Understanding & Improving Demand and Capacity

What are your **expectations** of the Demand and Capacity Management session?

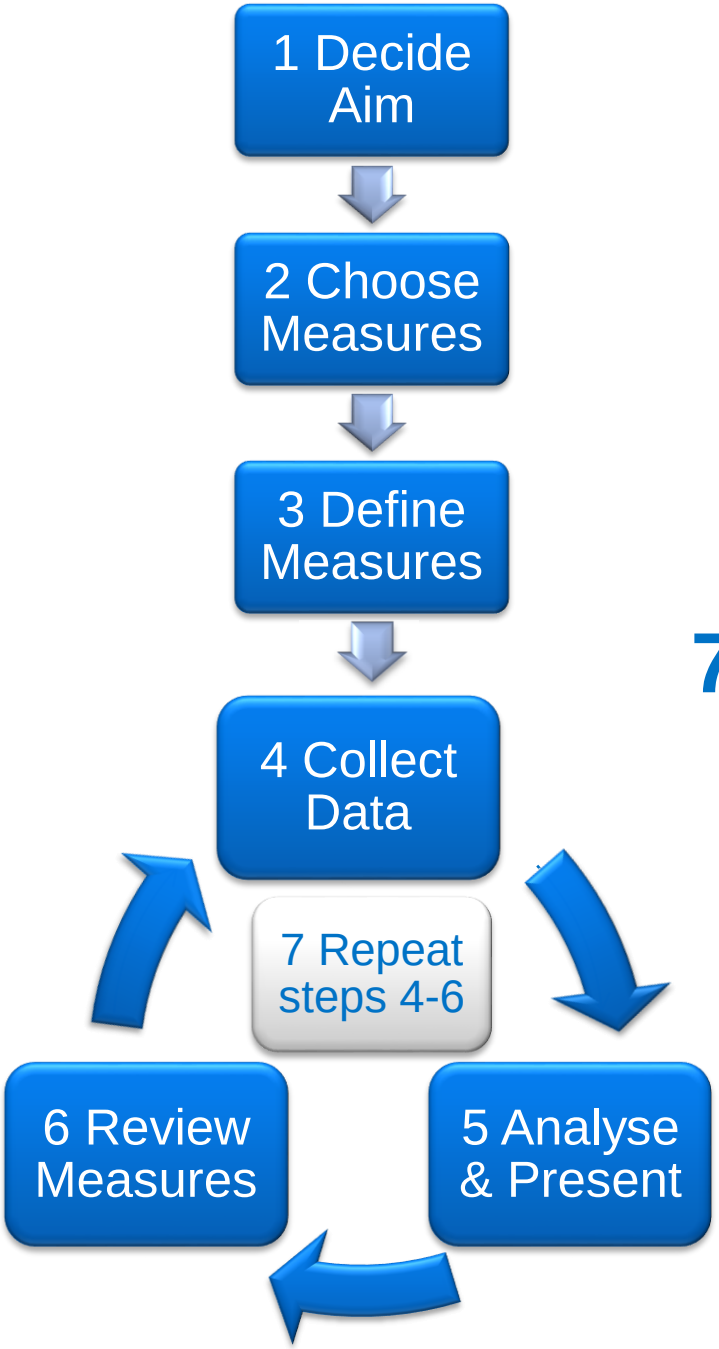
Aim and objectives

Aim

To have a greater knowledge and understanding of demand and capacity management

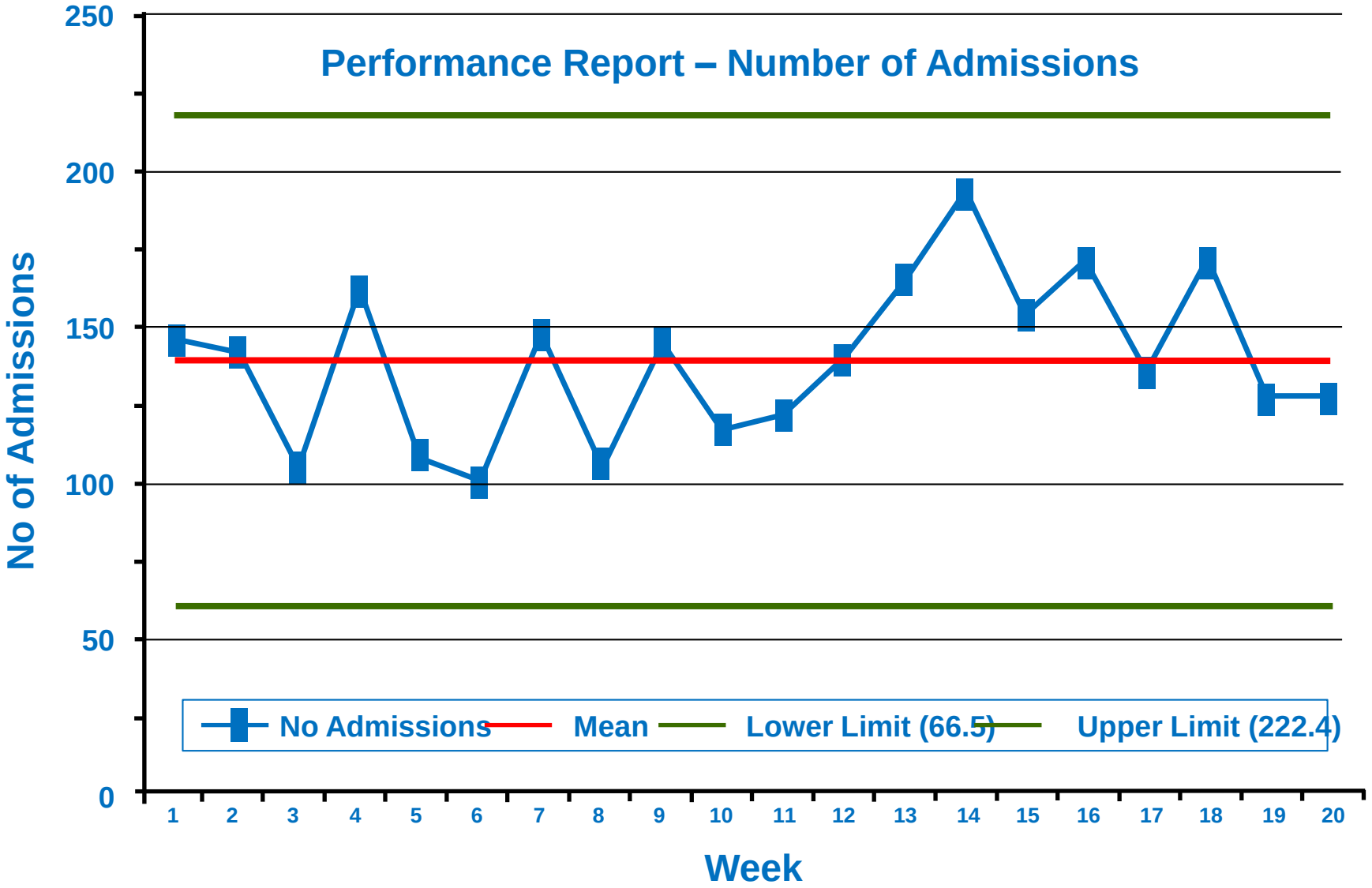
Objectives

1. Identify why managing demand and capacity is important
2. Learn about key definitions
3. Understand the common reactions to managing backlog
4. Identify what you could do differently
5. Apply the theory to a practical demand and capacity simulation exercise



7 Steps to measurement

Statistical Process Control (SPC) Charts:



Why is demand and capacity management important to the NHS?

Demand example

Six weeks later

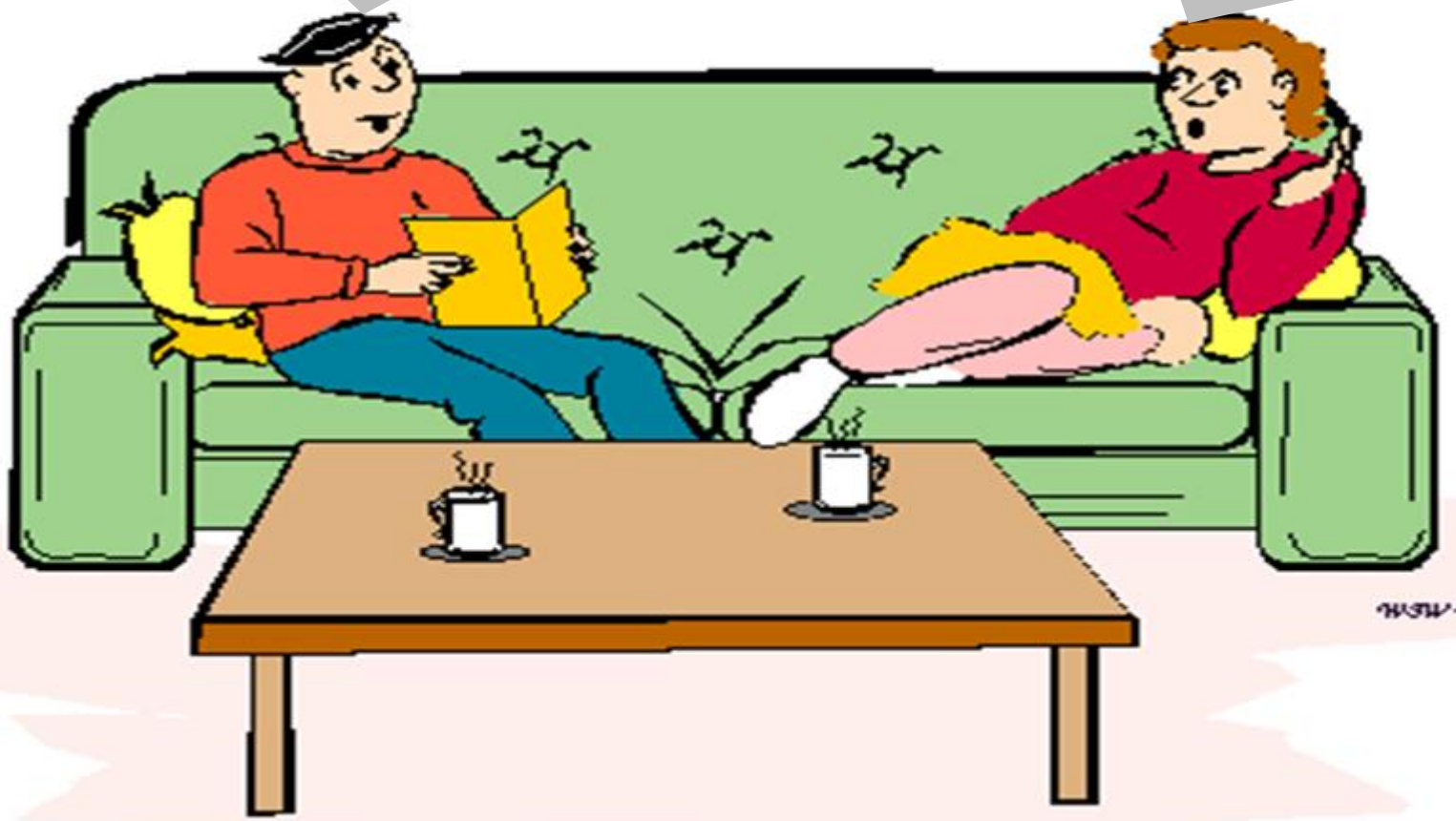
I haven't heard anything. I need to let my boss know when I'll be away. My wife's wondering if you ever did ring the travel company?

YES – of course I did!
They are a very busy company.
There's not a lot I can do to speed things up from this end

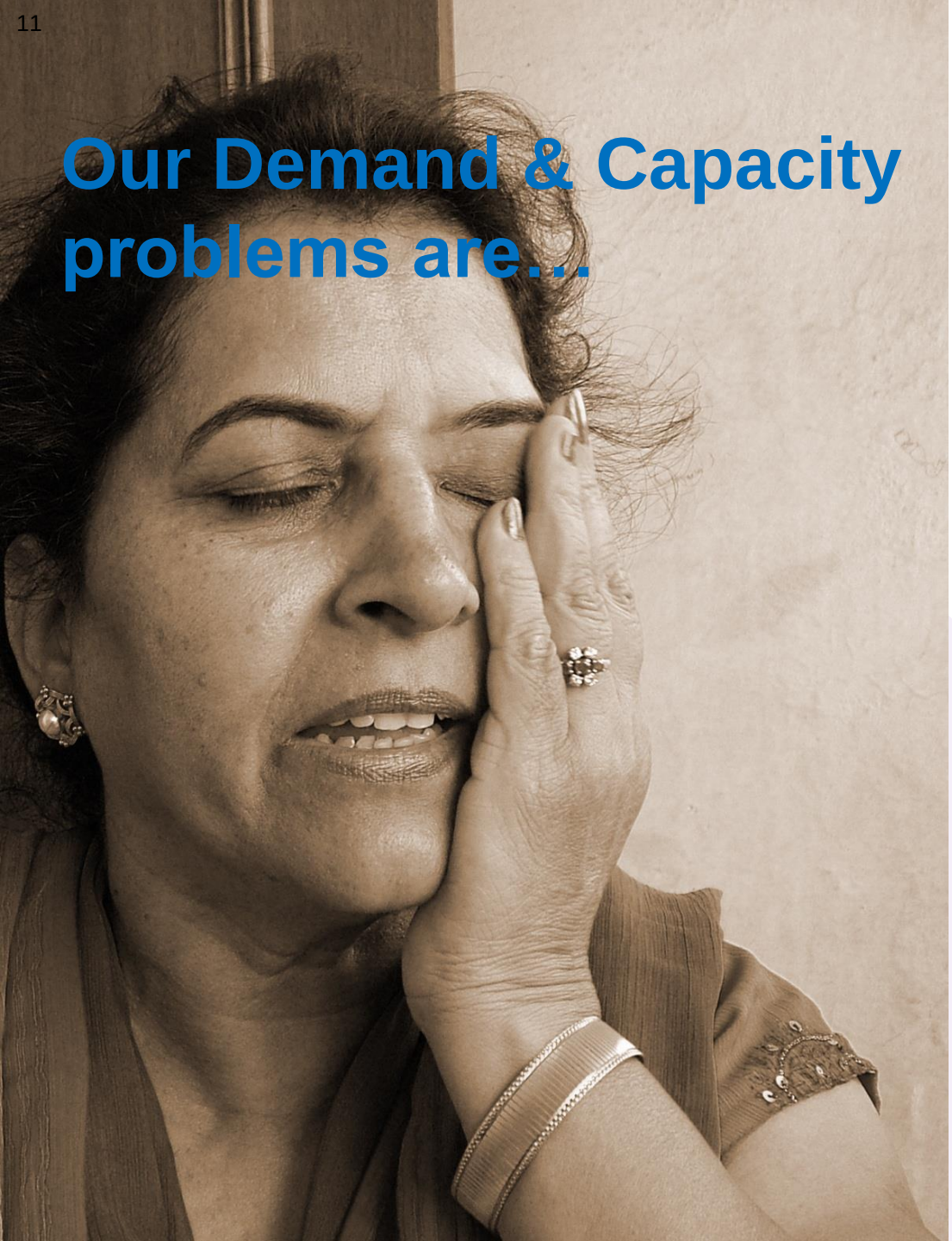
Two weeks after that…

Look what's arrived in the post today.
We fly to Majorca tomorrow at 6am!

What about the kids?!
What about the cat?!
Won't your boss go up the wall?!



At the airport



Our Demand & Capacity problems are…

…but for service users and patients?

Mindset that causes the problem?

We doubled resources….

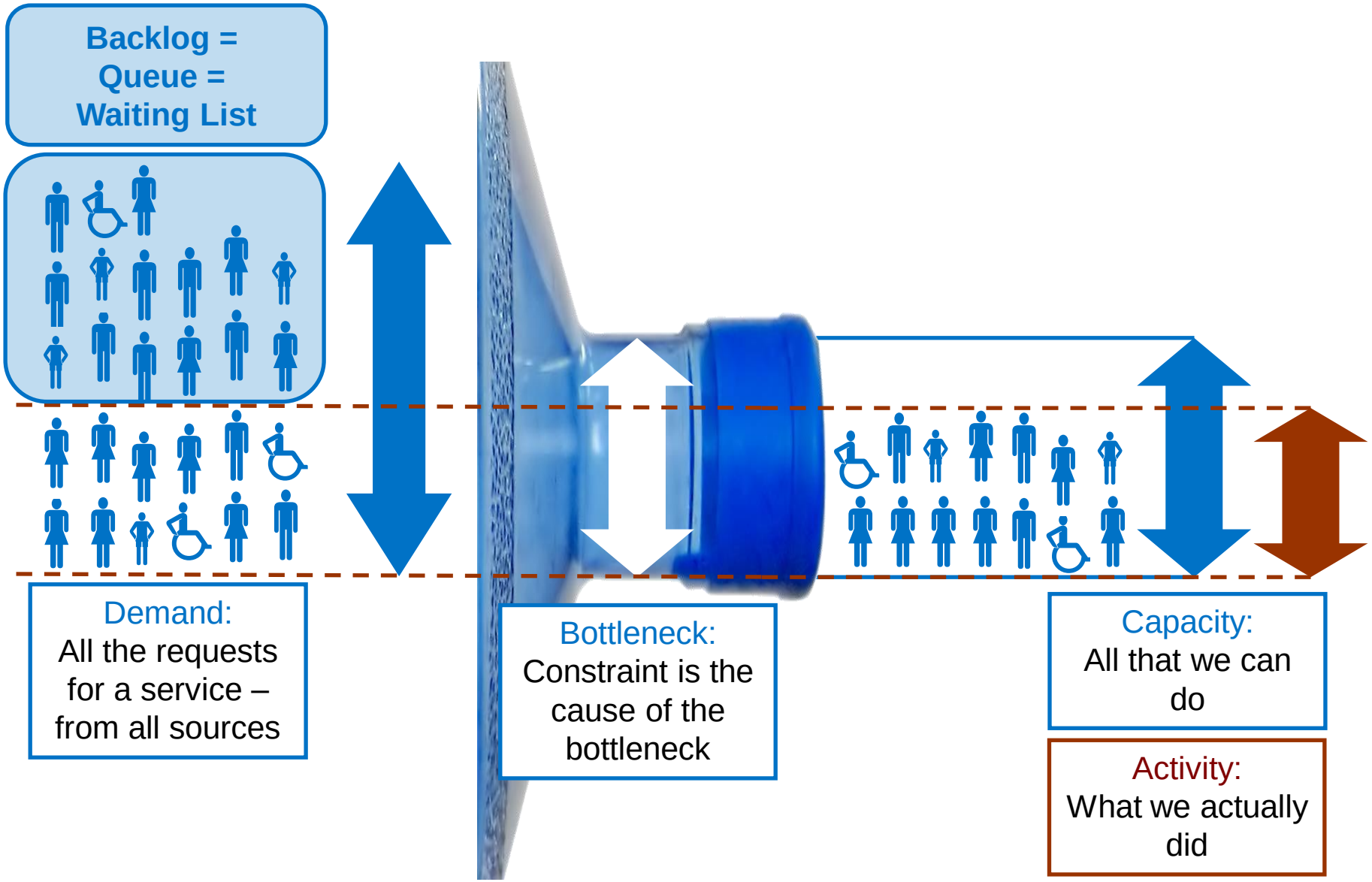
- £47 billion p.a. to £110 billion p.a. = 3% increase in activity
- 'We need value for money' but ……

….there is still a funding gap and patients are still waiting

Working definitions

Demand	what we *should* be doing
Capacity	what we *could* be doing
Activity	what we are *actually* doing
Backlog	what we *should* have done
Bottleneck	where the *queue* is seen
Constraint	what restricts capacity of the service at the *bottleneck*

How do they fit together?



How do we measure them?

Activity		
No. of patients* seen or processed	X	Time taken to process a patient
Backlog		
No. of patients in the queue	X	Time taken to process a patient
Demand		
All patients needing a service	X	Time taken to process a patient

* Or invoices or email or orders - etc etc etc

In the perfect NHS….?

We will collect perfect and complete data, for every patient , through every process…..

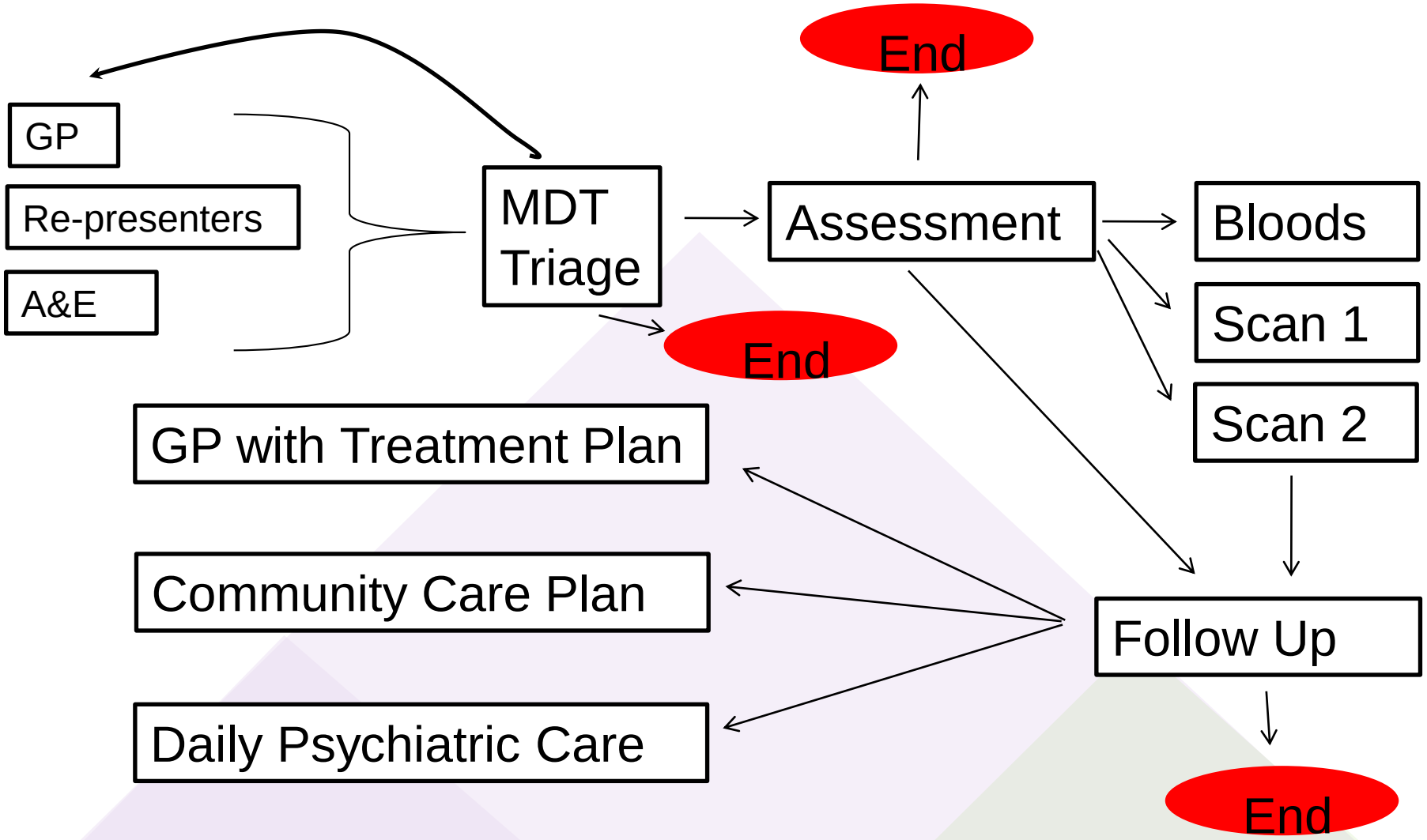
Would this be the best use of our time?

