NHS
Institute for Innovation and Improvement

Improvement Leaders' Guide

Matching capacity and demand

Process and systems thinking

Improvement Leaders' Guides

The ideas and advice in these Improvement Leaders' Guides will provide a foundation for all your improvement work:

- Improvement knowledge and skills
- Managing the human dimensions of change
- Building and nurturing an improvement culture
- Working with groups
- Evaluating improvement
- Leading improvement

These Improvement Leaders' Guides will give you the basic tools and techniques:

- Involving patients and carers
- Process mapping, analysis and redesign
- Measurement for improvement
- **Matching capacity and demand**

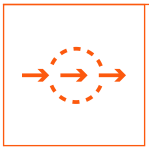
These Improvement Leaders' Guides build on the basic tools and techniques:

- Working in systems
- Redesigning roles
- Improving flow

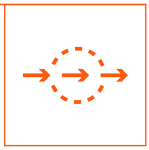
You will find all these Improvement Leaders' Guides at www.institute.nhs.uk/improvementguides

Every single person is enabled, encouraged and capable to work with others to improve their part of the service

Discipline of Improvement in Health and Social Care



Contents



1. Introduction

Improvement of a patient's healthcare journey will not necessarily improve with just more staff, more equipment and more facilities. It has been proved that our valuable resources are not always used wisely and if there is a need for investment, the location of that investment should be carefully considered.

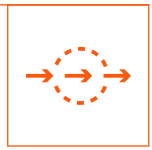
Matching capacity and demand will make some dramatic improvements but you need to start at the beginning:

- map and analyse the process to really understand what happens to the patient
- test out and implement changes that reduce the number of hand-offs and the number of non value added steps across the whole process
- now look very carefully at the process map and identify that stage in the patient journey where they have to queue or are put on a waiting list. This is a bottleneck:
  - map this part of the overall patient process in more detail: to the level of what one person does, in one place, with one piece of equipment, at one time
  - measure at the bottleneck to really understand the capacity and demand problems
- begin to test and implement the relevant change ideas as a result of what the measurement shows you
- create templates of the processes, begin to schedule those templates and watch the whole process improve

This Improvement Leaders' Guide: Matching capacity and demand will help you with all of this as will the other guides in the process and systems thinking group.

As long as we think we already know, we don't bother to rethink the situation

Eliyahu Goldrat



2. Mapping the patient's journey

Matching capacity and demand will make some dramatic improvements but, in order for it to be effective, you need to start by mapping the patient journey and establishing baselines and targets for your objectives.

We strongly recommend that you are familiar with the Improvement Leaders' Guides: Process mapping, analysis and redesign and Measurement for improvement www.institute.nhs.uk/improvementguides. These guides offer help and advice in the setting of aims, how to test change ideas, what and how to measure for improvement and how to present the data to interested parties. They also have a lot more information about the vital first stage, mapping your chosen patient process and analysing it to really understand what is happening.

In summary, you need to:

- define and agree which group of patients is to be mapped (the slice)
- define and agree the first and last step of the stage (the scope), for example from the date of GP referral to the date of discharge from hospital. Be careful not to limit the scope unnecessarily
- identify and involve all staff groups involved within the scope of the stage of the journey being considered
- map the patient's journey and any parallel processes such as making appointments, getting patients' notes and test results from a clinic, arranging an x-ray examination, arranging a pathology test
- identify those steps that do not add value in your process
- identify where the bottlenecks and constraints are in the patient's journey

Bottlenecks and constraints

A bottleneck is any part of the system where patient flow is obstructed causing waits and delays. It interrupts the natural flow and hinders movement along the care pathway. However there is usually something that is the actual cause of the bottleneck and is the constraint. This is usually a skill or piece of equipment.

Examples: Process mapping reveals that patients have to wait to:

- get an appointment at their surgery (bottleneck). The constraint could be the availability of the GP or practice nurse
- get their MRI appointment (bottleneck). The constraint may initially be thought to be the MRI scanner (equipment), but the scanner is there 24 hours each day. Again the constraint is probably the availability of the skill of radiographers and radiologists to operate the scanner

3. Analysing the patient's journey

Having mapped the patient journey, analyse it by considering the following:

- how many steps are there?
- what is the approximate time taken for each step (task time)?
- what is the approximate time between each step (waiting time)?
- what is the approximate time between the first and last step?
- when does the patient have to queue?
- where are the waiting lists?
- how many times is the patient passed from one person to another (hand-off)?
- how many steps add no value for the patient? Imagine that you, your parent or child is the patient, what steps add nothing to the care being received?
- where are there problems for patients? What do patients complain about?
- where are there problems for staff? What do staff complain about?
- ask:
  - is the patient getting the most appropriate care?
  - is the most appropriate person giving the care?
  - is the care being given at the most appropriate time?
  - is the care being given in the ideal place?
- look to see if work or patients are being batched. This is when the work accumulates for hours or even days before it is considered to be enough to attend to. For example, reporting a whole week's x-rays in one go, or allocating appointments for a whole week's referral letters at one time, rather than dealing with them as they come in
- is it someone with 'expert' skills causing the bottleneck? Look to see what the expert is doing. Are they doing what they should be doing, or do they have to do other things that take up their time? Experts include all staff with expertise including medical, nursing, administration and technical staff
- estimate the number of queues (groups of people waiting) at the bottleneck and the amount of time and effort required to manage those queues, as in the diagram below

For example:

Number of appointment types
- emergency
- urgent
- soon
- routine
- follow up

1 2 3 4 5 number of doctors

In one clinic if there are 5 doctors with 5 different appointment types, there are 25 queues to manage.
Two similar clinics per day, five days a week meaning 250 queues to manage each week

Case study
Endoscopy service in the West Midlands

This Endoscopy department started with long and variable waits for patients and the knowledge that they were a key bottleneck in the journey for cancer patients. The team mapped the endoscopy process to understand where the problem areas were for patients, and created what the ideal journey should be like. One of the problems was the way patients were allocated to the ten specialists. This meant that there were 73 different queues to be managed in their department.

