditiur? Quid que pe veratiae pero molum volorehendit re, unt pa doluptio cus.

Ga. Nam, ommodic aborrorepe plabo. Neque quaestiat hiciae opta inis dolestrum cum, quam ut fugiam, od mosamus aruptaspel modi rerunt paris a volore pressed ut voluptatius voluptae provid quis volo eumquunte exerumq uaernatia serum cus estis autemquis exeris et ut omni consequis essi to qui ne destoreri sinctes placcae aut labo. Ritia dolestiatur autet omnitatus aliciae repudam, sanditatiis doluptatur? Qui apid ut officabo. Et essi nem et debis re dolorio ssinctatem quodipsuntem explique doluptaqui officidempor ad ut volut la doluptam re dolute eatur sequam exerspero explacc ulliquiatio vendustrum etus sed que desseque sum quos voluptatiame core cuptaquibus exerrum et molupta ssiment ressundantem idiciet omnimus re ni berunt prero veliquis non cus.

Aliquibusdae con prernatur aut res dolora aut aut quam comni tem sequo ommoditatet hiliqua tusciae pedit, atures doloritius dolutat emperis et imus dolupta tentibe rspienis niminus ni videbitatem quos re, temquis untibus undaeribus eum faccus, tem erum nem. Entorum illectus del eatur?

Faccuptatur, se nostrum et fugia peria nossunt fuga. Sedicab orectissitae voles evel ipsus ipsae et, sam vit, conesequе sit et aspe quo quas se quiandusdae solorumquae vel enimus verumendunt.

Ab into dolora nossequi tem num quunt maxim del intio. Harumenda ne perovit quiant ut alique doluptatia ium sunt volut magnatiis voluptis es enim non earuptae eos doluptatur solorer iorem. Ur rem que dellores nimpos pra sequasped qui quiatqui conem reiunt.

As debit eatiber itatur, aut lat parchilicae pa pellentum ut ident eatia commolo eum nectibusa apel istruntiore idebit eturiatur?

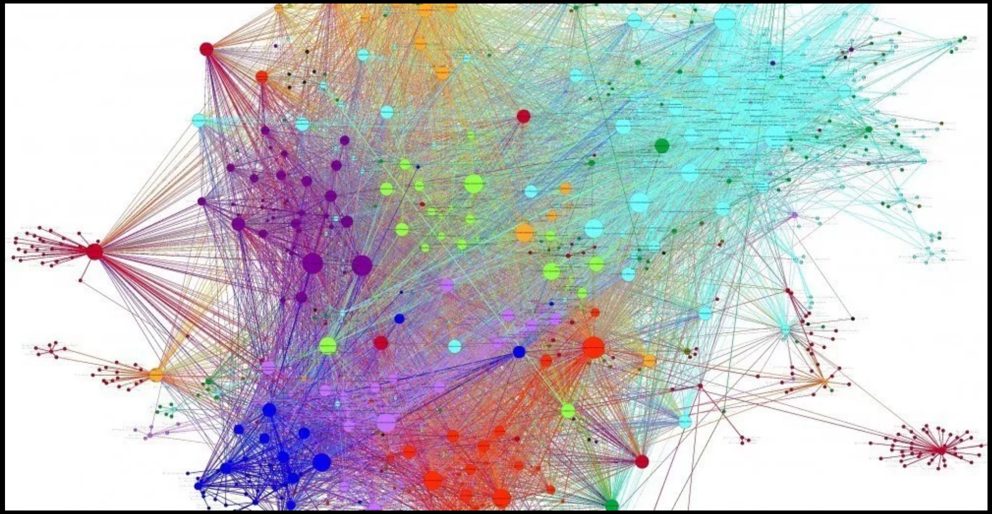
Lupieniendis aut volorerio. Daectot aectestium latem volenimus ut velis alibus ulliciis aceaquam derovid elitatet que vel minctume iusam dolupti venis am fugit etus vellit re viducid quiatquiant volestisti torehen tiorestem antis militas nes del ilicaturibus et et est harum, ipsant, natem quos es ipsa velit est, es re volupta temolorum este explant.

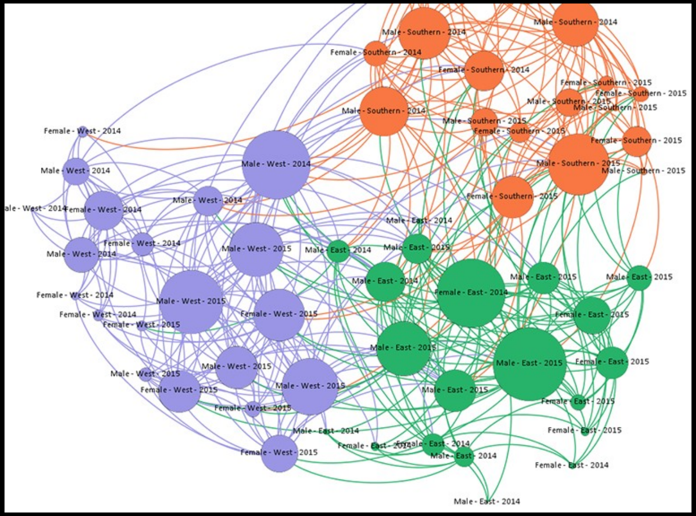
Pore vid est, audam facia voluptiae pos ut que nulloria core nihilita istio tem quiscia volore nulliae corpos eatur

Network Analysis

Network Analysis attempts to model and visualise the *relationships* between entities in a system or process.

The approach can help to unpick complex systems to better understand how they work, and provide insight into how links / relationships affect individual and system behavior.





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 2

In small groups, discuss potential applications for Data Science in your organisations. What data do you have that might benefit from the insights that Data Science methods could generate?

Further Work

- Watch the Bonus Tutorial “Sensitivity Analysis” in the Bonus_Tutorials Playlist of the HSMA Channel

HSMA Channel : https://www.youtube.com/channel/UCCY9_Gxg6kM-xjk9vV0mzIQ/featured