How do we *measure* them?

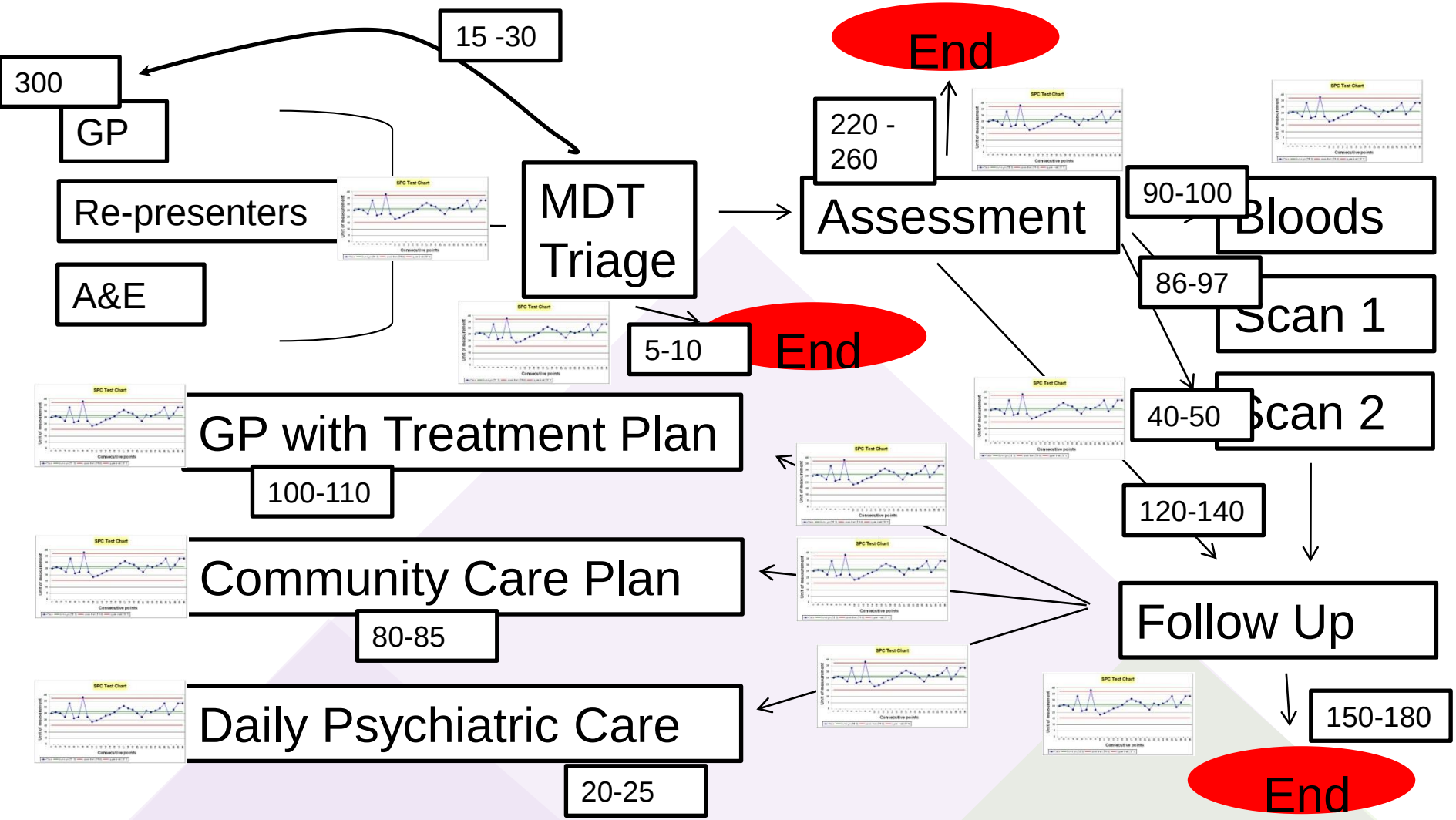
Stage 1 – Draw a process map

(Model developed by Matt Tite)

Psychiatric Process map of the system- Stage 1



Psychiatric Process map of the system Stage 2 – Add volume SPCs



Stage 3... Turn into Time

Although we now have the volumes at each stage (and they vary).

How long does it take to do each task? (this will vary too) This will depend on the patient and the person doing the task.

We must incorporate this in the calculations, not ignore it.

25 sets of data (shouldn't take long, as these are the tasks staff do repeatedly)

Remember A4 paper game!

Calculating Psychiatric Assessment times...

	Assessment time (mins)
1	**22**
2	**21**
3	**20**
4	**19**
5	**19**
6	**18**
7	**18**
8	**18**
9	**18**
10	**18**
11	**17**
12	**17**
13	**17**
14	**17**
15	**17**

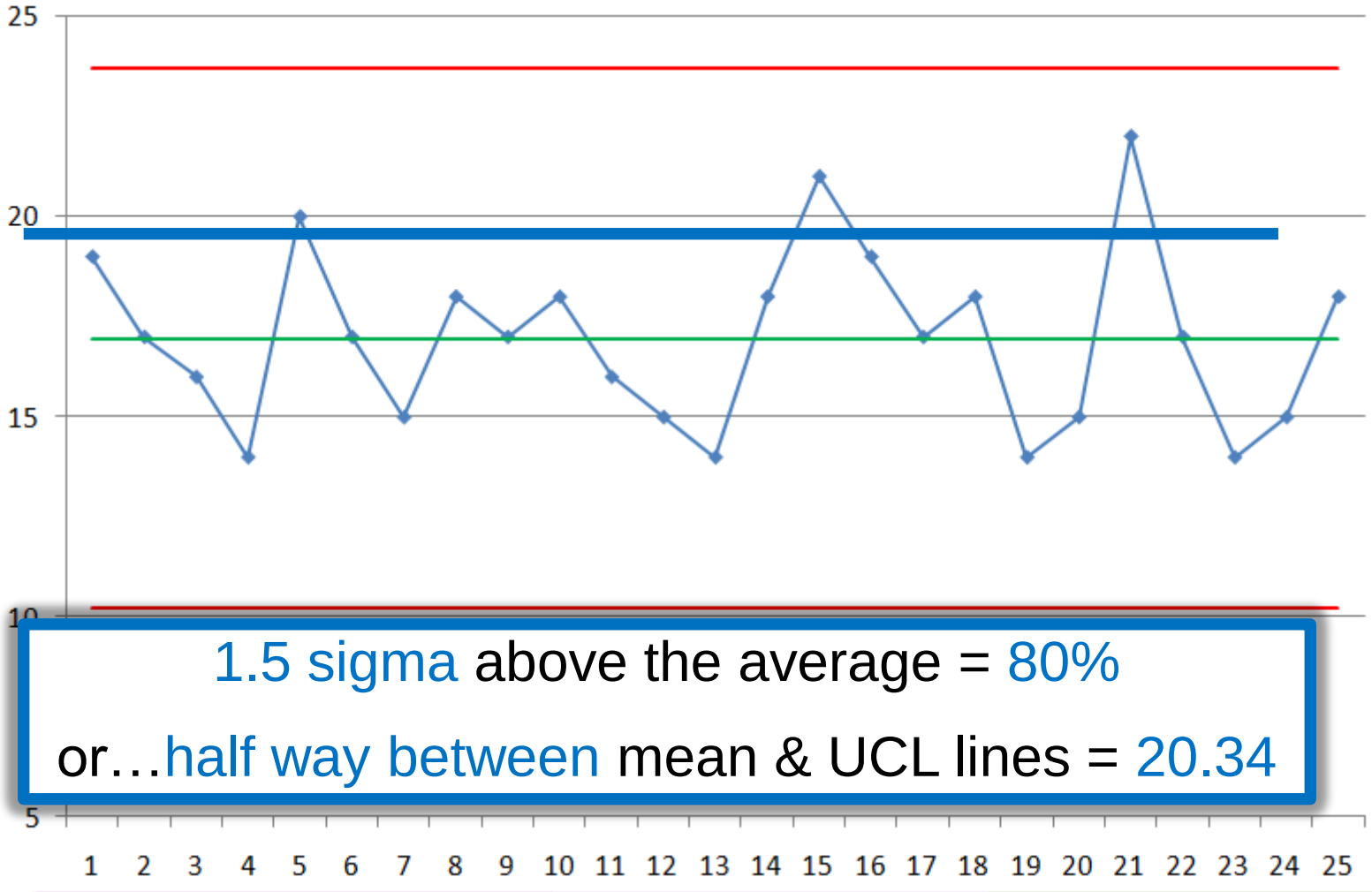
	Assessment time (mins)
16	**16**
17	**16**
18	**15**
19	**15**
20	**15**
21	**15**
22	**15**
23	**14**
24	**14**
25	**14**

Mean?	16.96
Median?	18
Mode?	18

Which of these times could we use for planning if we can only use one?

Calculating Psychiatric Assessment times...

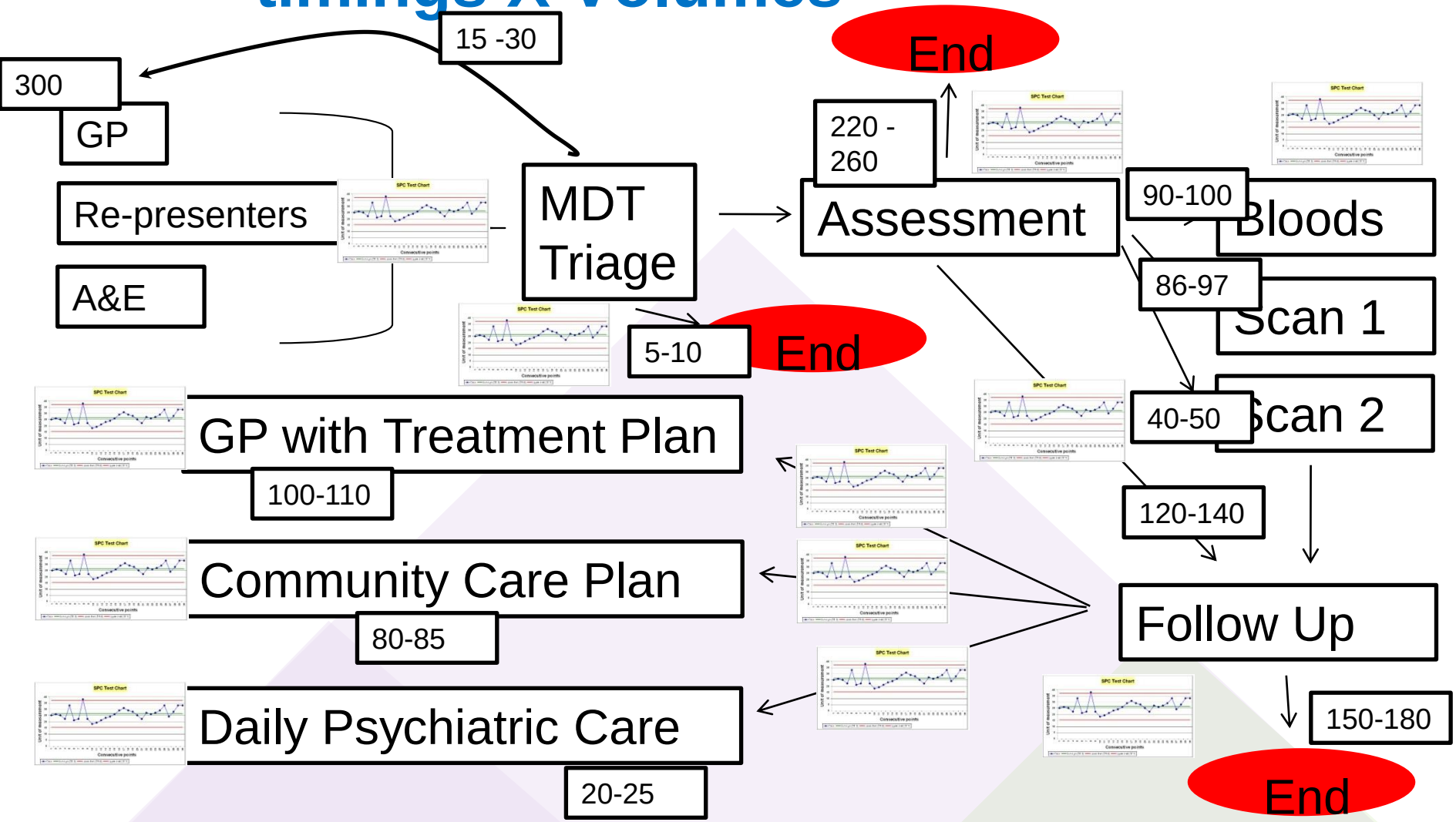
Even better if….use the 80th percentile



1.5 sigma above the average = 80%

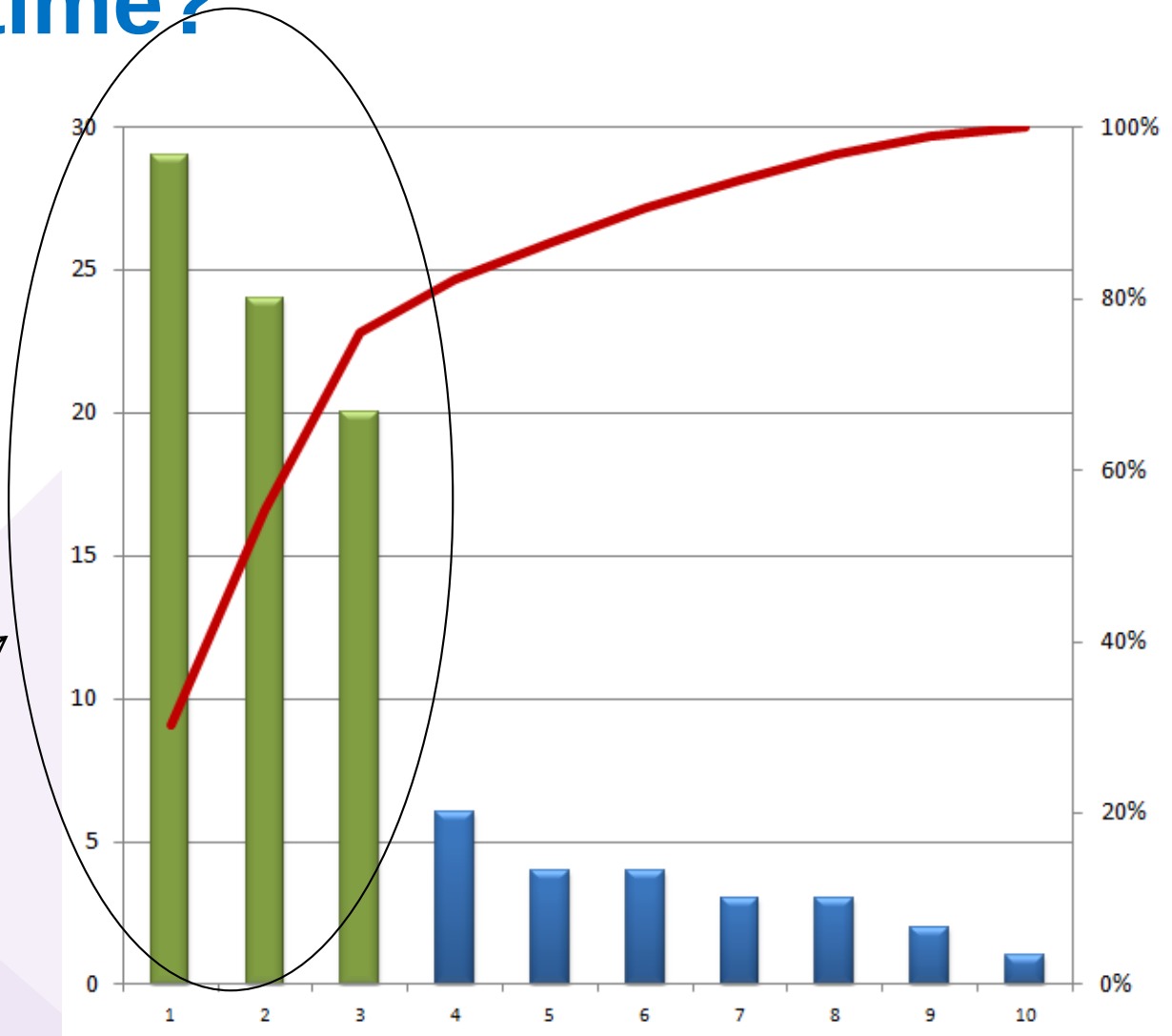
or…half way between mean & UCL lines = 20.34

Psychiatric Process map of the system Stage 3 – work out 80% timings X Volumes



Pareto chart from the process map – where do we spend most of the time?

80% of the patient journey is spent in these three stages, start here for big improvements



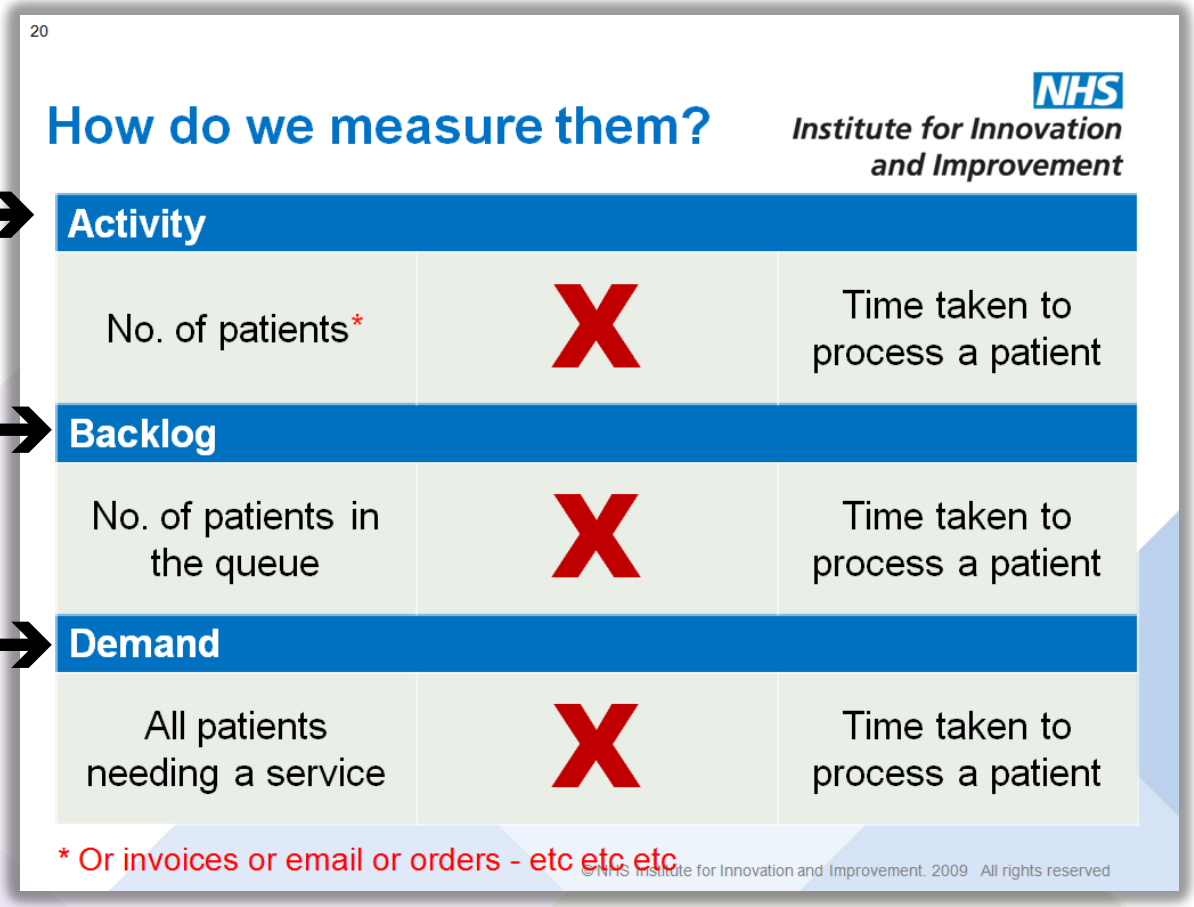
Calculating A, B & D for a CPN (Community Psychiatric Nurse)

Now use these timing figures across:

Activity →

Backlog →

Demand →

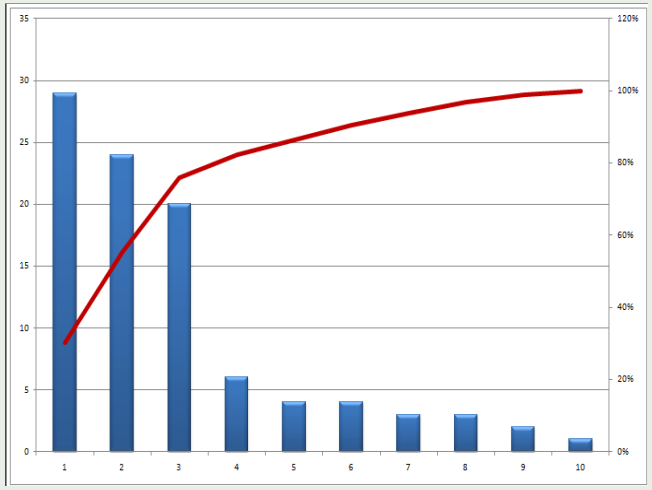


In the real NHS….