Number of specialists

	Surgeons				Physicians					Radio-logist
	1	2	3	4	1	2	3	4	5	1
Flexi. Sig.										
Urgent	x	x	x	x	x	x	x	x	x	
Soon	x	x	x	x	x	x	x	x	x	
Routine	x	x	x	x	x	x	x	x	x	
Colonoscopy										
Urgent	x	x	x	x				x	x	
Soon	x	x	x	x				x	x	
Routine	x	x	x	x				x	x	
O.G.D.										
Urgent	x	x	x	x	x	x	x	x	x	
Soon	x	x	x	x	x	x	x	x	x	
Routine	x	x	x	x	x	x	x	x	x	
ERCP										x

73 queues to manage in an endoscopy

A thought

We stack things everywhere: patients in waiting rooms, laundry on trolleys, referrals on lists, requests forms in piles, emails inboxes etc.

4. Identifying the bottlenecks

4.1 Concentrate on the bottlenecks

You need to:

- identify the steps where there are the longest delays for patients. These are likely to be the bottlenecks
- map that part of the process in more detail to make sure you really understand what is going on. Map to the level of what one person does, in one place, with one piece of equipment, at one time
- look carefully for the true constraint. The constraint is often a lack of availability of a specific skill or piece of equipment. Queues tend to occur before the bottleneck in the patient journey, and clear after the patient has gone past the stage with the constraint
- keep asking 'why' to try to discover the real reason for the delay. For example, if your starting point is 'the clinic always overruns and patients have to wait for a long time', ask why at least five times. Possible responses might be the consultant does not have time to see all their patients in clinic as they have to see everyone who attends including first visit assessments and follow-up patients or it is what has always been done
- keep a look out for other bottlenecks. In the whole patient journey, from visiting the GP to discharge after treatment, it is very likely that there is more than one bottleneck

Case study

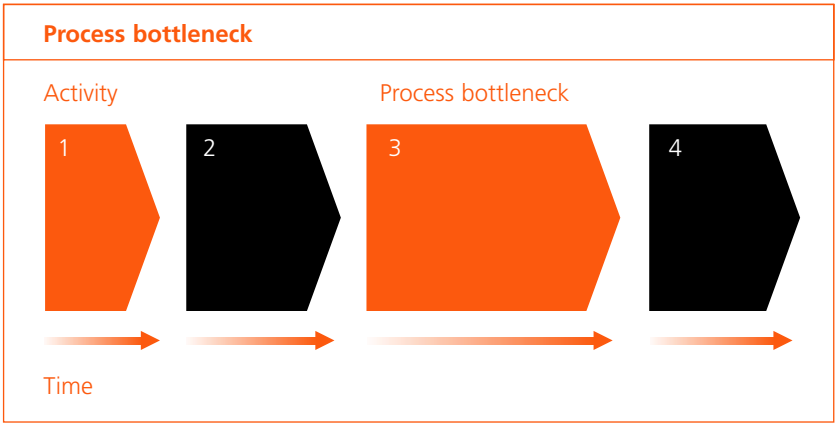
A London Podiatry service took a team approach to address their long waits for routine appointments. The team identified a backlog of patients waiting to receive an appointment, a process that could take up to three weeks.

The administration team suggested change ideas to reduce the unnecessary steps in the administration process. They reduced this stage in the process by ten days which has had a positive impact on the whole patient journey

4.2 Different types of bottlenecks

Bottlenecks are that part of the healthcare system with the smallest capacity relative to the demand on the system. There are two different types of bottlenecks: process bottlenecks and functional bottlenecks.

Process bottlenecks are that stage in a process that takes the longest time to complete. Process bottlenecks are often referred to as the 'rate limiting step or task' in a process. There is more about this in the Improvement Leaders' Guide: Improving flow www.institute.nhs.uk/improvementguides



In the example above, activity 3, is the process bottleneck as it takes the longest time. It could be the consultant seeing the patient in outpatients, or a GP seeing a patient in his surgery.

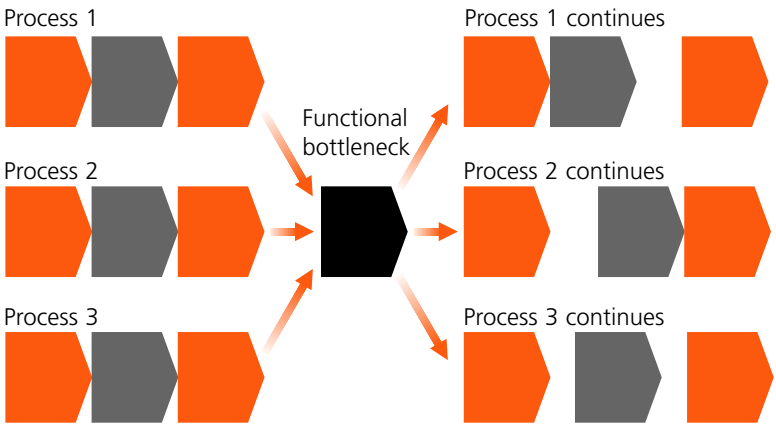
Functional bottlenecks are caused by services that have to cope with demand from several sources. Radiology, pathology, radiotherapy, and physiotherapy are often functional bottlenecks in healthcare processes.

Functional bottlenecks cause waits and delays for patients because:

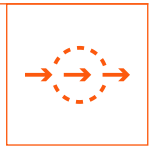
- one process, such as ENT surgery, might share a function, such as imaging with other processes, e.g. orthopaedic surgery, and medicine
- a surgeon may be called to theatre when he is also needed in outpatients
- a GP has to go out on an emergency call when they have patients waiting to be seen in surgery
- a social worker may be torn between representing an existing client in court or doing an assessment of a new patient

This type of bottleneck causes a disruption to the flow of all patient processes. Functional bottlenecks act like a set of traffic lights, stopping the flow of patients in one process while allowing the patients in another process to flow unheeded.

Functional bottleneck



The three processes queue for the functional resource and wait their turn.
The functional bottleneck interrupts the flow of the patients in all three processes.



5. Measuring demand, capacity, backlog and activity at the bottleneck

It is so important to measure demand, capacity, backlog and activity at the bottleneck. The data is really powerful in convincing people to change their practice.

5.1 Demand

Definition

Demand is all the requests and referrals coming in from all sources to the bottleneck step.

Measuring demand at the bottleneck step.

Multiply the number of patients referred by the time in minutes it takes to process (see or treat) a patient at the bottleneck step.

Example

20 referrals x consultation time of 30 minutes each = 600 minutes (10 hours) of demand each day.

Note

Make sure all demand is measured:

- all requests that come in by letter, phone call, fax, email, etc.
- consider hidden demand including those who are not referred but should be. This should be agreed between the referrer and the receiver and is called the referral threshold

> **Golden rule:** Measure demand, capacity, backlog and activity in the same units for the same period of time, for example in one 24-hour period, or over seven days.

5.2 Capacity

Definition

Capacity is the resources available to do the work at the bottleneck step. This includes all equipment and the staff hours available to care for patients.

Measuring capacity at the bottleneck step.

Multiply the number of pieces of equipment by the time in minutes available to the people with the necessary skills to use it at the bottleneck step.

Example

2 treatment machines x 480 minutes (8 hours) of session time = 960 minutes (16 hours) of capacity each day.

Note
Capacity can then be converted into the number of patients that could be seen. So if a patient takes 20 minutes to process, then the capacity is 960/20, that is 48 patients.

Ensure that you measure all available capacity. Many people do lots of different things, so you need to make sure you measure any hidden capacity. For example, if you wanted to calculate the capacity for a pharmacist dispensing drugs or preparing chemical substances, you would need to know all the activities they currently do and understand the proportion of their time devoted to each task.

5.3 Backlog

Definition
Backlog is the previous demand that has not yet been dealt with, showing itself as a queue or waiting list. In a community setting, this includes all patients waiting to be assessed by the occupational therapist. Including those on waiting lists and new patients in the system.
Measuring backlog at the bottleneck step.
Multiply the number of patients waiting by the time in minutes it will take to process a patient through the bottleneck step.
Example
100 patients on the waiting list x 30 minute treatment time each = 3,000 minutes (500 hours) backlog.
Note
Take care when measuring the backlog and ensure that you don't count the same patient more than once. There may be patients who have been put on waiting lists at different parts of the same process, e.g. patients requiring radiotherapy treatment can be waiting in queues for their pre-treatment, planning, and simulation at the same time. Only count them in the earliest queue to avoid recounting them at a later stage in the process. In the radiotherapy example given, it will be at the planning stage.

5.4 Activity

Definition

Activity is all the work done at the bottleneck step. It is the actual work carried out by staff including the time spent with patients, carers and liaising with colleagues.

Measuring activity

Multiply the number of patients processed through the bottleneck by the time in minutes it took to process each patient.

Example

100 patients processed x 15 minutes each = 1,500 minutes (250 hours) of work done each day.

Warning: Measures of activity numbers are misleading as this does not necessarily reflect demand or capacity:

- the activity in the month of June may well include demand carried over from May, April or even March
- staff may have not been fully utilised. They may have been kept waiting for the patient, specialised pieces of equipment or test results

Note

You can measure demand, capacity, backlog and activity either in patient numbers or in minutes. If patients take different times to process, then it is often easier to calculate everything in minutes or hours.

Estimating backlog for a CT scanner

Backlog

Requests and time taken for procedure	Backlog in minutes (hours)
524 head scans @ 30 minutes	524 x 30 minutes = 15,720 minutes (262 hours)
129 limb scans @ 20 minutes	129 x 20 minutes = 2,580 minutes (43 hours)
356 chest scans @ 15 minutes	356 x 15 minutes = 5,340 minutes (89 hours)

Total backlog = 1,009 patients on the waiting list who would take 23,640 minutes (394 hours) to process

In this example, the patients take different times to process, so the demand and capacity have been calculated in minutes (hours) of work required.

5.5 Working out the flow through the bottleneck

If the demand is 6 patients / hour and the capacity is 4 patients / hour then, the activity will be 4 patients / hour and the backlog will grow by 2 patients every hour.