We should collect just enough data and use this to project over all the patients we see…

How?

We will use Pareto…



“80 / 20” Rule

Calculating A, B & D for a Community Nurse

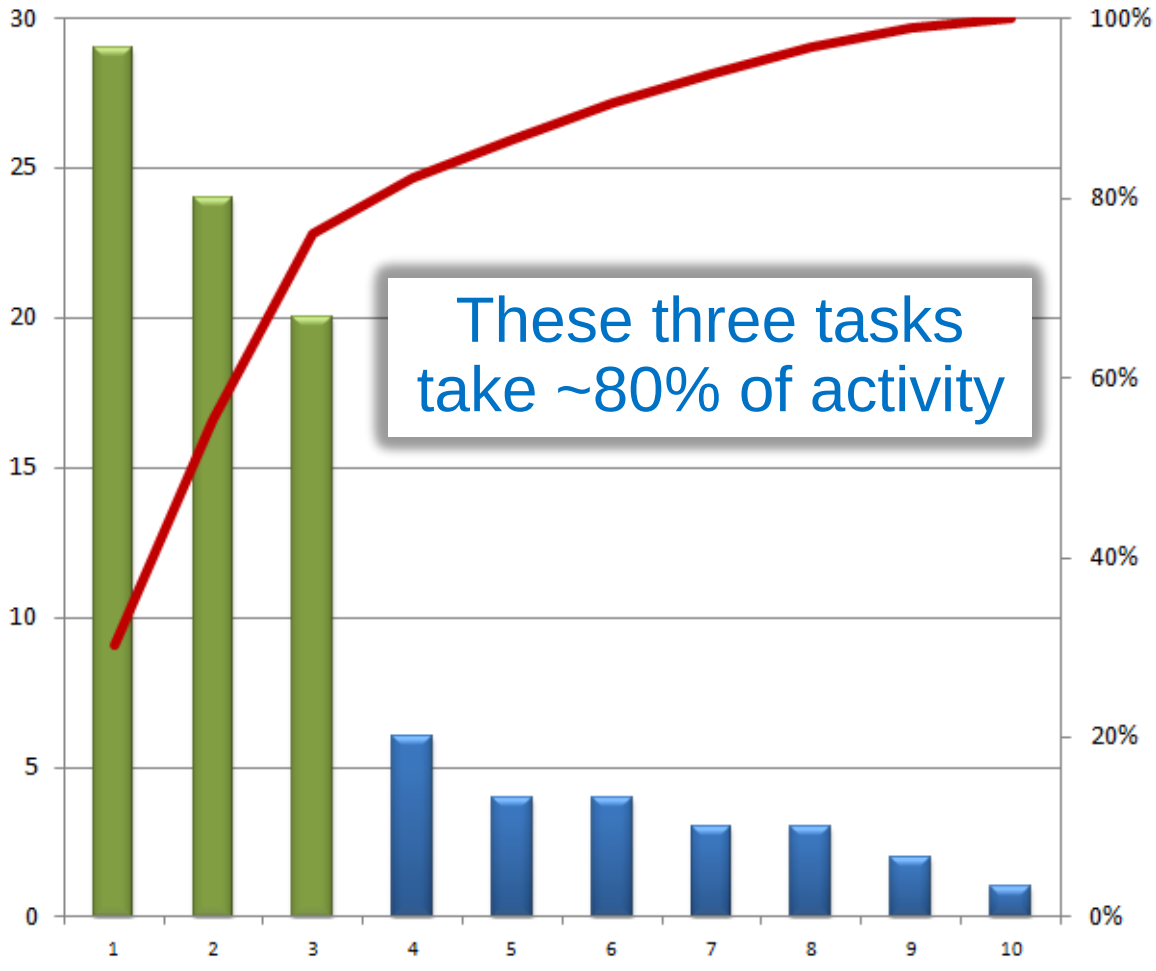
For 3 days we collected a tally chart of activities for each of the 4 community nurses:

Nurse Activity:	Tally Chart	Totals:
Eye drops	𝍸	3
Creating nursing plans	𝍷	2
Taking temp and BP	𝍸 𝍸 𝍸 𝍸 𝍷	24
Giving advise	𝍷	4
Taking blood samples	𝍸 𝍸 𝍸 𝍸	20
Prescribing aids	𝍵	1
Setting up drips	𝍸	3
Dressing wounds	𝍷	4
Morphine pump	𝍸 𝍸 𝍸 𝍸 𝍸 𝍷	29
Taking urine samples	𝍸 𝍵	6

Calculating A, B & D for a Community Nurse

From the tally chart we create a Pareto chart - to work out what activities make up 80% of the community nurses total

Sort by size:
29
24
20
6
4
4
3
3
2
1



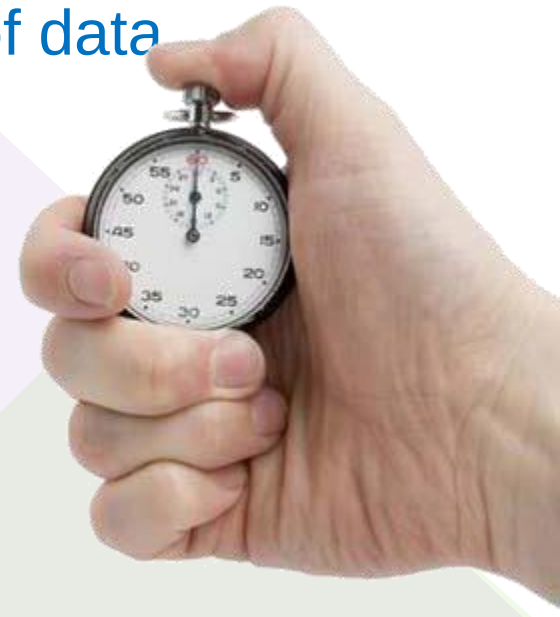
Calculating A, B & D for a Community Nurse

Each task that takes 80% of the nurses time, will not only vary between nurses, but for the same nurse the task time will vary too.

We must incorporate this in the calculations, not ignore it.

Collect for each of the “80%” tasks - 25 sets of data
(shouldn’t take long, as these are
the tasks nurses do repeatedly)

Remember M&M game…



Calculating A, B & D for a Community Nurse

Morphine pump (mins)	
1	**22**
2	**21**
3	**20**
4	**19**
5	**19**
6	**18**
7	**18**
8	**18**
9	**18**
10	**18**
11	**17**
12	**17**
13	**17**
14	**17**
15	**17**

Morphine pump (mins)	
16	**16**
17	**16**
18	**15**
19	**15**
20	**15**
21	**15**
22	**15**
23	**14**
24	**14**
25	**14**

If we put the 25 timings in order

LONGEST

↓

SHORTEST

Calculating A, B & D for a Community Nurse

	Morphine pump (mins)
1	**22**
2	**21**
3	**20**
4	**19**
5	**19**
6	**18**
7	**18**
8	**18**
9	**18**
10	**18**
11	**17**
12	**17**
13	**17**
14	**17**
15	**17**

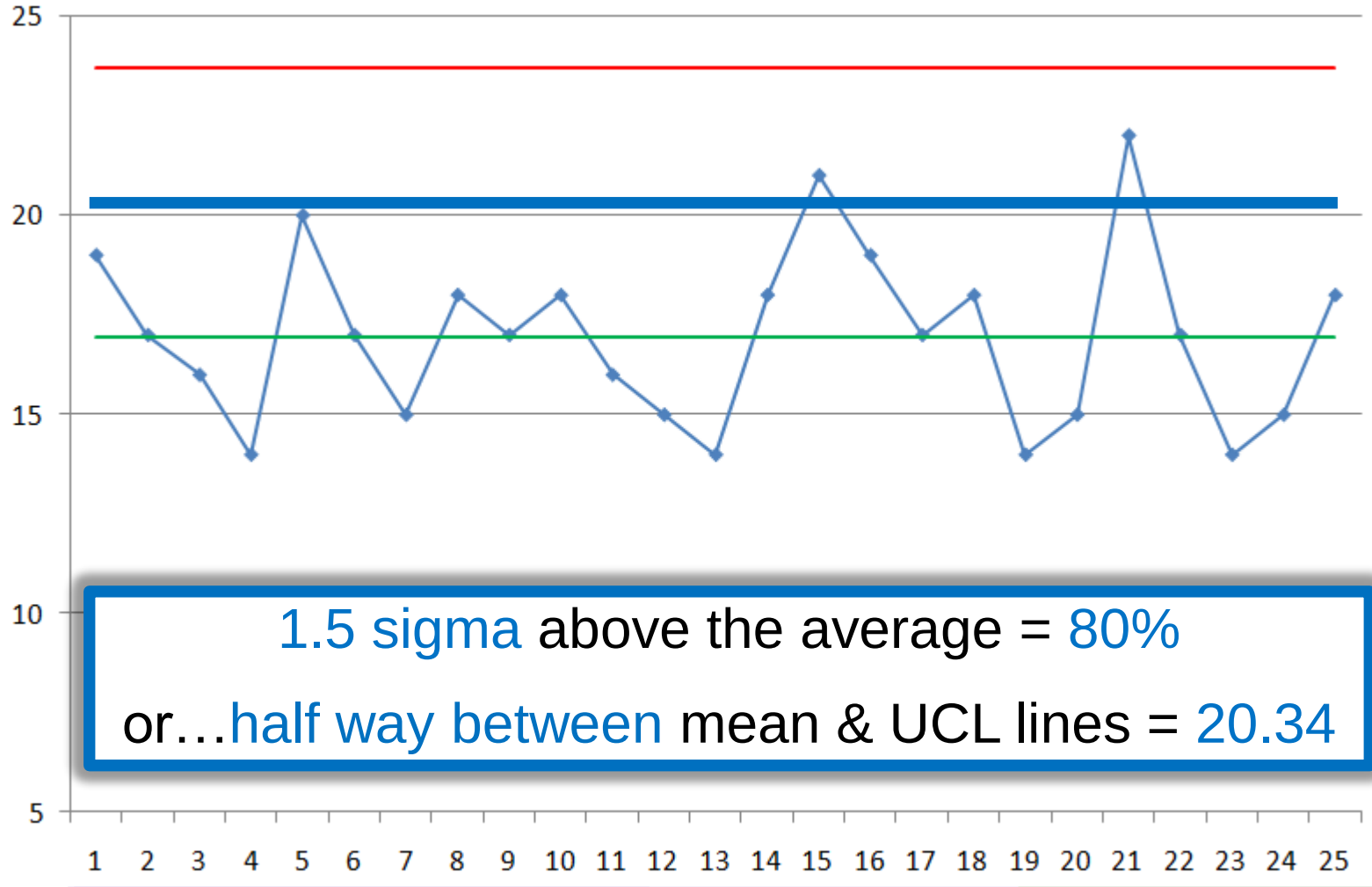
	Morphine pump (mins)
16	**16**
17	**16**
18	**15**
19	**15**
20	**15**
21	**15**
22	**15**
23	**14**
24	**14**
25	**14**

Mean?	16.96
Median?	18
Mode?	18

Which of these times could we use for planning if we can only use one?

Calculating A, B & D for a Community Nurse

Even better if….use the 80th percentile



1.5 sigma above the average = 80%

or…half way between mean & UCL lines = 20.34

BaseLine 1.00 from www.SAASoft.com Project: ""

Data

Metric ? Upper Spec

Order ? Lower Spec

Format hh:mm

Order	Time	Group	Metric	mR	Special
8			18.00	3.000	
9			17.00	1.000	
10			18.00	1.000	
11			16.00	2.000	
12			15.00	1.000	
13			14.00	1.000	
14			18.00	4.000	
15			21.00	3.000	
16			19.00	2.000	
17			17.00	2.000	
18			18.00	1.000	
19			14.00	4.000	
20			15.00	1.000	
21			22.00	7.000	
22			17.00	5.000	
23			14.00	3.000	
24			15.00	1.000	

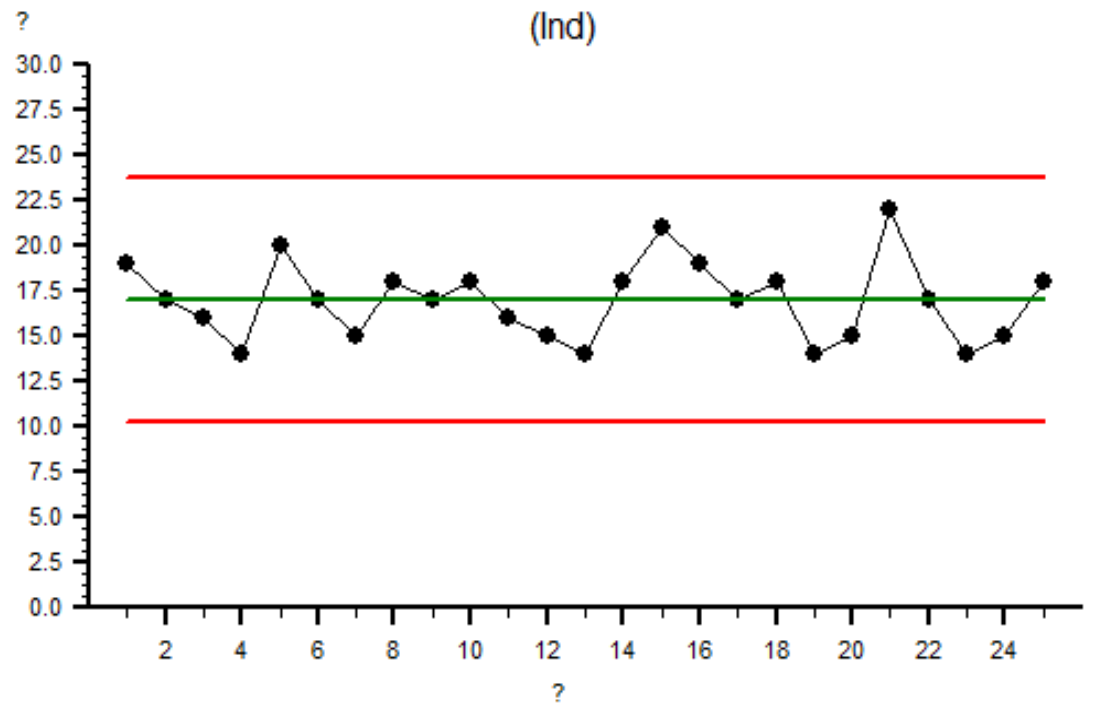
Order	Time	Group	Metric
22			17.00

Add Set Ins Excl Incl Del Splt Mrg Lock L2S

Help User Guide Copy Paste Append Undo

Individuals Chart

Right Click on Chart for Options



- + Axes Edit Note

Segment	Start	Finish	Mean	Sigma	Count	LCL	UCL	Stable
1	1	25	16.96	2.253	25	10.20	23.72	Yes

Half way between = 20.34

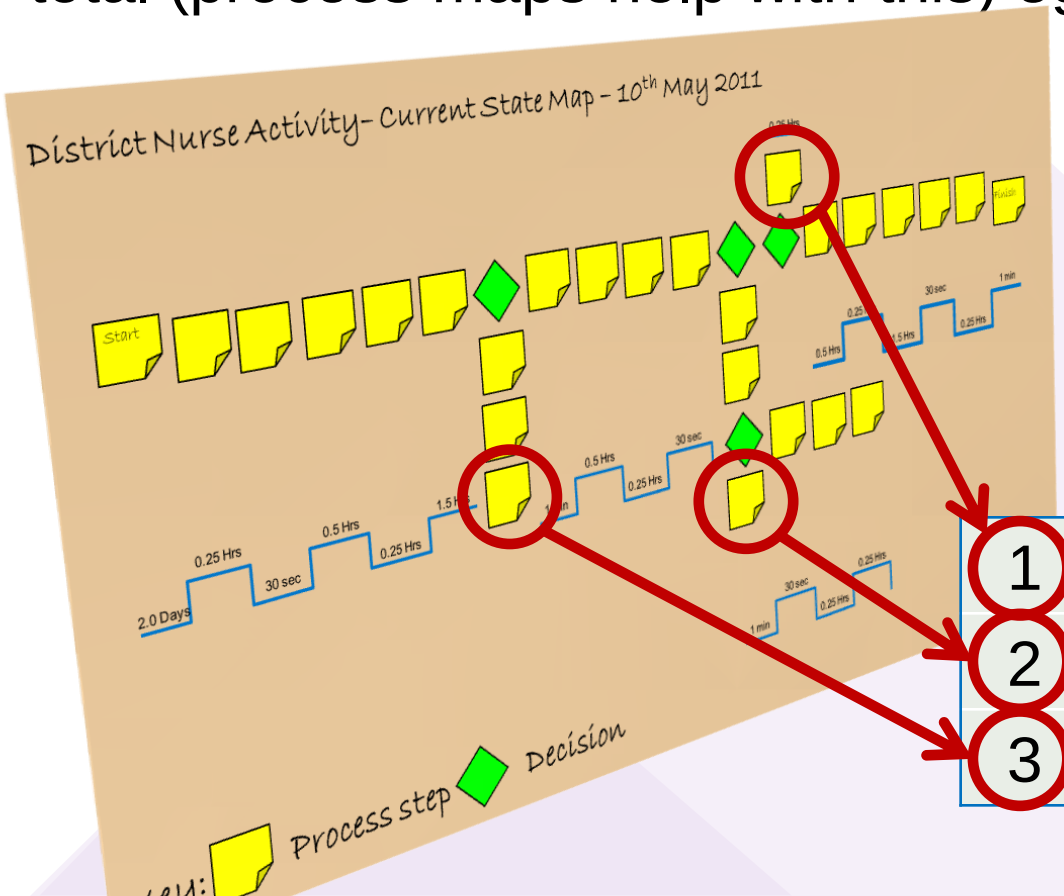
☐ Sigma Levels ☐ Specification ☐ Histogram ☐ Groups ☑ Flags

Data Run Ind mR Notes New Load Save Export Exit

Calculating A, B & D for a Community Nurse

Repeat for the other activities making up 80% of the time

Measure the proportions of those key activities as part of the total (process maps help with this) eg:



1	Morphine pump	30%
2	Temp & BP	25%
3	Bloods	20%

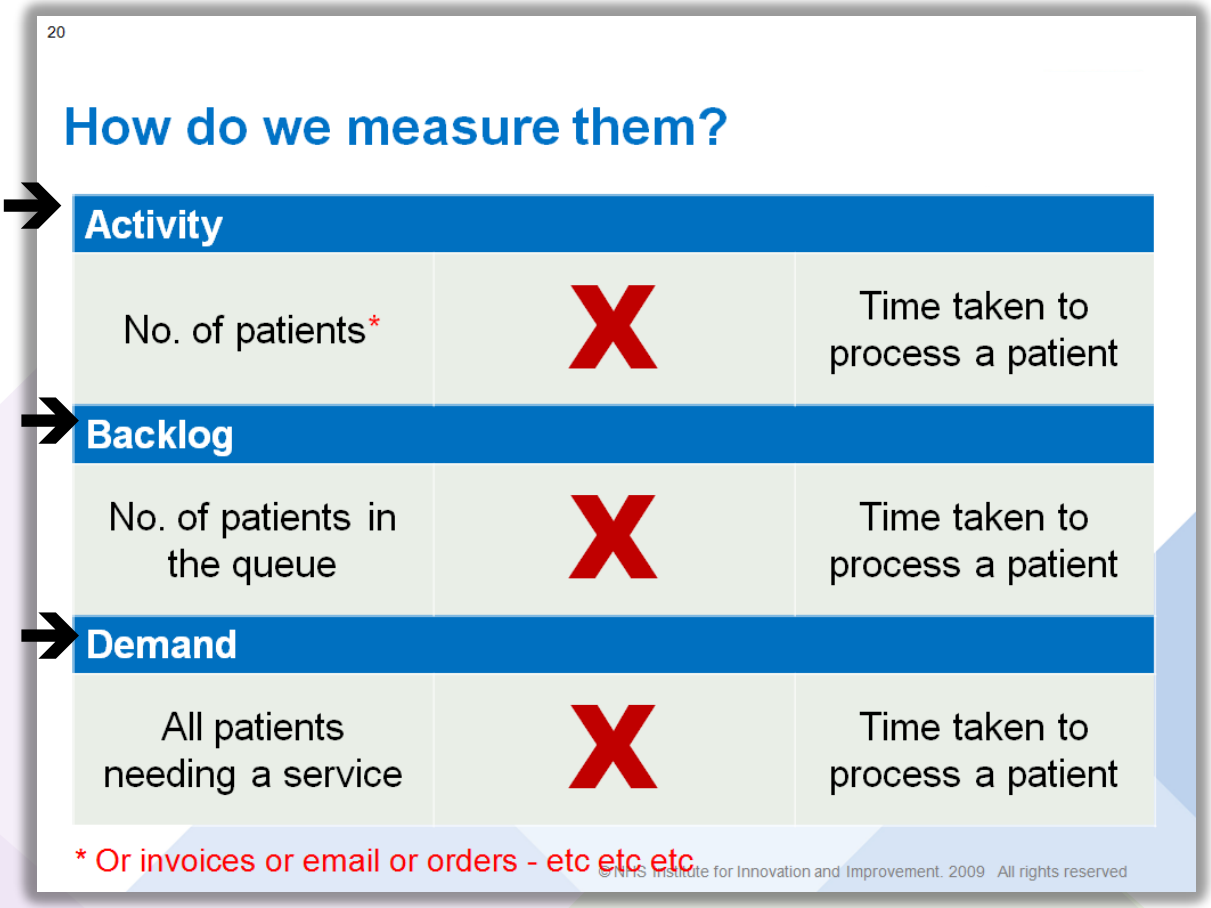
Calculating A, B & D for a Community Nurse

Now use these timing figures across:

Activity

Backlog

Demand



Key message

Measure everything in the same units for the same period

Key message

Keep it simple!