Example of demand and capacity for a CT scanner measured over seven days.	
Weekly demand	
Requests and time taken for procedure	Requests per week x time taken for procedure
20 head scans @ 30 minutes each	20 x 30 = 600 minutes (10 hours)
18 limb scans @ 20 minutes each	18 X 20 = 360 minutes (6 hours)
4 chest scans @ 15 minutes each	4 X 15 = 60 minutes (1 hour)
Total demand in minutes	**1,020 minutes (17 hours)**
Weekly capacity	
Equipment and staff available	Equipment x amount of time people with the necessary skills are available to use it per week
Monday morning: 1 CT scanner and 1 radiologist for 240 minutes	1 x 240 = 240 minutes (4 hours)
Wednesday all day: 1 CT scanner and 1 radiologist for 480 minutes	1 x 480 = 480 minutes (8 hours)
Friday afternoon: 1 CT scanner and 1 radiologist for 240 minutes	1 x 240 = 240 minutes (4 hours)
Total capacity in minutes	**960 minutes (16 hours)**
In this example, demand exceeds capacity by 60 minutes (1 hour) each week	

5.6 Identifying queues

Queues occur where demand has not been dealt with and results in a backlog. The main reasons why queues occur is because:

- demand exceeds the available capacity. For example, the number of patients referred to the local dentist is greater than the available resources in terms of qualified staff or appropriate equipment
- there is a mismatch between variation in demand and capacity at specific times, because the right people or equipment are not always available to deal with the demand in a timely manner
- patients are not always discharged to accommodate admissions

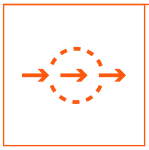
Every time the demand exceeds the capacity, the queue is carried forward to the following day. However every time the capacity exceeds the demand, the extra capacity is lost in the fixed session, or it is filled from the queue. So plans based on matching the average daily demand to the average daily capacity are fundamentally flawed: they guarantee the very queue they are trying to eliminate.

There are two ways of dealing with this problem:

- planning extra capacity into the system to keep the queue under control, however this may not be affordable
- reducing the variation mismatches in the system:
  - manage the capacity to meet the peaks and troughs in demand
  - reduce the peaks and troughs in the demand, but especially the capacity, by understanding what is causing the variation and eliminating it

There is more on queues and variation in the Improvement Leaders' Guide: Improving flow www.institute.nhs.uk/improvementguides

Beware: There are inappropriate measures of performance or perverse incentives in the system. A queue allows a resource to appear busy, important and in need of more resources.



6. Redesign to match demand and capacity

In order to make the most of patient flow through a healthcare system, it is necessary to address the entire patient process. You need to analyse and understand capacity, demand, backlog and activity wherever there are queues or waiting lists. This will help you to understand the impact of changes to patient process at previous stages (upstream) and at future stages (downstream).

6.1 Use measurement to predict and manage

Once you have started to measure demand, capacity, backlog and activity in the same units over the same amounts of time, you can use the data and the patterns that emerge to start predicting and managing the capacity, demand, activity and backlog at the bottleneck.

Change ideas

Look carefully at patterns for:

- variation in referral protocols and referral thresholds
- daily, weekly, monthly, and seasonal variations in demand
- types of requests and who is making the request
- who receives the requests and what they do with them
- any patterns of did not attends (DNAs) or cancellations
- look carefully at the backlog patterns. Where are the queues? If the backlog numbers remain constant over time, then demand and capacity are equal, however the waiting times might not be

6.2 Manage the bottleneck in the patient flow

The bottleneck determines the pace at which the whole of the healthcare process can work. If changes are made to improve parts of the care process without addressing the bottleneck, improvement initiatives are unlikely to succeed. When you have identified the bottleneck, look for changes to maximise the work of the bottleneck and/or to drive work away from it.

Change ideas

Ensure that the bottleneck, whoever or whatever it is, has no idle time:

- if the bottleneck is the CT scanner, make sure the next patient is prepared and waiting at all times
- schedule routine maintenance of specialised equipment, e.g. MR scanners and radiotherapy machines, for weekends or evenings
- ring patients up and ask them if they still intend to come

Put an inspection or checking stage in front of the bottleneck:

- if the bottleneck is the doctor in the clinic or surgery, check that all test results are available before each patient goes in to see the doctor

If the bottleneck is the expert skill, they should only be doing work for which their expertise is needed:

- doctors in a clinic or nurses on a ward should not be using their time chasing notes or test results
- separate the responsibilities for patient flow and paper flow. Clinicians and clerks/administrative staff should work as a team but have clear and different responsibilities

Don't make the bottleneck an 'inspector'. Someone else in the team should do any inspection of the patient or paperwork prior to the bottleneck:

- arrange pre-assessment of patients and their fitness to proceed before surgery

Consider if someone can help free up an individual who is currently overloaded:

- this is usually done by sharing skills in the team, e.g. development of a nurse specialist role or team coordinator
- consider if the patient can provide their own care, e.g. putting in ear drops prior to an aural examination
- re-think follow up appointments: is the follow up visit really necessary or can they be seen by a nurse specialist or another specialist in the community

Sharing referrals amongst clinicians, known as team referrals or pooling, can reduce waiting times by:

- distributing the work amongst the clinical team
- everyone working to their highest level of skill and expertise
- each having a similar waiting list/queue

Golden rule:
There are only two ways to make improvements at a bottleneck: make changes to reduce demand or make changes to increase capacity.