"Use crude measures of the right thing rather than precise measures of the wrong thing"

Richard Steyn

Practical exercise

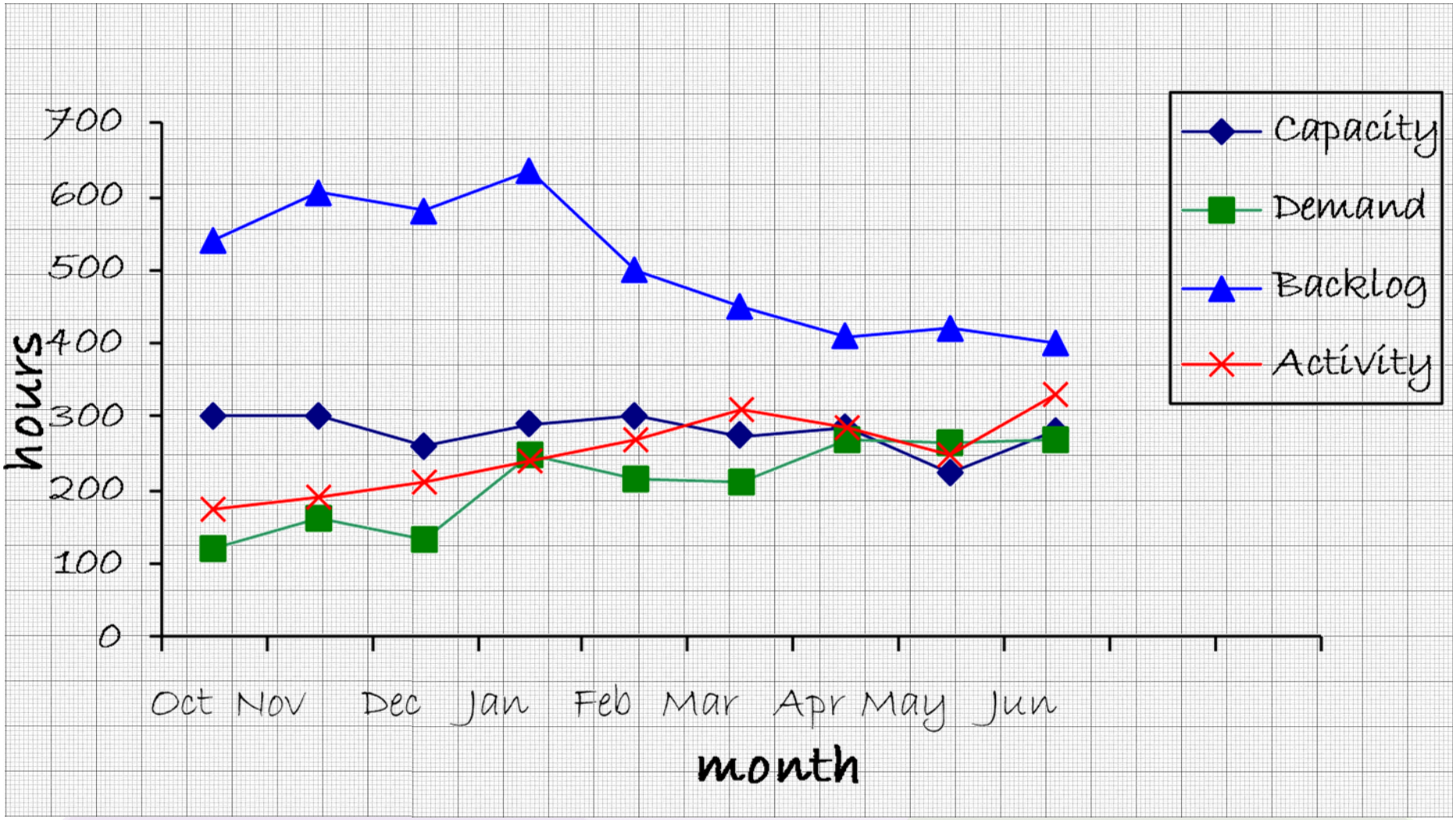
We used the following data (see handout) to plot C,D,B,A:

Measure = hours				
	Capacity	Demand	Backlog	Activity
Oct	300	120	540	175
Nov	300	160	605	190
Dec	260	135	580	210
Jan	290	250	635	240
Feb	300	215	500	270
Mar	275	210	450	310
Apr	285	270	410	285
May	225	265	420	250
Jun	280	270	400	330

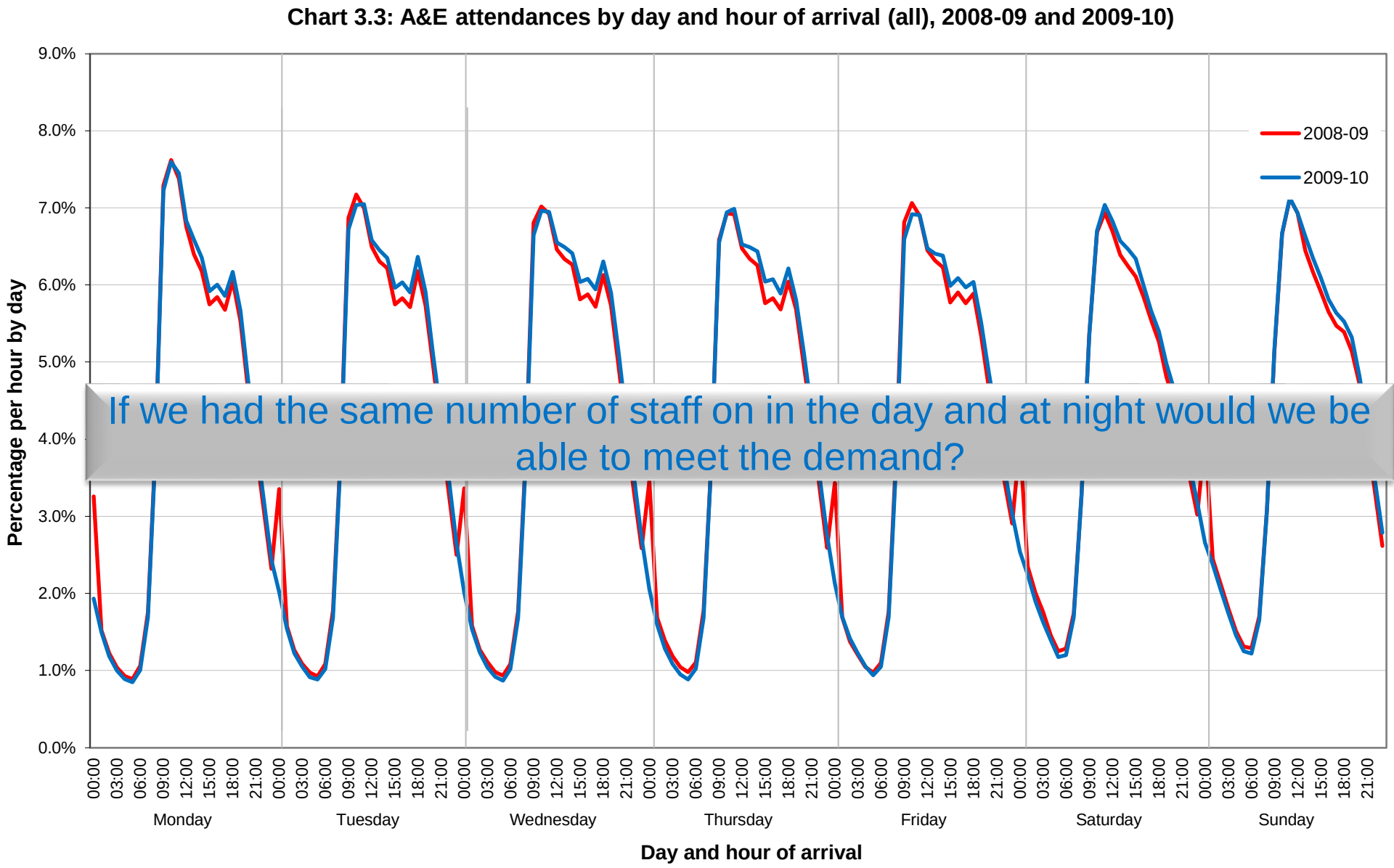
Now that you are all expert analysts we would like you to analyse the results on the following slide, and feedback your findings….

D, A, C and B

Demand, Activity, Capacity and Backlog



Predictable demand?



If we had the same number of staff on in the day and at night would we be able to meet the demand?

Chart 3.8: A&E attendances by day and hour of arrival (Road accidents), 2008-09 and 2009-10

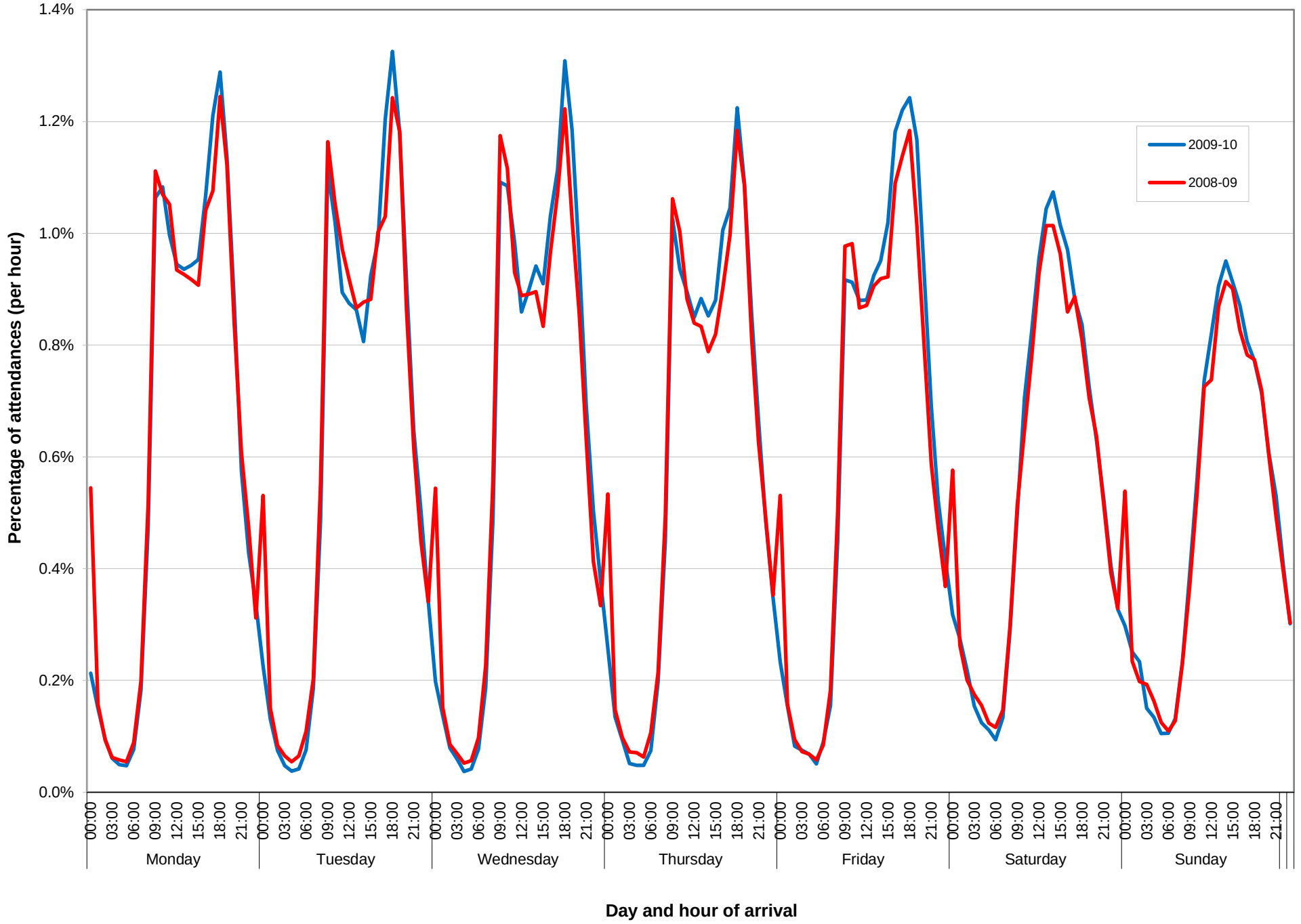


Chart 3.9: A&E attendances by day and hour of arrival (Assaults), 2008-09 and 2009-10

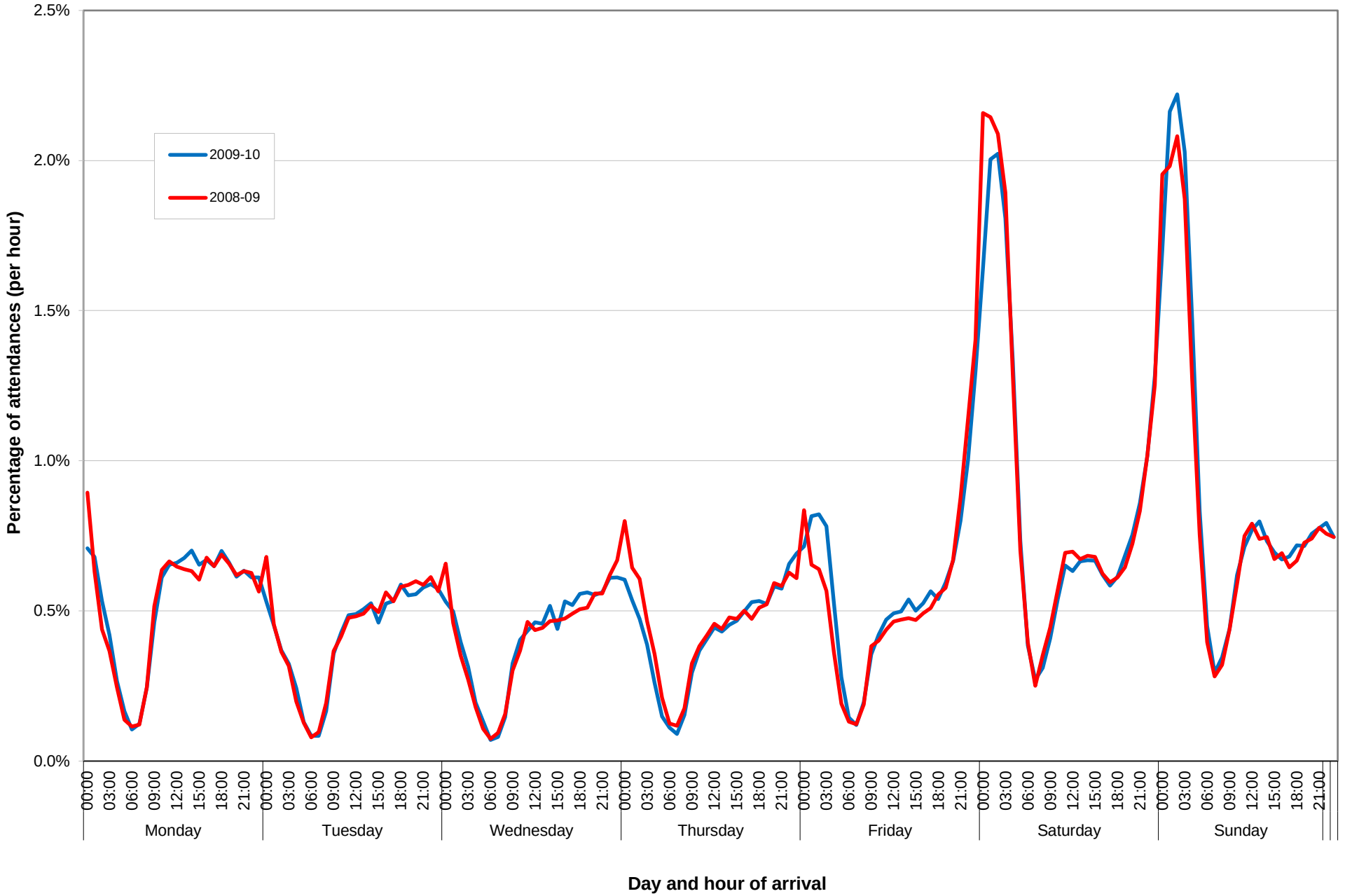


Chart 3.10: A&E attendances by day and hour of arrival (Self harm), 2008-09 and 2009-10

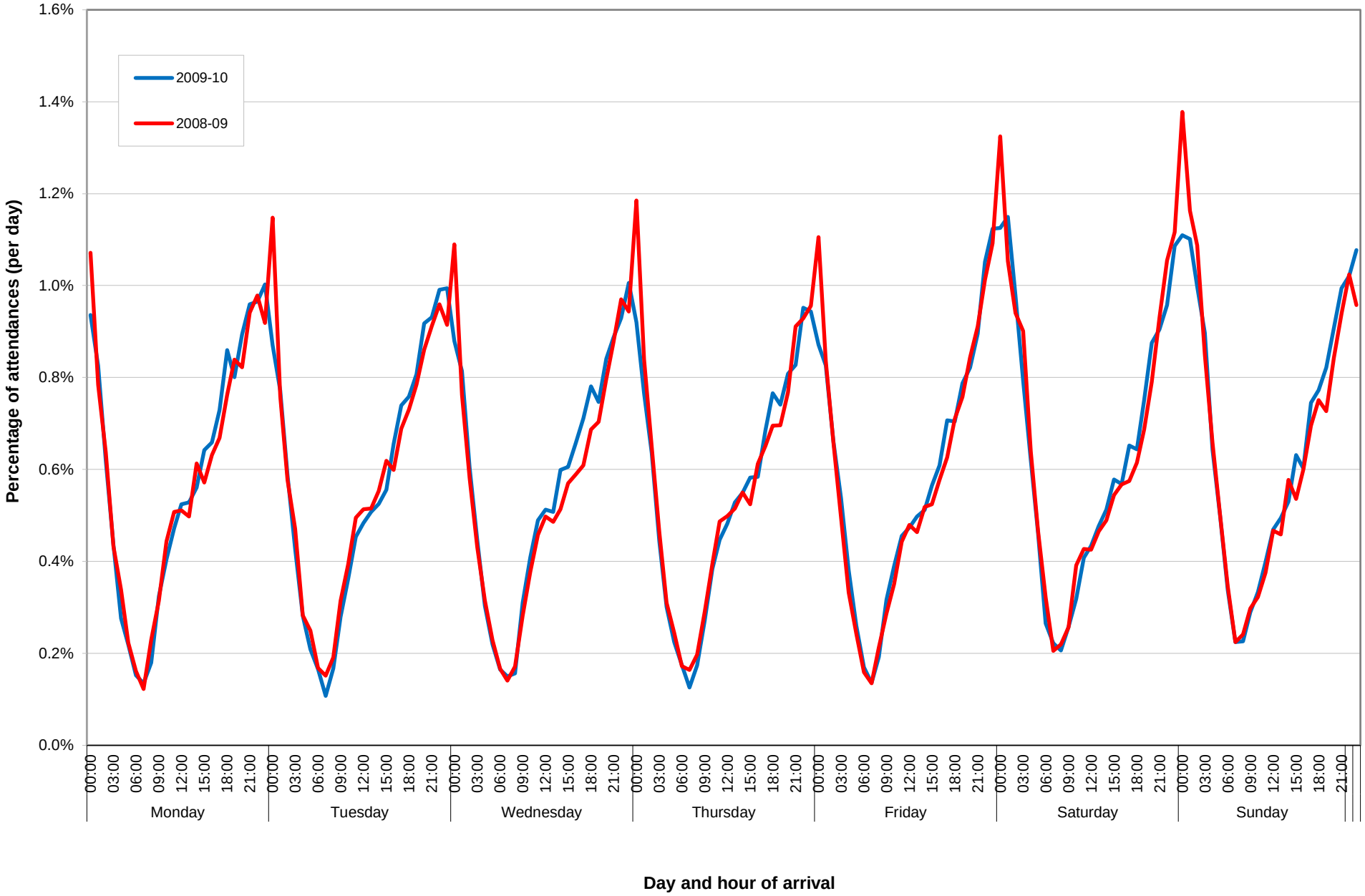
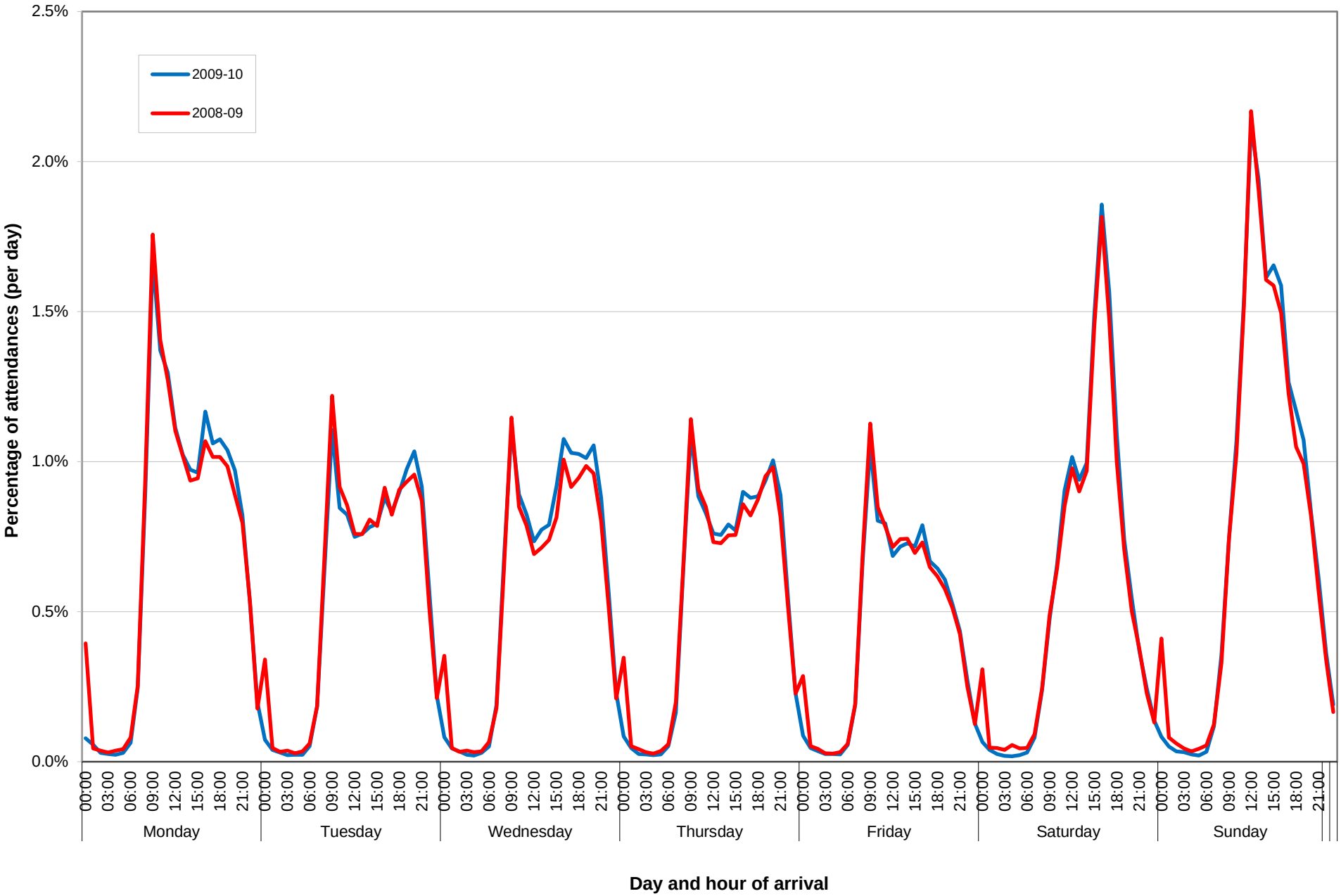


Chart 3.11: A&E attendances by day and hour of arrival (Sports Injuries), 2008-09 and 2009-10



Measuring capacity

Capacity is all of the resources required to do the work and includes equipment, rooms and the people with the necessary skills to use it.