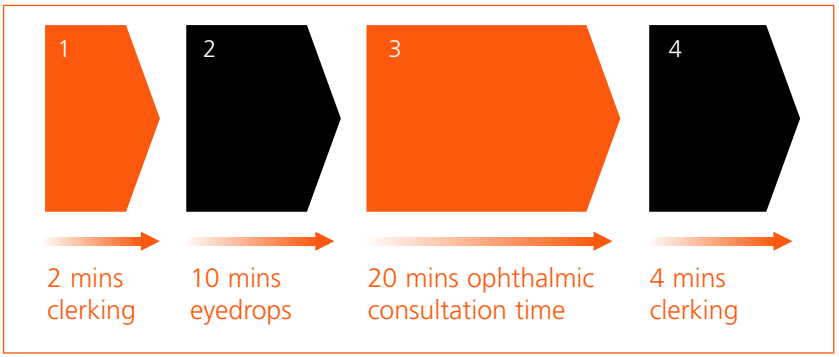
6.3 Resolve capacity problems at the appropriate point in the system

Often changes are introduced in the wrong place. Focusing on and speeding up the beginning of the patient journey will cause a build-up of patients further along the process at the site of the constraint:

- getting orthopaedic patients referred more efficiently to physiotherapy will not speed up the whole patient journey if the physiotherapy department doesn't have sufficient capacity to cope with demand
- extending the car parking or waiting area at a hospital or surgery may be of some benefit but long patient waits will still happen when the actual problem is a bottleneck further on in the system caused by the nurse, doctor, or a piece of equipment

Change ideas

- increase capacity at the stage of the process where it will create the greatest outcome. For example, the ophthalmic process shown below has four steps and takes a total of 36 minutes not including the delays between steps. There is little point in increasing the capacity at task 1, the initial clerking, because it will create a wait for tasks 2 and 3. However, if the capacity is doubled at task 3, the consultation, it will improve the throughput of the whole process
- increase capacity of the bottleneck by moving resources from previous (upstream) steps or future (downstream) steps of the process
- reduce inappropriate demand to the constraint effectively by:
  - evaluating the clinical merit of current procedures
  - educating the referring clinician to use referral pro-formas and agreed referral thresholds
  - ensuring the referrals match the expertise of the constraint
  - providing feedback on inappropriate referrals



Case study
Dermatology service in the Midlands

Traditionally, all patients had been referred to the consultant dermatologist for an opinion. Following process mapping and analysis, many changes have been introduced, including nurse led clinics for children with eczema and psoriasis, and patients with viral warts. This has freed 25 consultant slots each month.

6.4 Reduce all unnecessary waits and delays

Waits and delays are not inevitable features of healthcare services. They are symptoms of systems that are poorly designed. They cause increased anxiety to patients and staff as well as increasing overall costs.

Change ideas

- synchronise the first morning and first afternoon appointment. Make sure that the patient, the clinician(s), the necessary equipment, and the paperwork are all in the room at the same time, ready for a prompt start
- reduce hand-offs. This is when we hand responsibility for the patient from one person, department or organisation to another, which often causes waits, delays and mistakes
- instead of a referral being received, inspected and dealt with by three different people, develop an extended administration role to deal with all three stages
- do all or some of the tasks simultaneously. Many healthcare processes are designed so that tasks are carried out in a step-by-step sequence. The second task in the process is not begun until the first task is completed. Begin to plan discharge as soon as a patient is admitted by referring them to the appropriate professionals soon after they are admitted onto the ward
- reduce or eliminate batching. Do work when it arrives rather than waiting to deal with a whole set of similar tasks at the same time, e.g. report x-rays as they are done, or make appointments as they come in rather than them piling up for a day or even a week
- move the physical location of adjacent steps in a patient process closer together so that work can be passed directly from one step to the next. Develop a 'near patient' test centre close to the outpatient department where the more common and straightforward tests, e.g. blood tests, ECGs etc., can be done at the same time and results can be easily transported back into the clinic

- develop **pull** systems instead of pushing the patient and other work along the process. In a **push** system, transferring patients from one step of the process to the next is the responsibility of the earlier part of the process. They will push the patient to the next stage. For instance, GPs push urgent referrals to cancer units. Cancer units 'push' patients requiring specialist radiotherapy to cancer centres. Most healthcare organisations and systems operate push systems. The trouble is that patient flow stops when it reaches a bottleneck where queues and waiting lists (backlog) build up.

  In a **pull** system, the bottleneck governs the rate that patients flow through the whole process. In this system it is the responsibility of the later parts of the process to pull patients towards them by asking for the work when they have the capacity to do it. Pull systems are particularly effective when patients are transferred from one care setting to another.

Example
One healthcare organisation introduced a Hospital at Home Scheme based in the community where staff went into their local orthopaedic wards and pulled eligible patients out into the community, enabling the wards to pull other patients into empty beds

6.5 Eliminate backlogs

The aim is to avoid having every stage in the patient journey so busy that there is no room for flexibility. When the staff and resources in general are so busy with work that was created in the past, they cannot respond to today's requests as they are doing last month's work today.

When huge backlogs (waiting lists or queues) accumulate, they take a lot of effort to manage and often create more work, including dealing with complaints and having to re-schedule appointments.

In order to work down backlogs, a conscious and intentional plan is required. This must include:

- an understanding of the true extent of the backlog. Establishing if the waiting list is constant over time or growing. You will need to measure all waiting lists using the same units used to measure demand and capacity
- a plan to add capacity on a temporary basis. However, this should be done over time with careful attention to other parts of the system. A massive initiative addressing the backlog in one part of the system will cause a tidal wave of patients arriving at later stages. For example, a big initiative on surgical outpatients will increase the waiting lists for surgical treatment

- ways to introduce more flexibility and consider new ways of working. One radiotherapy department introduced seven day working with flexible hours to accommodate the needs of the patients and staff
- commitment by senior clinical and managerial leaders across departmental and organisational boundaries

The overall goal is to reduce demand appropriately, effectively and permanently and increase capacity accordingly. The elimination of backlogs should be an early stage in the improvement initiative. This may require additional resources in the short term with strategies addressing both demand and capacity.

However, once the backlog is eliminated, the organisation should work hard to maintain that position. It is only at the stage when backlogs have been eliminated at specific points in the whole healthcare system that capacity and demand can be matched and true gains in the overall system made.

Change ideas

- increase the number of procedures undertaken to more than current demand. If there is already a waiting list (backlog) for a test and 26 new patients are being referred each week, then performing more than 26 tests a week will start to clear the backlog and reduce the waiting list
- create centralised resources. Employ or train someone who has several clinical skills. They can then add capacity in different areas, as required
- introduce weekend shift work and extended hours in departments that normally work 9-5 Monday to Friday

Golden rule:
Every time demand exceeds capacity, you carry forward the excess demand as backlog. But you cannot carry unused capacity forward to the following day or next week. As a consequence, if you plan the average capacity to equal the average demand you will always end up with a queue. Therefore plan the average capacity to be slightly greater than the average demand so that this queue can be eliminated quickly

Case study
ENT service in the Midlands

The team looked at the clinic booking rules for each of the ENT consultants and measured the demand, capacity and activity over an eight week period. The data helped convince the clinicians to change their booking rules. This has resulted in finding many additional new patient slots and reduced waiting time.

6.6 Match capacity and demand on a daily basis

Once the backlog has been eliminated, the next aim is to ensure that demand and capacity are in equilibrium and that the backlog has stabilised. This requires that demand and capacity is 'matched' on a daily basis.

Clinical teams should consider patient demand issues and plan capacity to meet them on a regular basis such as each day or each week. This way, the team can avoid backlogs of work building up.

Change ideas

- ensure that the patient schedule in the surgery, in the clinic, in the theatre, etc. reflects what is really happening
- use process templates to plan and schedule work accurately (section 7)
- make pauses in the schedule to catch up
- plan average capacity at 80–85% of the normal fluctuation in demand. This ensures that queues and waiting lists rarely build up and that there is the flexibility to cope with unexpected demand instantly
- determine the minimum number of clinicians required, and do not fall below that level. Consider if sessional work is still appropriate
- develop and agree ways to meet the unexpected and the expected situations that occur
  - add more appointment slots or clinicians as needed when demand goes up unexpectedly
  - plan for the increased demand on Mondays, for example in primary care and in the fracture clinic
  - plan additional clinics or sessions to compensate for lost Bank Holiday Mondays
  - plan clinics for each day to eliminate batching and sessional working patterns

More change ideas
You can find a lot more information and examples of improvement in a variety of settings in the process and systems thinking group in the Improvement Leaders' Guides, especially Improving flow
www.institute.nhs.uk/improvementguides

Case study
Echocardiography service in London

As test requests came in, they were batched for sorting each day. Outpatients were advised of their appointment date in writing. The waiting time was 130 days, there was a high non-attendance rate and, after tests were done, there was often a further delay before the requesting clinician received a result.

The team got together and tested a variety of ideas. The changes introduced included a new appointments system to 'deal with today's work today' and a new staff rota which staggers break times to enable the machines to run throughout the day, instead of closing over lunch.

Now results are reported within 24 hours, outpatient waits have been reduced to 7 days, and inpatient waiting times have been virtually eliminated.

7. Process templates to match capacity and demand

A process template describes the process in terms of what happens to one patient at one point in time. The following example shows the development of a process template for endoscopy and how it can be used to schedule resources.

Example of a process template

Step 1: map the process and allocate the time it takes in minutes to complete each step, as shown in the table.

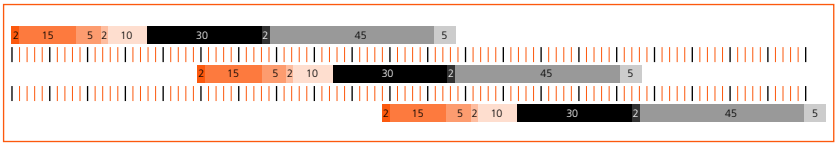
Step	Time
Clerk in (reception)	2
Clerk in (nursing)	15
Patient gets changed	5
Pre observations	2
Consent	10
Procedure	30
Post observations	2
Type up report	10
Patient in recovery	45
Discharge	5

Step 2: allocate a colour to each step

Step	Time	Code
Clerk in (reception)	2	
Clerk in (nursing)	15	
Patient gets changed	5	
Pre observations	2	
Consent	10	
Procedure	30	
Post observations	2	
Type up report	10	
Patient in recovery	45	
Discharge	5	

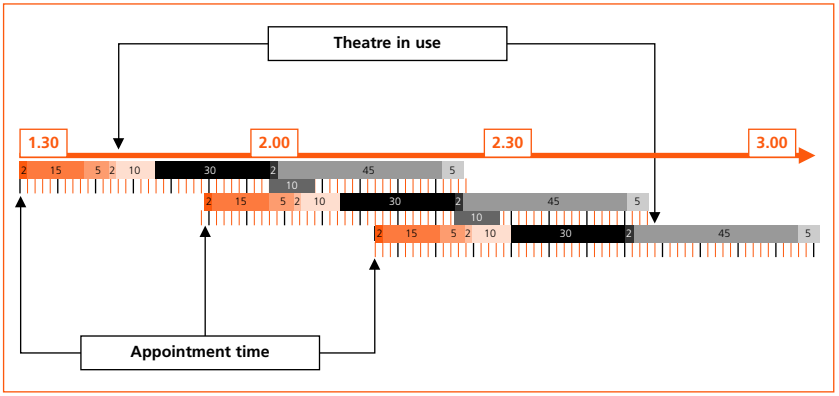
Step 3: line up the colour steps in sequence in blocks of colour proportional to the time scale where each block represents two minutes of time. By doing this you can see that (in this example) there are two process bottlenecks for the patient, i.e. take the longest time.