Capacity		
No. of resources available – eg autoclaves, MRI scanners etc	X	Staff time available to run those resources

Bottlenecks

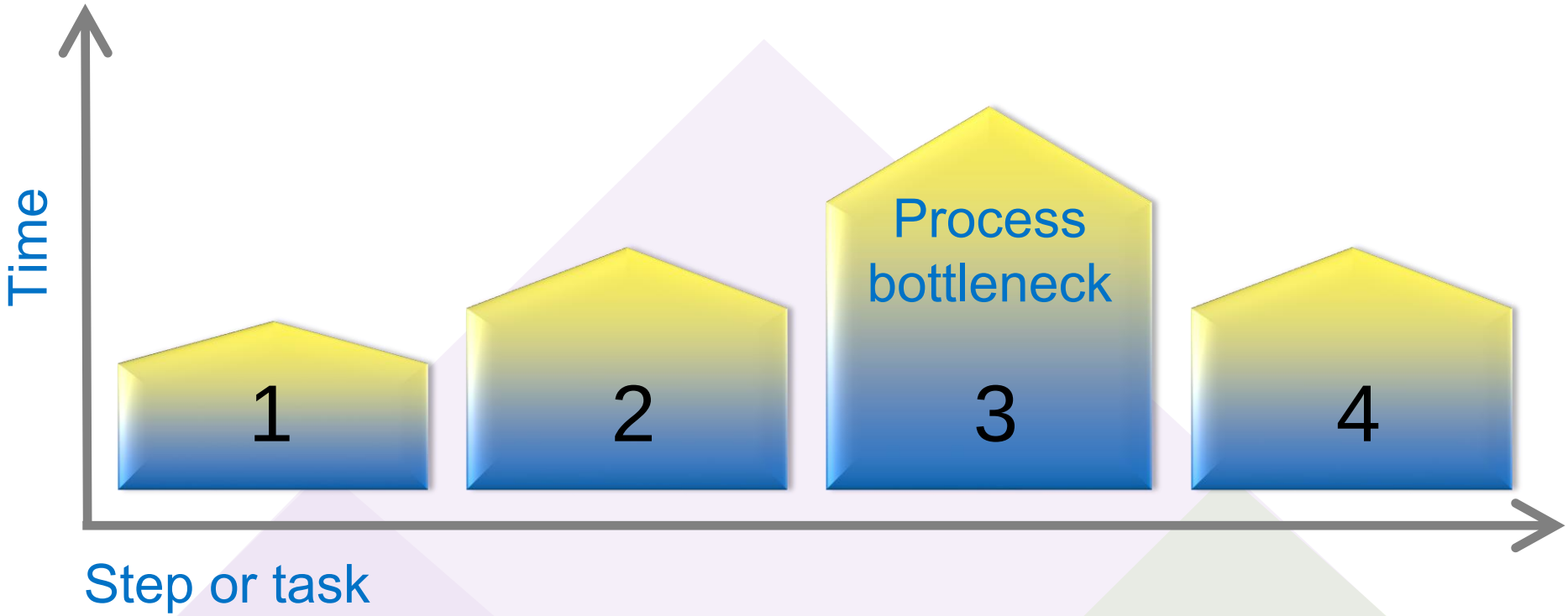
A bottleneck is where the queues form - it will slow down the whole process

Two types of bottleneck:

1. Process
2. Functional

Process bottleneck

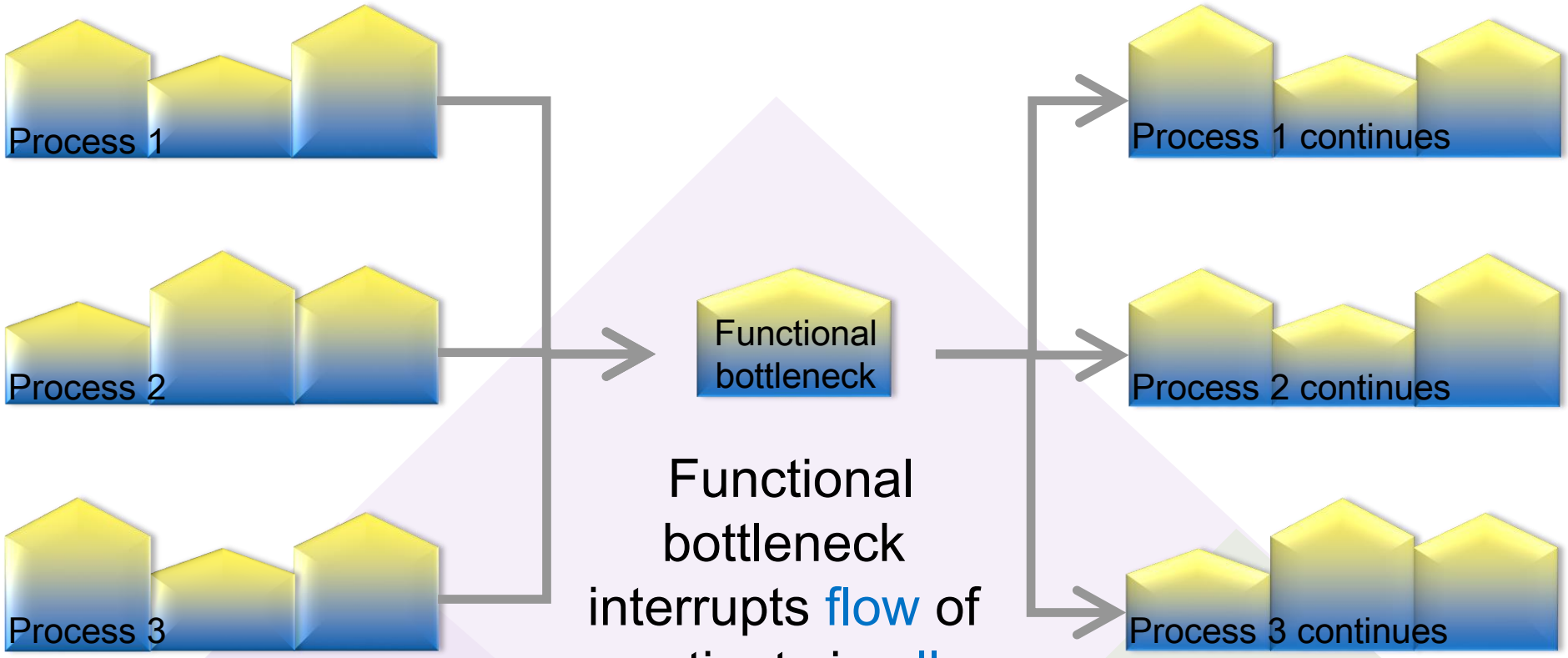
Process bottlenecks are the process stage that takes longest to complete. Sometimes referred to as 'rate limiting' step or task



Which step or task is the bottleneck?

Functional bottleneck

Functional bottlenecks caused by services that have demand from a number of sources e.g. radiology, pathology, porters



Functional bottleneck interrupts flow of patients in all three processes

Constraints

Bottlenecks caused by a constraint
This restricts the capacity (flow) of the service
It may be a particular skill or piece of equipment

Constraint examples:

- Number of treatment rooms
- Specialist skills i.e. surgeon, radiologist
- Decontamination washer/machine
- Theatres
- CT scanner
- Phlebotomist

How to deal with them:
Maximise utilisation of constraint - as not easily increased?

Apply pressure?

On the system and on the people

Performance management (at all levels)

- Changes in behaviour

Targets

- Waiting list initiatives
- Carve out capacity
- Forced booking
- High utilization of capacity

The affects of applied pressure?

Poor decisions?

- Guesswork - single figure decision making
- Decisions based on special cause variation
- Anecdotal data
- “Known” solutions or assumptions

Cost cutting?

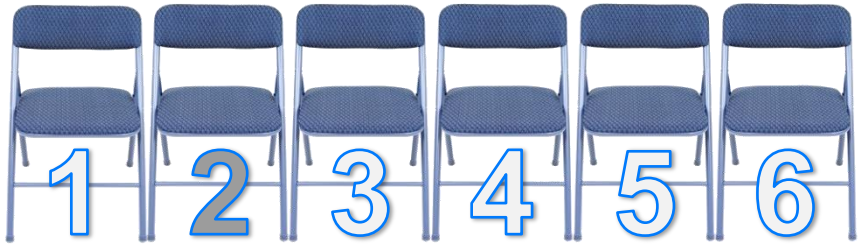
- becomes indiscriminate

Quality drops?

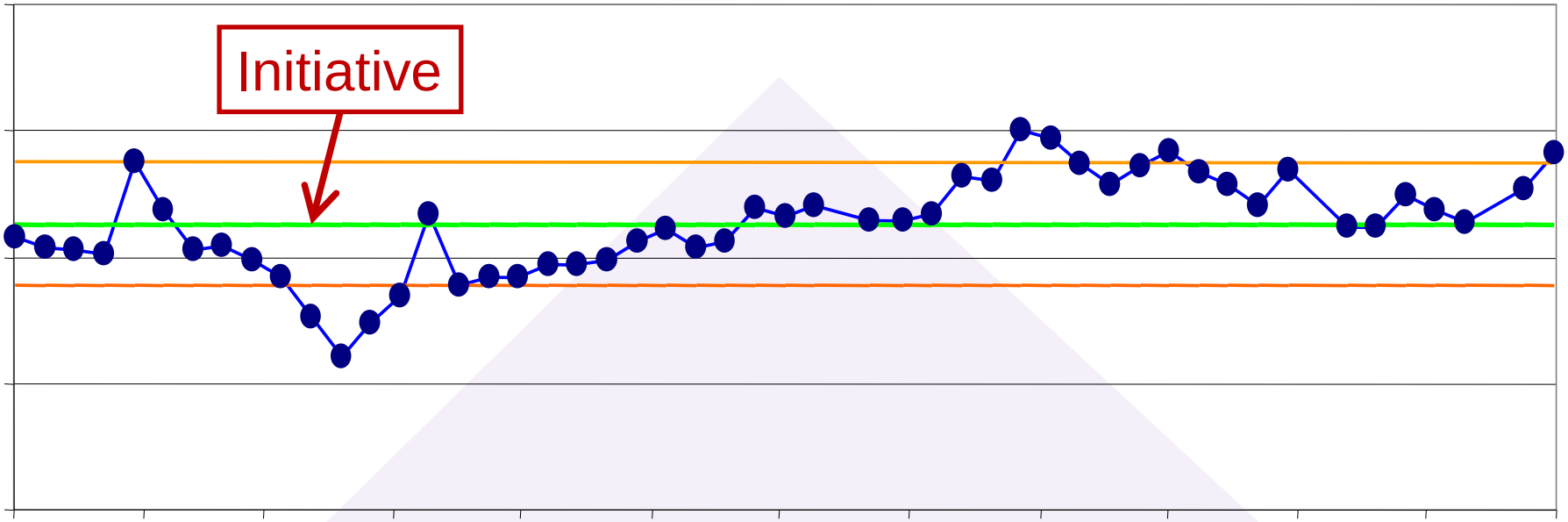
- re-work, increased clinical risks to patients, complaints, litigation, increased costs

Demoralised staff?

- staff turnover, sickness and absenteeism



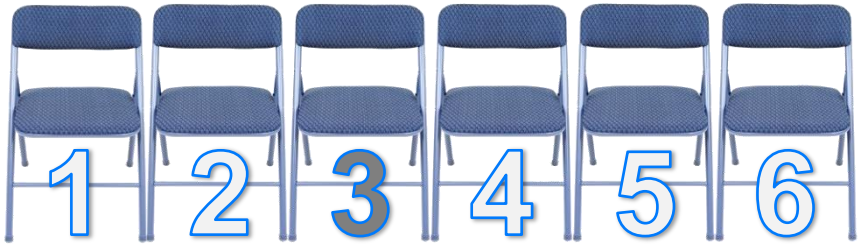
Waiting list initiatives



Luton and Dunstable NHS Trust

Waiting list numbers - weekly from Jan to Jan

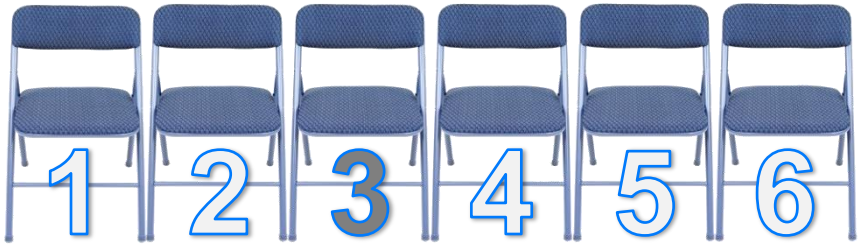
Was the initiative sustained?



High Utilisation

Utilisation is a measure of how much capacity is used (%age).

What is an appropriate utilisation target for a service?



High Utilisation

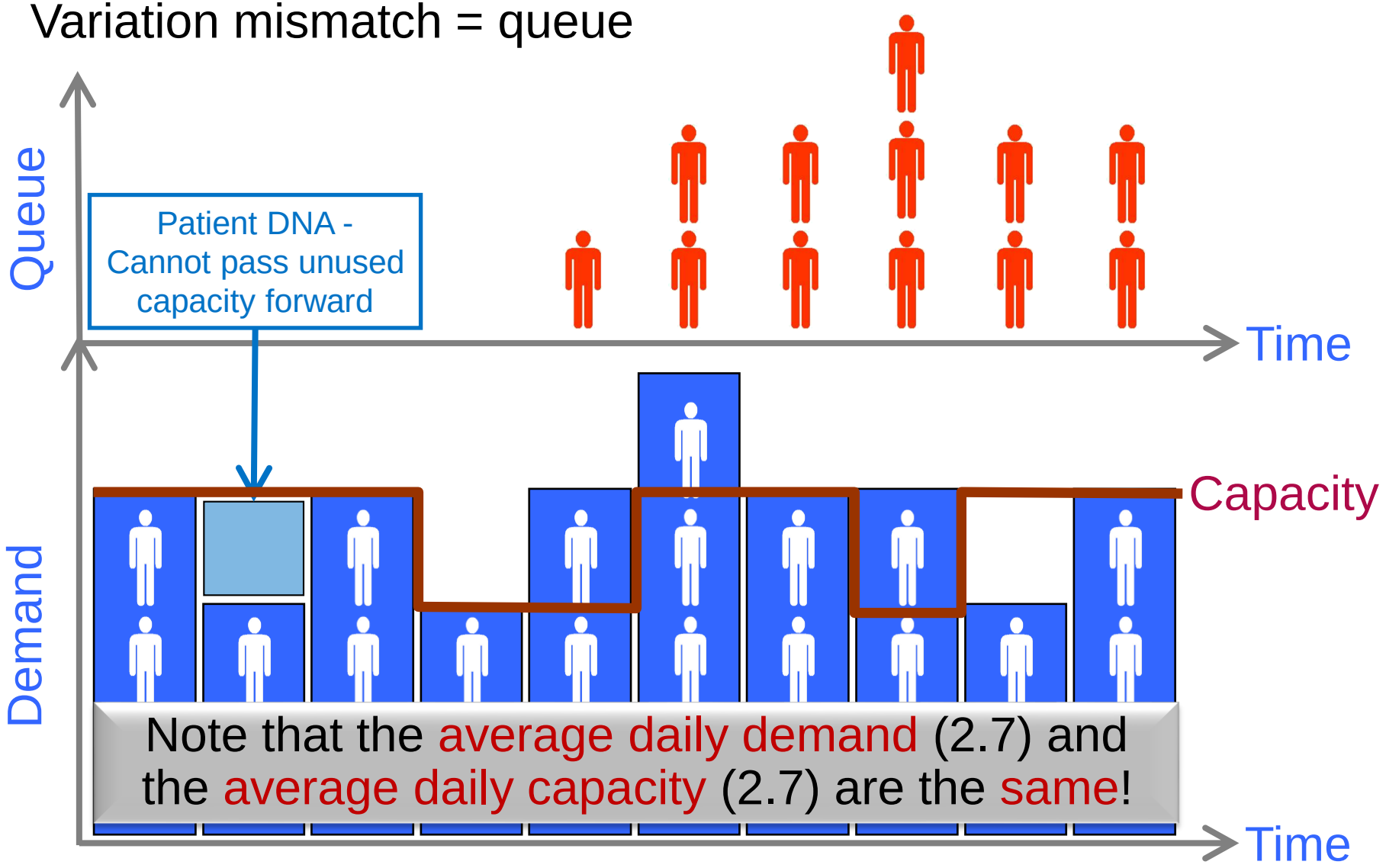
Services often aim for 90 - 100%

These services are set up to fail if they do not take into account inherent variation in demand

Pressure to fully utilise resources can also lower staff morale and trigger adverse changes in behaviour