Step 4: line up several templates so that bottleneck(s) are fully utilised, i.e. there is only ever one patient having the procedure and one in recovery.



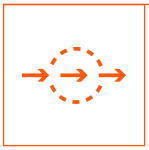
NOTE: In steps 3 and 4 'type up report' is not part of the patient process shown

Step 5: position on time line, to determine patient appointment time, optimum theatre usage and list composition.



Step 6: use the process template to schedule resources and staff for the required number of procedures

Source: Cancer Services Improvement Partnership (CSIP)



8. Beware the dangers of variation

8.1 Carve out and segmentation

Often we have dealt with healthcare problems by prioritising, ring fencing or **carving out** the time of an expert, the time of specialised equipment or by keeping resources or facilities only for one particular group of patients. By carving out in this way, the process of care for one group of patients is prioritised over another irrespective of their needs.

For example, a GP practice might give priority to all pregnant women with diabetes and offer them urgent appointments whereas, another diabetic patient will have to wait. Prioritising in this way, or carving out capacity for one group of patients, interrupts the flow for other patients who inevitably end up waiting longer.

Accurate measuring of the backlog or waiting time for other groups of patients has shown that carving out capacity significantly increases waiting times overall and creates a very difficult system to manage effectively.

Segmentation is about the separation of the whole process of care for one group of patients but not at the expense of other patients.

There is more about carve out in the Improvement Leaders' Guide: Improving flow www.institute.nhs.uk/improvementguides

The difference between segmentation and **carve out**		
	Segmentation	Carve out
Objective	• to improve the flow for all patients	• to improve the flow for a specific group of patients at one bottleneck
Principles	• looks at the whole patient process • groups patients with similar processes • keeps the flow through the process bottleneck constant • matches demand to capacity along the process	• looks at one bottleneck, e.g. CT, theatre, outpatients • prioritises the queue irrespective of the patient need • interrupts the flow of patients and keeps them waiting at all steps in the process
Effect on waiting time	• reduces/eliminates waiting for patients	• makes the waiting time worse
Effect on other patient groups	• none	• other patients, e.g. with non urgent chronic diseases, have a much longer wait
Examples	• opticians diagnose, prepare and add cataract patients directly onto the surgical list. This removes them from other ophthalmic patients in the outpatient clinic	• reserving slots on a CT scanner for certain groups of patients, e.g. urology • ring fencing beds for certain groups of patients • reserving clinic/surgery slots for specific groups of patients, e.g. urgent, soon, routine • creating different queues for different consultants even when the process is the same, e.g. endoscopy
More examples - non healthcare	• cash dispensers where customers are totally removed from other customers at the counter in the bank	• bus lanes at peak times on busy roads • parking for mother and babies in supermarket car parks • first and second class post
Summary definition	• **segmentation** is when the separation of the process of care along the whole pathway for one group of patients is **not at the expense** of other groups of patients	• **carve out** is when the flow of one group of patients is improved **at one bottleneck at the expense** of another group of patients

Case study
Segmentation in Accident & Emergency (A&E)

Previously, A&E was managed as a carve out in which patient were triaged by a nurse into one of five categories. Categories one and two were prioritised and patients with minor injuries, in categories four and five, were told to wait.

The A&E department has now changed, replacing carve out with segmentation. Apart from the real emergencies, all patients are put into a single queue at the reception desk and the next available member of staff takes the next patient through to the appropriate treatment area: minors or majors.

The minor area is fitted with chairs in cubicles and trolleys with all the dressings required to treat any minor injury. The major area is fitted with trolleys and the equipment required to diagnose and treat an emergency, e.g. ECG machines, blood bottles, etc.

Since the majority of A&E attendances are minor injury patients, there is a very fast flow through the minors areas with the major areas moving at a slower rate. If the demand and capacity is matched by ensuring the maximum number of staff are available during the times of peak demand, 12.00 noon to 2.30 pm and again from 4 to 6 pm, then the waiting room is empty.

8.2 Understanding variation

Variation and Statistical Process Control (SPC)

Variation is a part of everyday life and occurs naturally in most processes. In healthcare, there is huge variation in the patient population from their clinical condition, to the level of support needed and their social background. Managing such variation is a real challenge for healthcare professionals who often make the situation worse in the way they organise processes within the system.

In healthcare measuring, understanding and reducing variation is key to improving patient flow. Variation and statistical process control (SPC), an effective way to measure variation, are considered in more detail in the Improvement Leaders' Guides: Measurement for Improvement and Improving flow www.institute.nhs.uk/improvementguides. For a practical demonstration of variation, demand and capacity visit www.steyn.org.uk

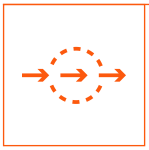
Variation and care pathways

The development, agreement, use and monitoring of care pathways based on sound national guidelines and protocols are key to reducing variation. Care pathways express locally agreed multidisciplinary practice. They are based on guidelines and evidence for a specific patient group. They form all, or part of the clinical record, document the care given, and help in the evaluation of outcomes.