Wasted capacity

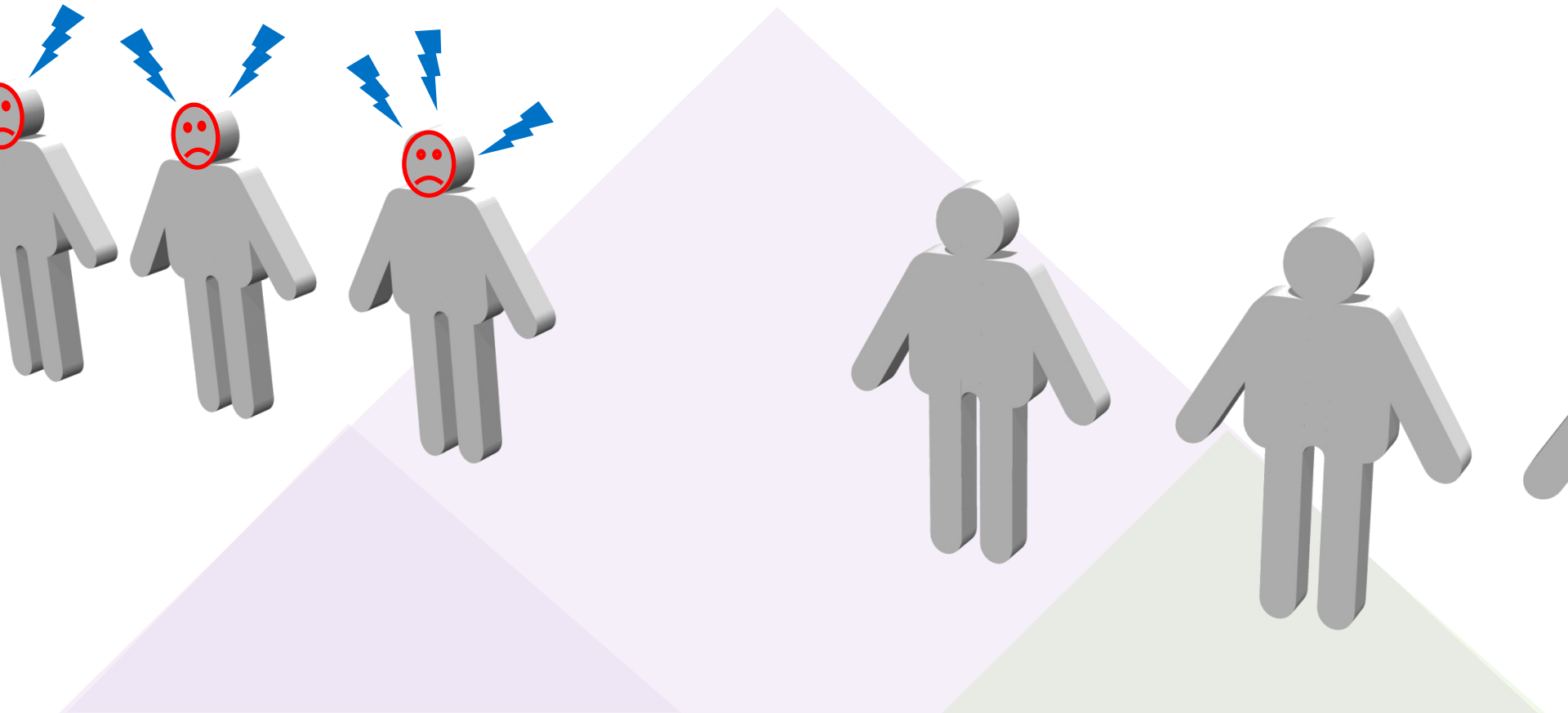
Variation mismatch = queue



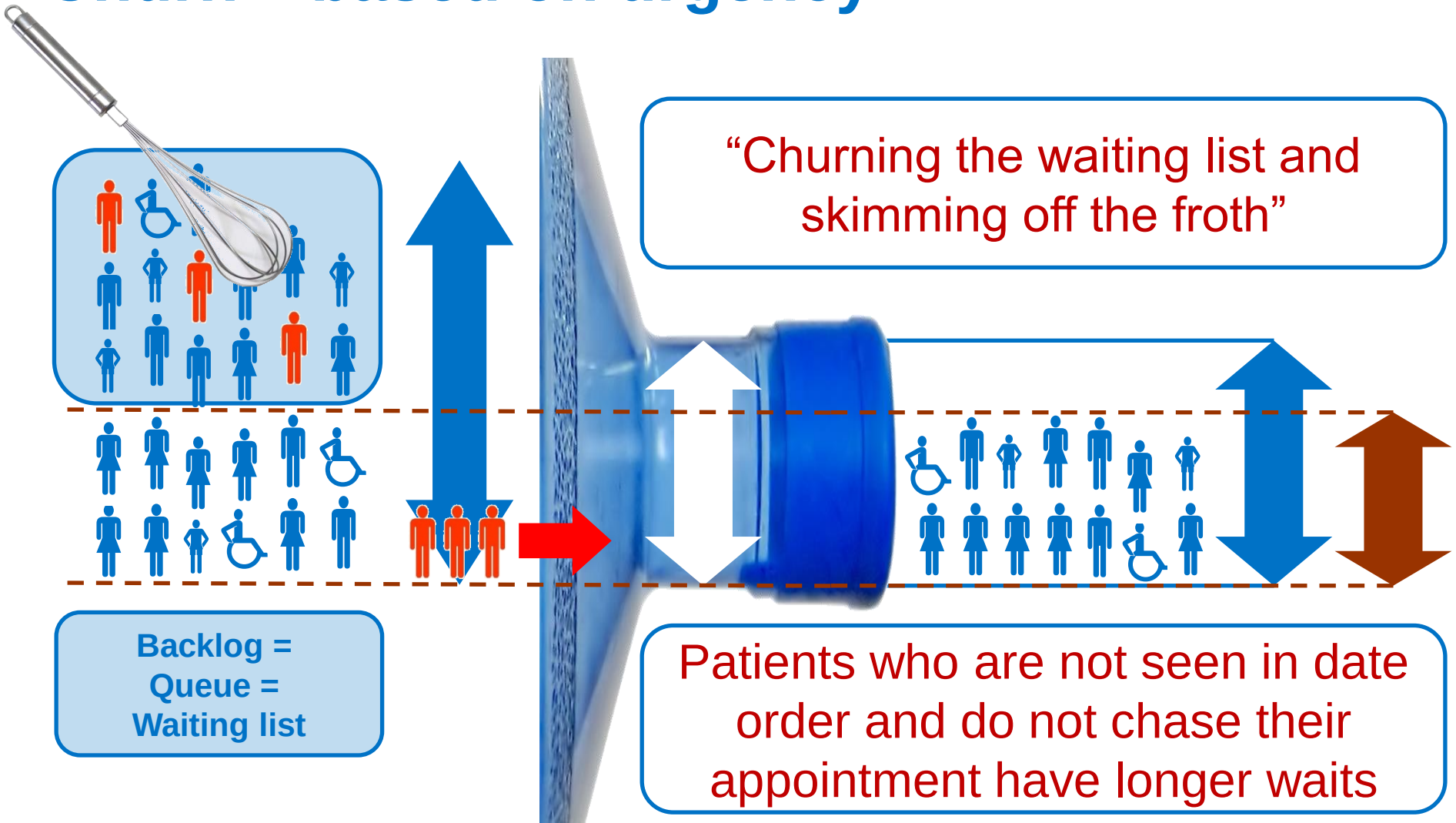
Problem with averages

"If I stick my right foot in a bucket of boiling water and my left foot in a bucket of ice water, on average, I'm pretty comfortable."

Forced booking



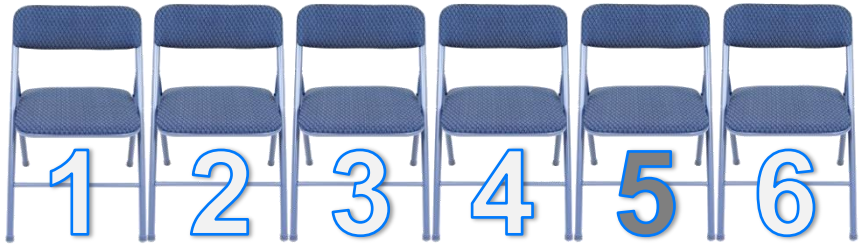
Churn – based on urgency



"Churning the waiting list and skimming off the froth"

Backlog =
Queue =
Waiting list

Patients who are not seen in date order and do not chase their appointment have longer waits



Carve out

- Carve out is capacity retained for a particular type of patient, operation or test
- Causes patients in the routine queue to wait longer
- Slots go unused or are misused as patients jump the queue - leading to increased variation
- Flow of one group of patients is improved to the detriment of others…

Carve out example : Endoscopy service

Specialists →		Surgeon				Physician					R
Appointment type ↓		1	2	3	4	1	2	3	4	5	1
Flexi Sig	Urgent	◎	◎	◎	◎	◎	◎	◎	◎	◎	
	Soon	◎	◎	◎	◎	◎	◎	◎	◎	◎	
	Routine	◎	◎	◎	◎	◎	◎	◎	◎	◎	
Colonoscopy	Urgent	◎	◎	◎	◎				◎	◎	
	Soon	◎	◎	◎	◎				◎	◎	
	Routine	◎	◎	◎	◎				◎	◎	
OGD	Urgent	◎	◎	◎	◎	◎	◎	◎	◎	◎	
	Soon	◎	◎	◎	◎	◎	◎	◎	◎	◎	
	Routine	◎	◎	◎	◎	◎	◎	◎	◎	◎	
ERCP	--										◎

73 queues!

Carve out example

Specialists →		Surgeon & Physician Pool									R
Appointment type ↓		1	2	3	4	1	2	3	4	5	1
Flexi Sig	Urgent					◎					
	Routine					◎					
Colonoscopy	Urgent					◎					
	Routine					◎					
OGD	Urgent					◎					
	Routine					◎					
ERCP	--										◎

7 Queues!

Next step – pool sigmoidoscopy and colonoscopy – 5 queues….

Batch demand

Batching is a method of processing patients, material or information – where they are accumulated into a lot, then pushed through as a group.

Also called the *batch and queue* method – for good reason!

- Treatment or operation types
- Laboratory samples
- Discharges
- Information (weekly decision meetings)
- Emails

Backlogs form because….

1. Demand exceeds capacity?

(but if this were the case then waiting lists would continue to rise without control)

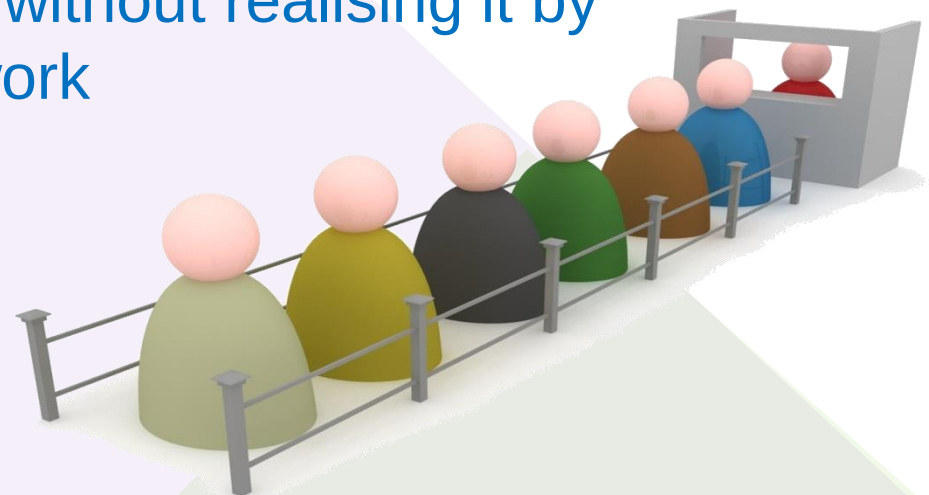
2. There are mismatches in capacity and demand

(caused by variation)

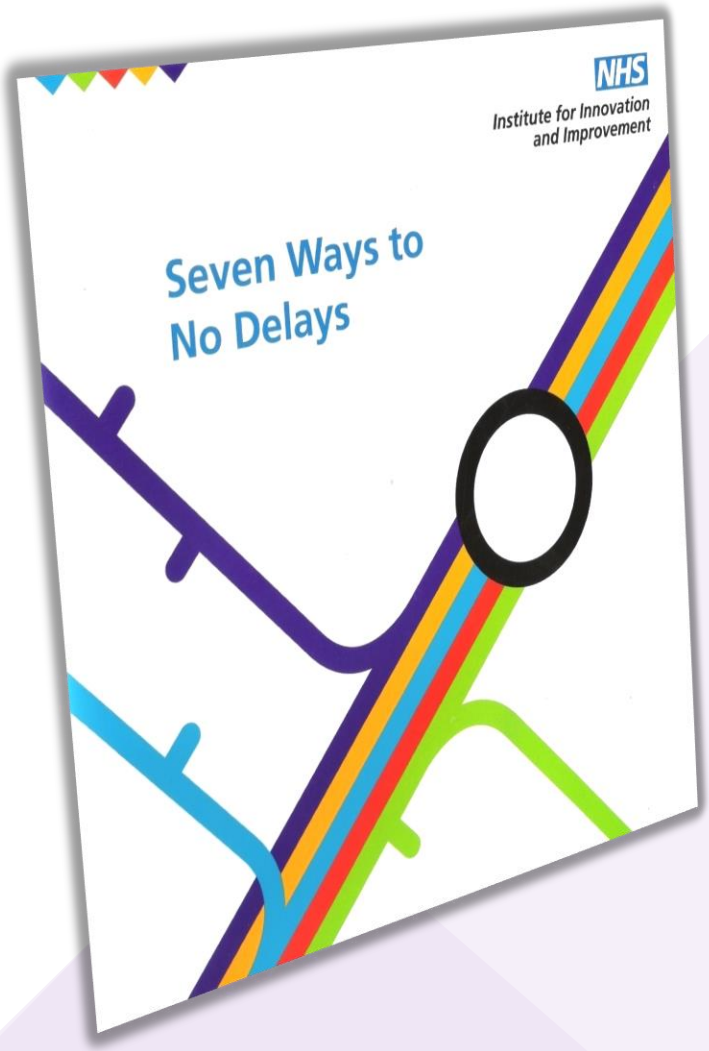
3. We make patients wait more without realising it by batching and ring-fencing our work

Queues keep us busy?

High utilisation?



Seven Ways to No Delays



A guide for clinical teams to help them understand Demand and Capacity - with a strong focus on reducing variation and looking at the bigger picture to improve service for patients

01 Focus on the whole patient pathway

Imagine if one team reduced its backlog of work and saw more patients, but the next team along the patients pathway did not make any changes.

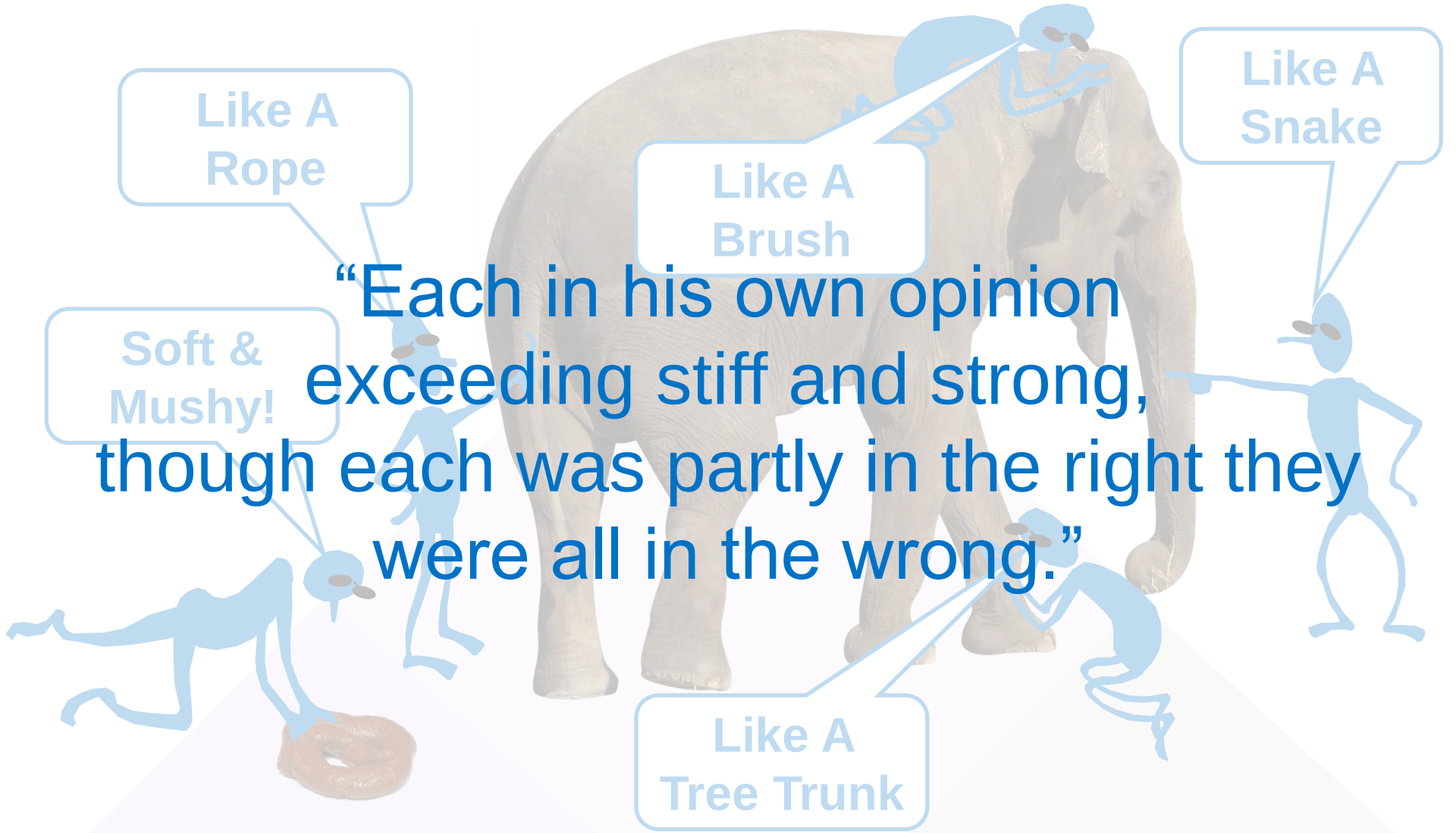
The effect would be for the firsts teams backlog to become the problem of the second team.

If the second team was the bottleneck, no more patients would be seen by the “pathway” as a whole.

Waiting times would not be reduced.

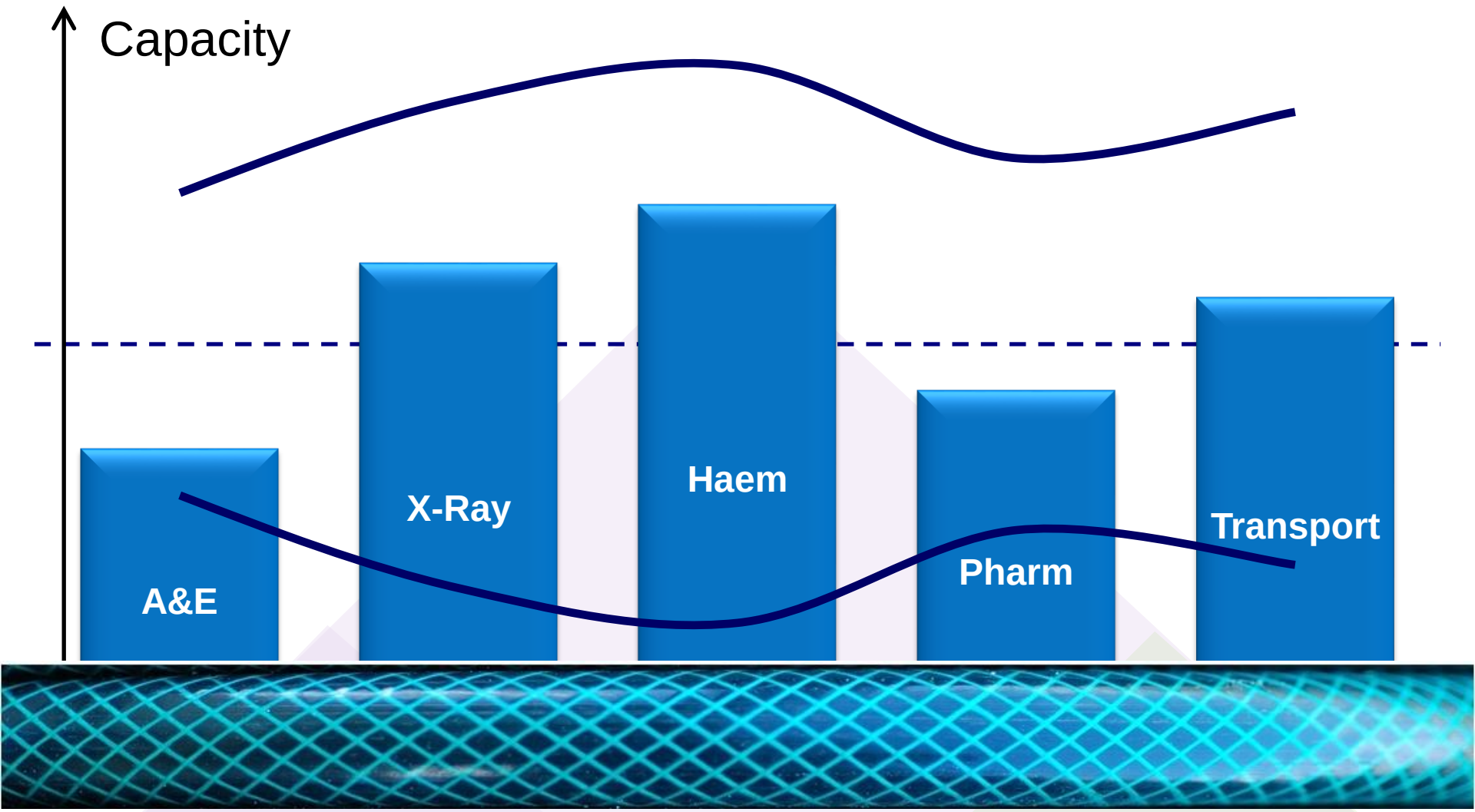
Lets explore further…

What is an elephant like?

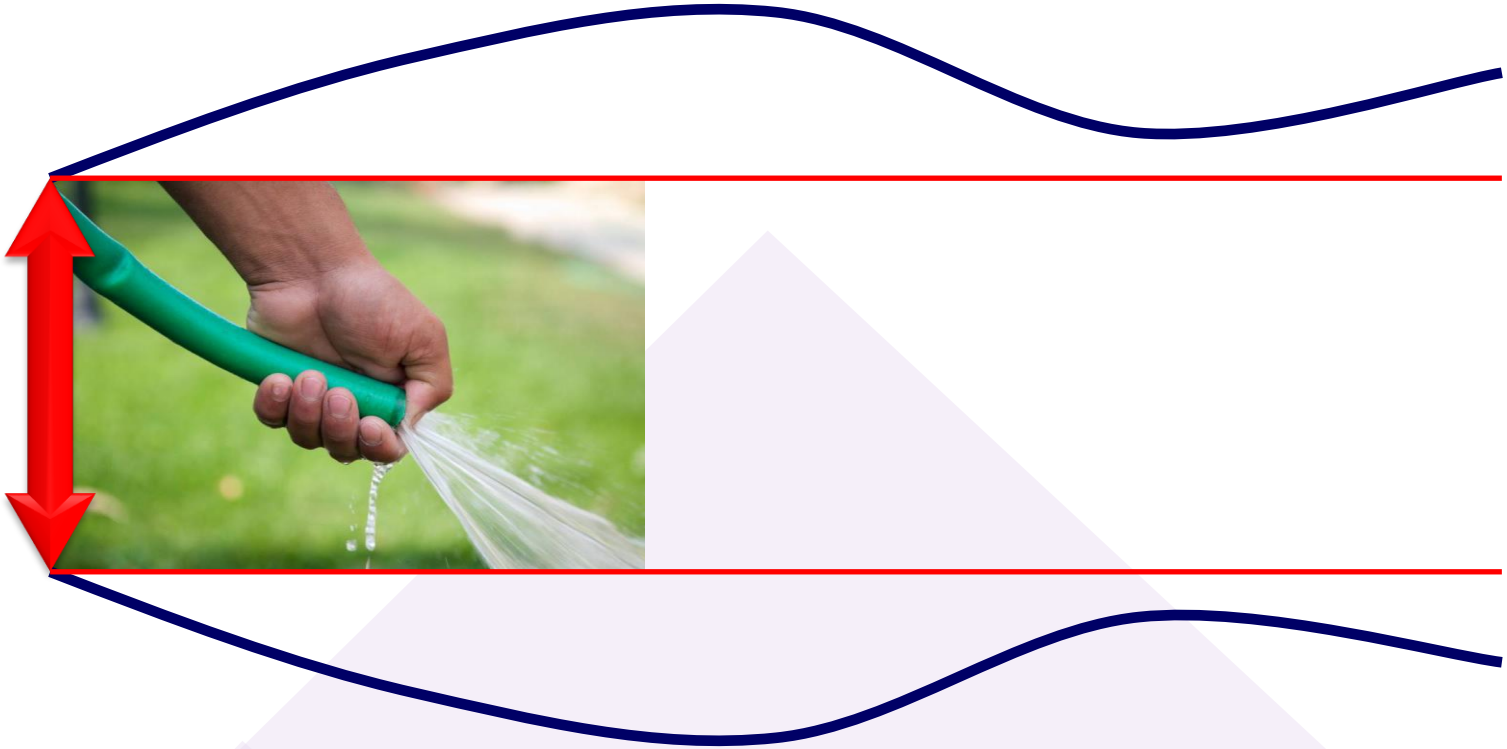


"Each in his own opinion exceeding stiff and strong, though each was partly in the right they were all in the wrong."

Managing constraints

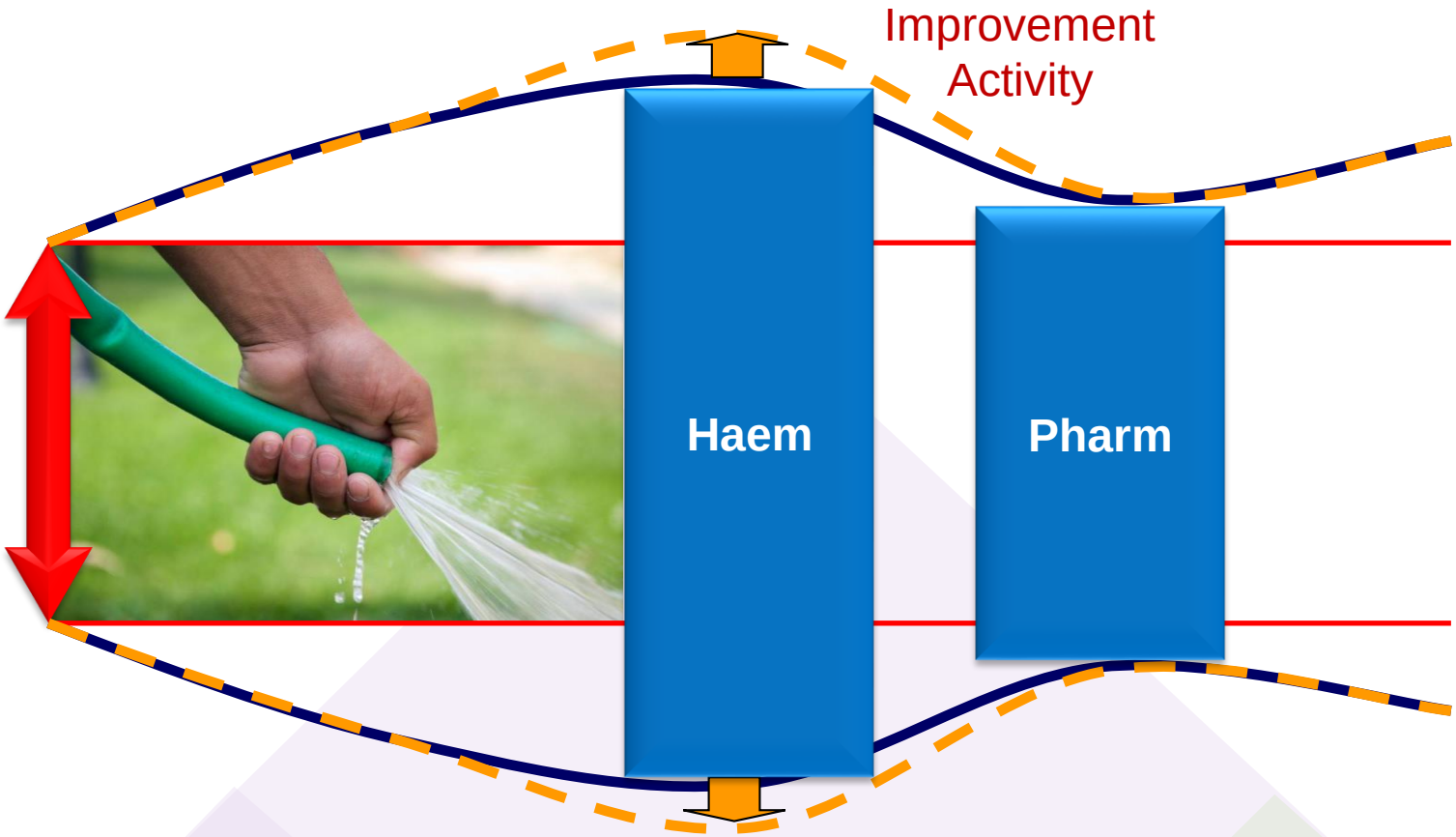


Managing constraints



Area with least capacity is the bottleneck
This constrains the entire end-to-end process:
Think - "*thumb on hosepipe!*"

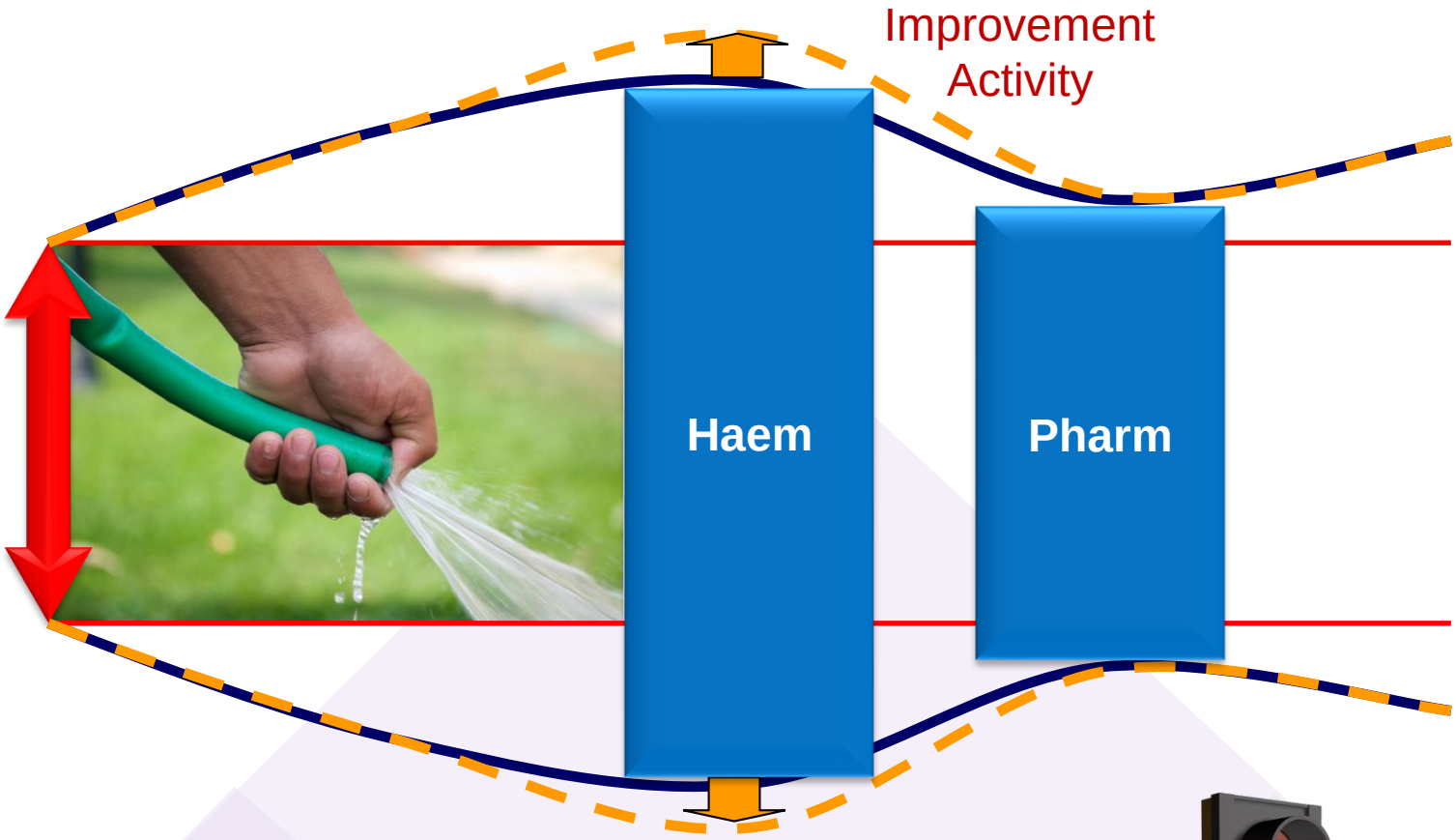
Managing constraints



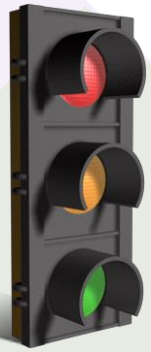
Tackling backlog in non-bottleneck areas

Increased overall flow? Or patients just moving quicker to wait longer at the next backlog?

Managing constraints

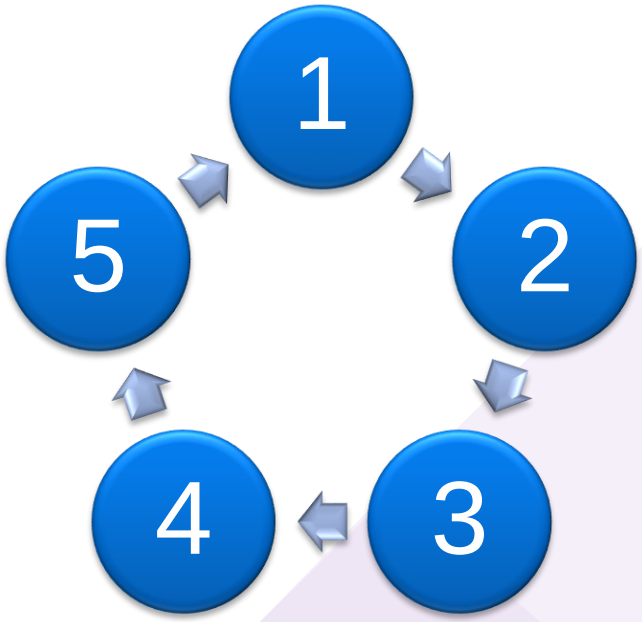


= Rushing to the next red light ?!
= Increase in Pharmacy queue / backlog



Theory of constraints

The theory of constraints identifies a five step process to achieve continuous flow and improve throughput:



1. Identify the system's constraint
2. Get the most out of constraint: exploit it
3. Support system's constraint: subordinate everything else to it
4. Elevate system's constraint
5. Go back to step one, don't allow inertia to become the system constraint

02 Plan ahead along all stages of a patients pathway

eg – the “Enhanced recovery programme*”.

Pre operative stage triggered plans and schedules for patients and staff

Length of stay reduced 12.6 to 6.0 days

Readmissions reduced

* Search “Enhanced recovery programme” on the NHS Institute website

03 Balance demand and capacity

There are two key strategies:

- Look for ways of gaining capacity within the system
- Look for ways of increasing the flexibility of the capacity

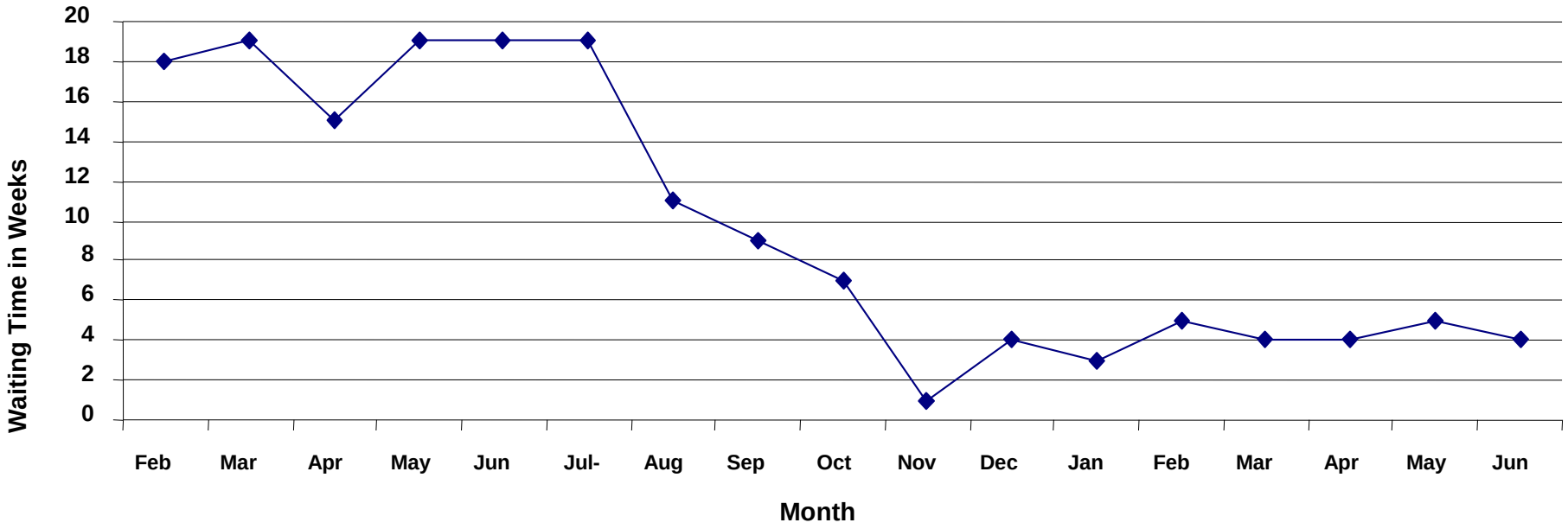
Identify and plan for known changes in available capacity – eg. staff leave, training, equipment maintenance

Maximise capacity by role redesign & releasing time to care

03 Balance demand and capacity

Role Re-design: City Hospitals Sunderland NHS Trust

Barium Enema Waiting List January - June (18 months)



An advanced practitioner role in radiology was introduced and reduced waiting times

03 Balance demand and capacity

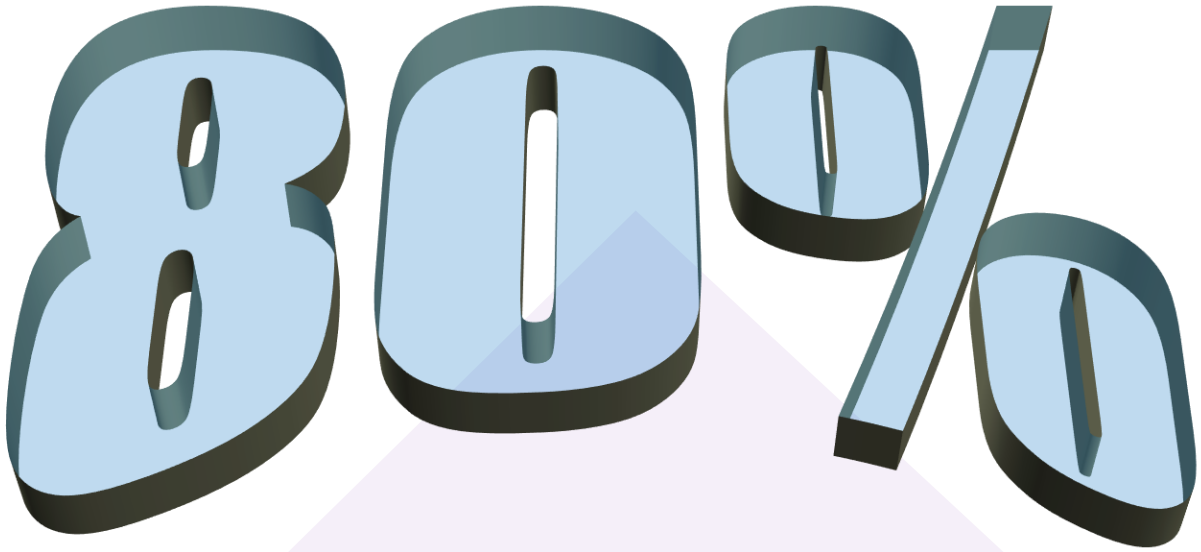
Consider reducing demand:

- What can we do to prevent demand on services?
- Should we see all these patients?
  - Implement protocols
- What can we do to prevent demand on services?
  - Patient education / health promotion?

03 Balance demand and capacity

- Use *process mapping* to find and ease the constraint
- Reduce the number of appointments types complexity and carve out
- Work differently – flexible hours, weekend, pre-plan and cover annual leave, extended roles, etc
- Bid for more resources only when constraint is equipment or staff and working differently will not help
- Reduce DNA's

Theoretical capacity



Set capacity (theoretical) at 80% of the variation in demand to allow for flexibility in the demand

Do not set capacity at the average demand

Practical exercise

Calculating capacity at 80% of the variation in demand

Calculating theoretical capacity - 80% rule

Theoretical capacity = [(max demand – min demand) x 0.8] + Min demand

What is the theoretical capacity of the:

1. Demand between 100 – 120 per week?
2. Demand between 150 – 180 per week?
3. Demand between 50 – 75 per week?

04 Pool similar work together and share staff resources

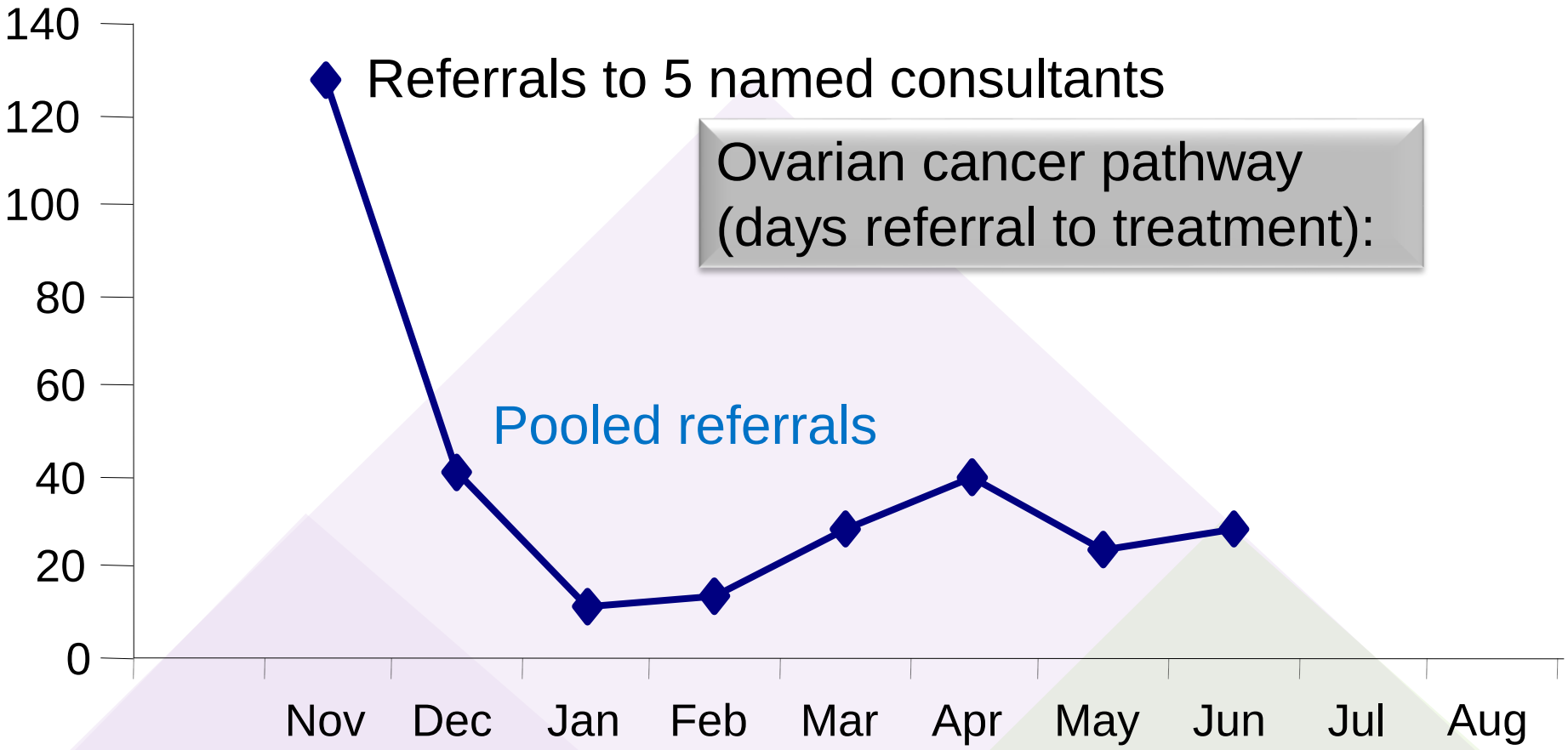
Remember the 73 queues!

Specialists →		Surgeon				Physician					R
Appointment type ↓		1	2	3	4	1	2	3	4	5	1
Flexi Sig	Urgent	◎	◎	◎	◎	◎	◎	◎	◎	◎	
	Soon	◎	◎	◎	◎	◎	◎	◎	◎	◎	
	Routine	◎	◎	◎	◎	◎	◎	◎	◎	◎	
Colonoscopy	Urgent	◎	◎	◎	◎				◎	◎	
	Soon	◎	◎	◎	◎				◎	◎	
	Routine	◎	◎	◎	◎				◎	◎	
OGD	Urgent	◎	◎	◎	◎	◎	◎	◎	◎	◎	
	Soon	◎	◎	◎	◎	◎	◎	◎	◎	◎	
	Routine	◎	◎	◎	◎	◎	◎	◎	◎	◎	
ERCP	--										◎

Pooling solves out “carve out” issues - it reduces both waiting times and number of queues

Remember Pareto:
high volume cases have potential for pooling

04 Pool similar work together and share staff resources

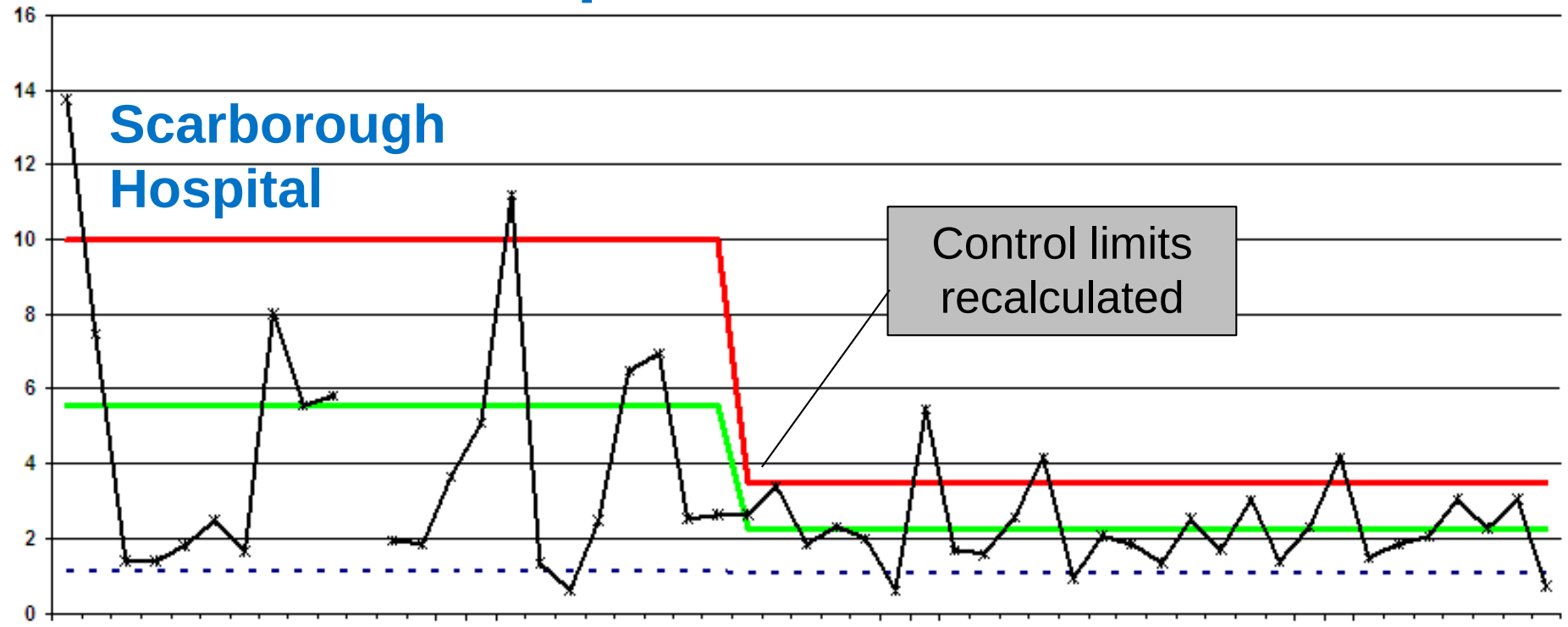


05 Keep things moving – see and treat patients in order

If one person jumps the queue for non-clinical reasons it means everyone else behind them waits longer:

- Blood samples getting trapped at the bottom of the “drop off” bucket until it is emptied
- Seeing patients in the order a report is written rather than the order of referral
- A consultant picks out a case that is interesting and brings it forward
- A GP had 5 different priorities for typing up his patient letters
- One consultant has longer waiting list times than her colleagues
- Patient DNA’s

05 Keep things moving – see and treat patients in order

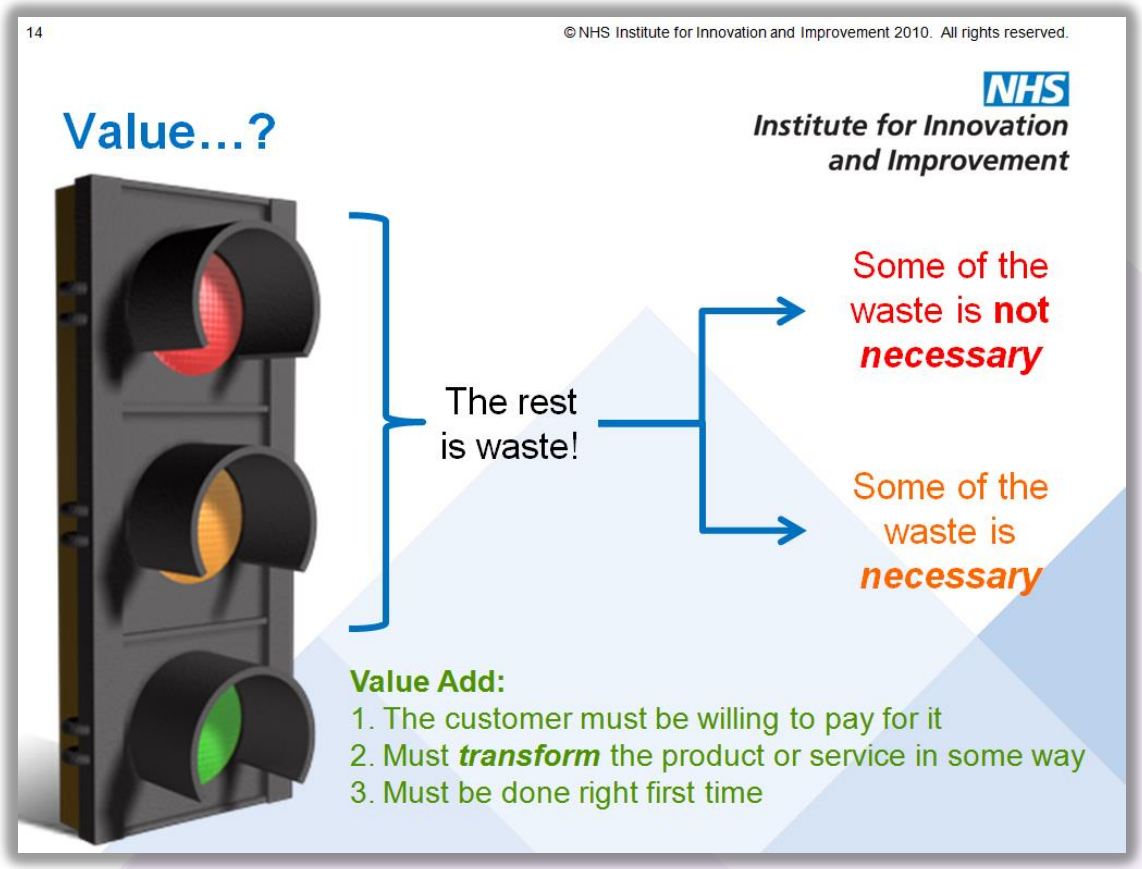


Patients treated by a single consultant – Trans-urethral Resection of Prostate (TURP).

12 months data followed by recalculation of control limits after patient seen in order

Note: 1 data point absent due to a single patient that would be off the scale

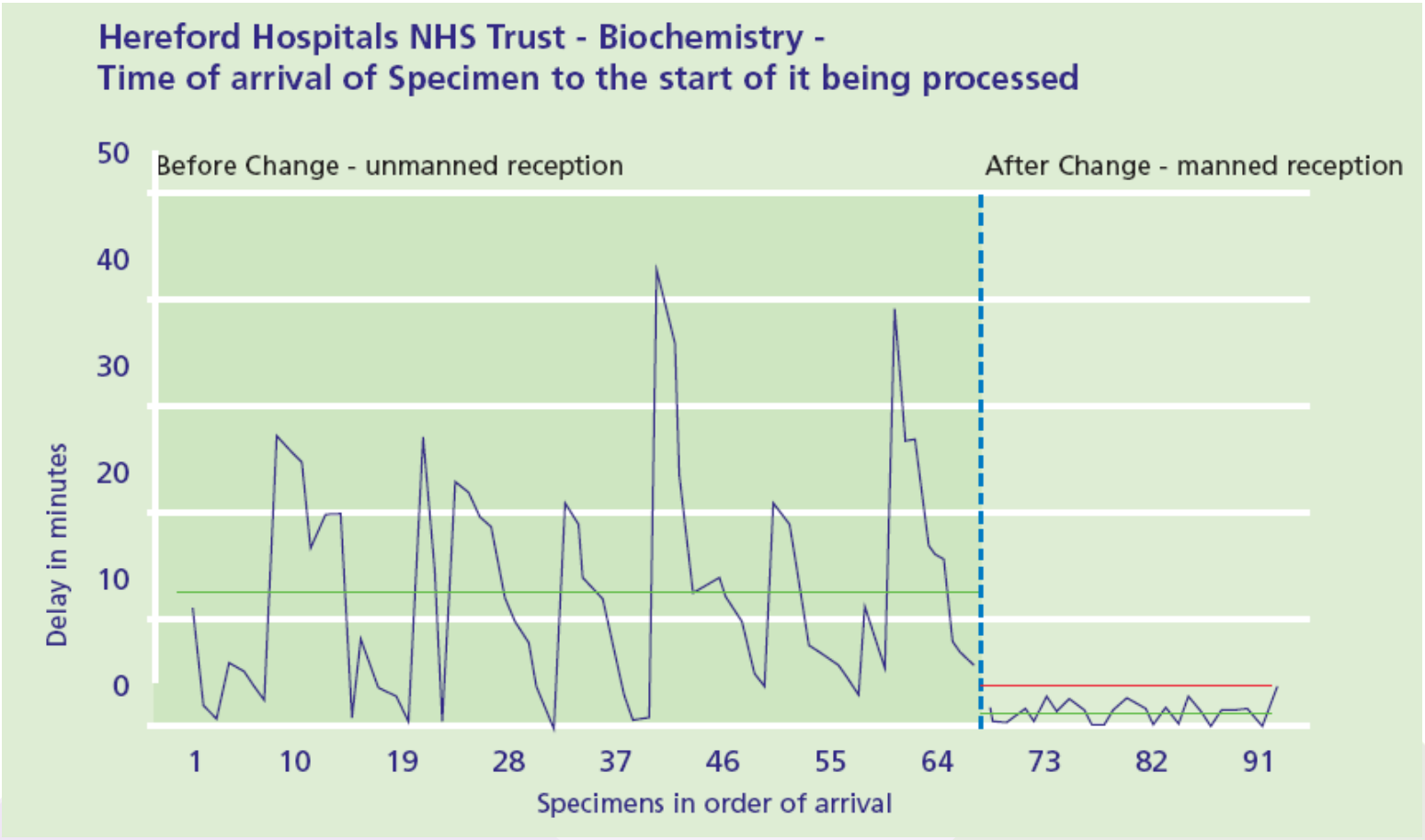
06 Reduce things that do not add value to patients



Remember the value add and waste discussions we had during process mapping?

06

Reduce things that do not add value to patients

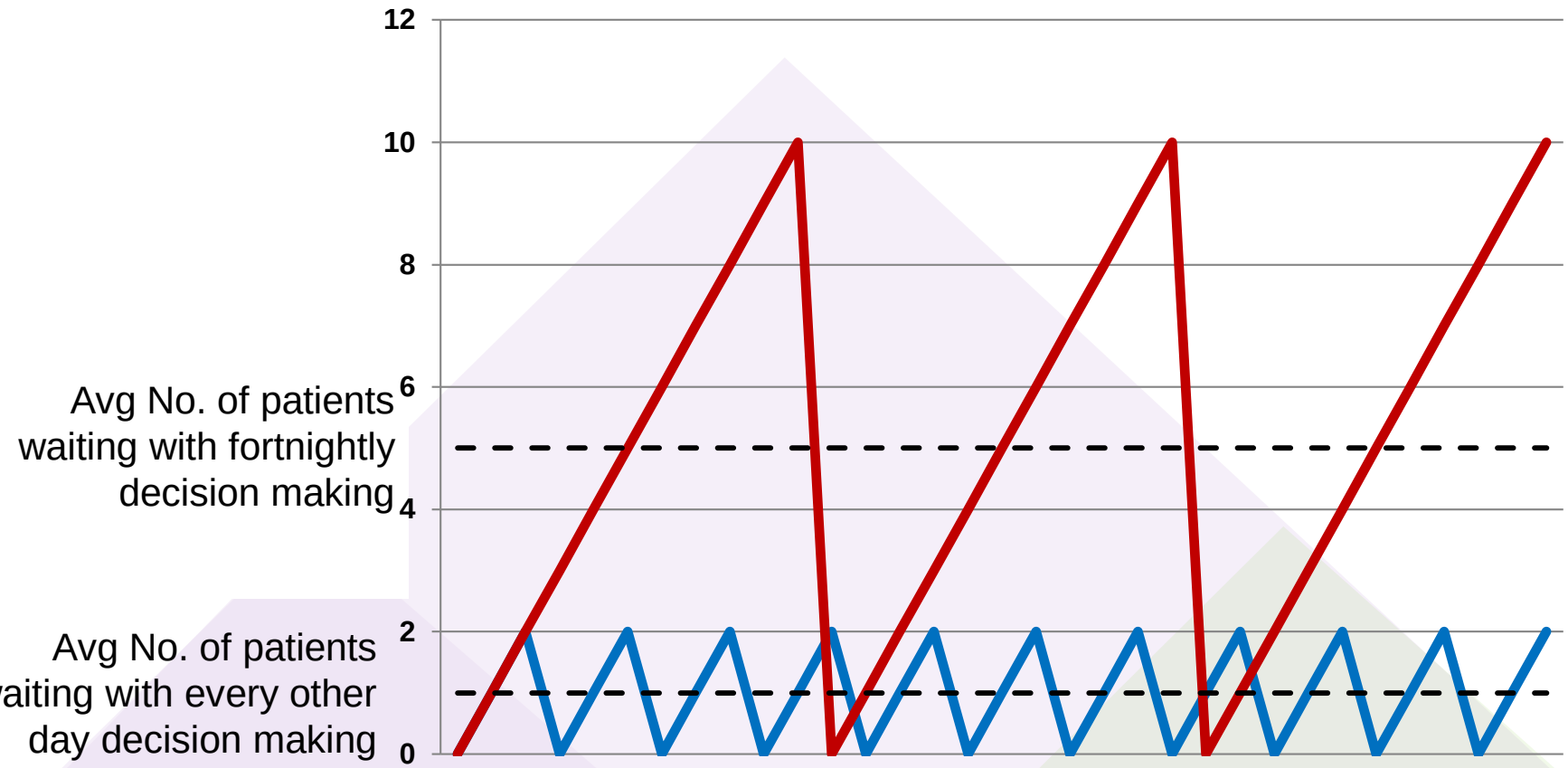


07 Keep the flow – reduce unnecessary waits

The availability / timeliness of *decision making* directly impacts the number of patients in hospital

Every decision, every:	Month
	Fortnight
	Week
	Day
	Hour

07 Keep the flow – reduce unnecessary waits

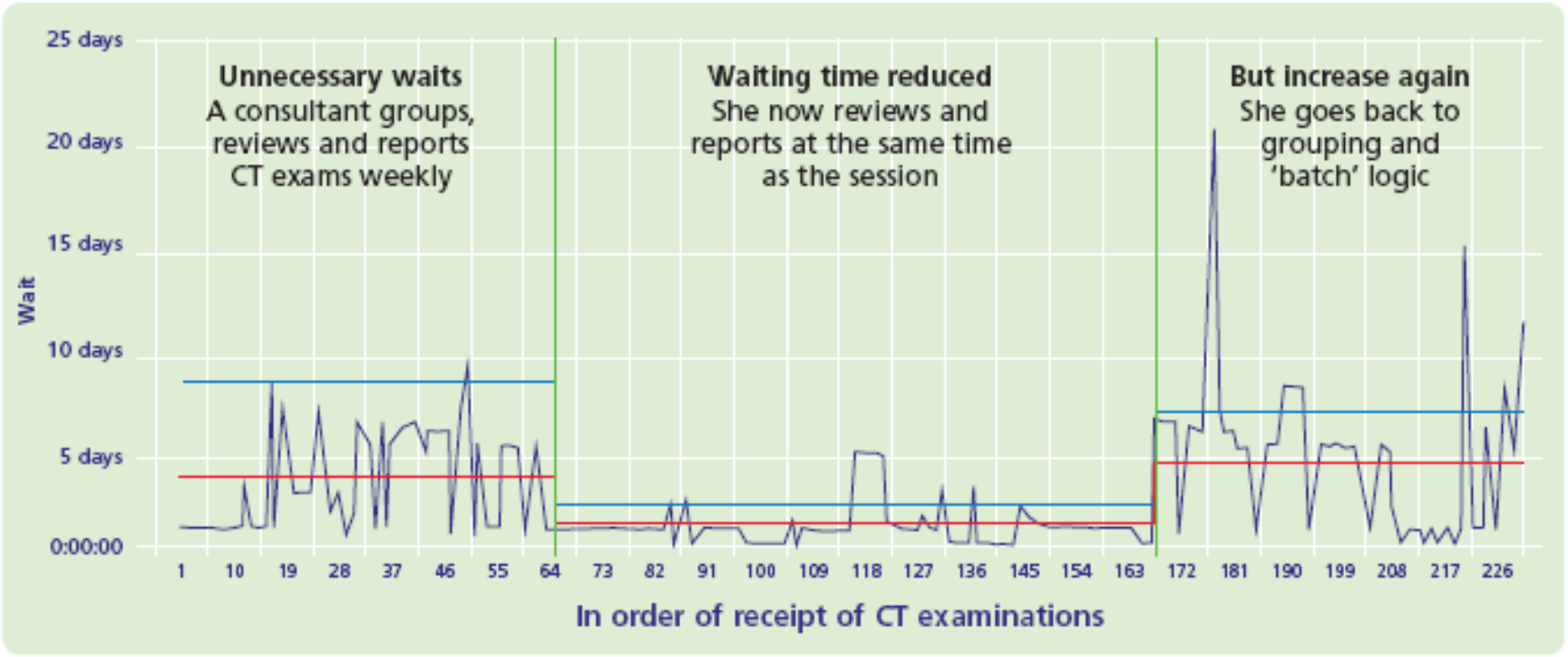


Keep the flow

Reduce piles of paperwork, ensure frequent decision making, reduce batching or batch sizes in diagnostics.

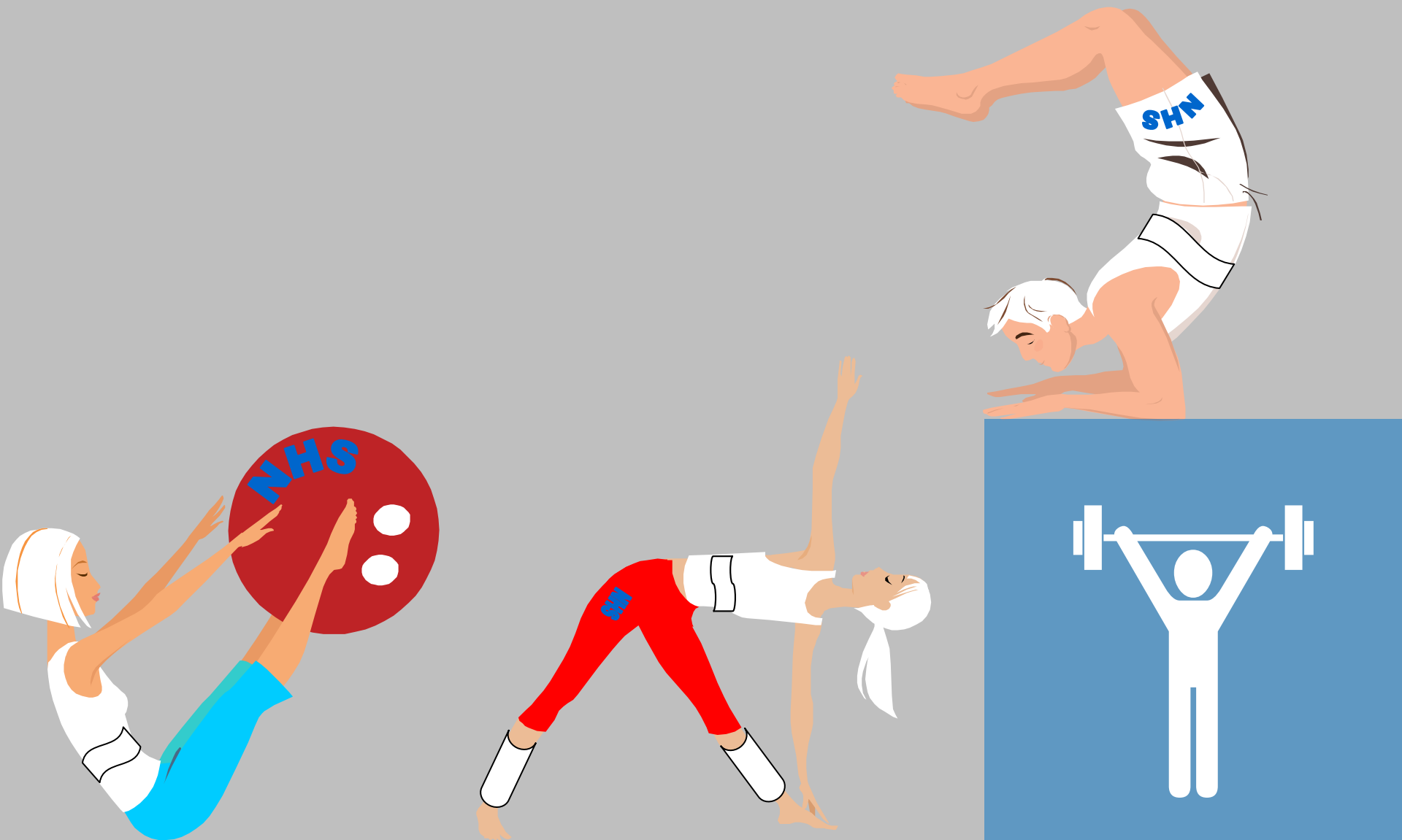
Why?

- Prevents demand amplification
- Reduces waiting times without the need for additional capacity
- More important as waiting times reduce



Any Questions ?

Group Simulation Exercise



Thank You

www.qihc.co.uk