Care pathways include:

- a clear description of the ideal patient journey
- what is to happen, where, and by whom
- clear, specific and measurable goals for each step of the patient journey based on good clinical evidence
- notes on patient care

There should also be space to record variations from the planned procedures. It is in the examination and actions taken to reduce or eliminate variation that the improvements will be made. Therefore, any variation from the desired plan should be clearly stated and the cause identified. Variations from the plan should be discussed by the team and actions agreed as part of a continuous quality improvement programme.



9. Activities

Before organising any activity, consider the following:
- who is the audience?
- what is their prior knowledge?
- how will they use the information provided?
- is the location and timing of the activity correct?
- recognise and value that participants will want to work and learn in different ways. Try to provide information and activities to suit all learning preferences

Why is this important?
Some of us take to the idea of change more easily than others. Some like to develop ideas though activities and discussions, while others prefer to have time to think by themselves. We are all different and need to be valued for our differences. The Improvement Leaders' Guide to Managing the Human Dimensions of Change gives ideas on how to ensure the best possible outcome when working with different people.
www.institute.nhs.uk/improvementguides

9.1 Carve out and segmentation

Objective
- to develop an understanding of carve out and segmentation and to get people thinking and talking

Benefits
- light relief during a capacity and demand session

Time required
- ten minutes maximum

Preparation
- participants to work in small groups of about five
- each group has a copy of the following sheet – without the answers of course
- facilitator to be judge – don't get side-tracked into too much discussion

Instructions to participants

- consider each of the examples opposite and decide if they are
  - carve out
  - segmentation
  - smoothing demand
  - flexing capacity

Learning points

- develops thinking about carve out and segmentation and the implications of each

Example for discussion	
Two slots in each clinic kept open for suspected urgent cancer patients	Carve out
Separate Integrated Care Pathways for neck of femur patients (or glue ear, or pigmented lesions)	Segmentation
Dedicated theatre lists for trauma patients (instead of putting emergency cases on to planned lists)	Carve out
Bus lanes	Carve out
Cash dispensers in banks	Segmentation
Named car park spaces for Trust Chair and Chief Executive	Carve out
"Please give this seat up for older people or mothers with children"	Carve out, but good carve out!
"Mr X's theatre shoes – do not touch"	Carve-out, but people tend to ignore it and wear them anyway
1st and 2nd class post	Carve out
Ring-fenced beds for elective admissions	Carve out
'This queue 8 items or less'	Segmentation, but more likely carve out – it's debatable
Happy hour – half price drinks from 6 till 7	Smoothing demand
'Slot' system for referrals – each GP can only refer a certain number of cases a month	Smoothing demand but not a good way to do so – doesn't take account of natural variation
Multi-function biochemistry analysis equipment for routine investigations; dedicated machines for specific tests	Segmentation
Fracture clinic held 7 days a week instead of Mon-Fri	Flexing capacity
Cheap off-peak rail fares	Smoothing demand but is it just an advertising ploy?
Wards based on patient dependency, rather than specialty	Segmentation
Nurse led diabetic clinics in General Practice	Segmentation
Think of some examples of your own	

9.2 Dice game

Objective
- demonstrates the effect of variation in clinical or other process

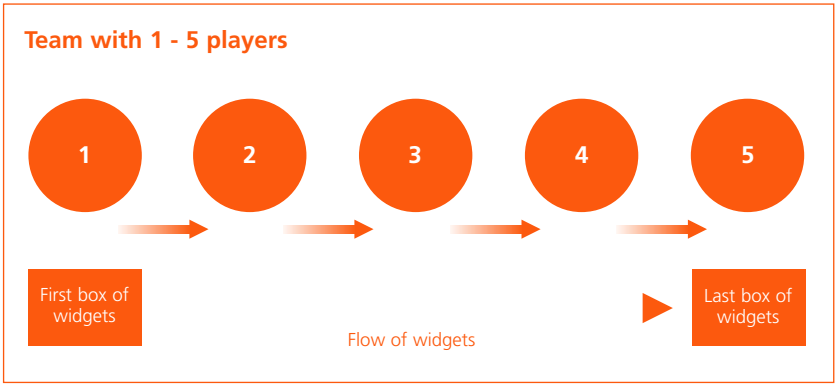
Benefits
- interactive for participants
- easy to set up

Level – advanced
- practice first on a group of willing volunteers
- warning: needs a good understanding of variation: the causes and effects
- do not attempt this activity without a good background knowledge

Time required
- 45 minutes, 15 minutes with 30 minutes for discussion

Preparation
- participants: teams of five players
- resources: flip charts and pens one die and 50 'widgets', e.g. buttons, pennies, etc. per team, score sheet for each player

Instructions to participants
- ask each team of five players to sit in a row and give each player a score sheet
- number the players 1 to 5
- put the pile of widgets on the left of player 1 and explain that the widgets should be passed to the next player to his right



- explain that the players represent a healthcare process (e.g. GP to outpatients to x-ray to theatre to rehabilitation) and they have been contracted to process 35 patients at the end of 10 weeks. This is based on the fact that each 'department' can throw anything from a 1 to a 6, and on average, will throw 3.5, which is equivalent to 35 at the end of 10 weeks.
- give the die to player 1
- player 1 rolls the die and takes the relevant number of widgets (patients) from the box and passes them and the die to player 2. Player 2 rolls the die and passes what they can to the third person, and so on
- when player 5 has rolled the die and delivered as many as he can to the far end of the row (last widgets box), you can start round 2
- the game lasts for 10 rounds – each person will have the opportunity of throwing the die 10 times

Completing the score sheet

Each person records on the scoresheet (see following pages):

- their position in the process (1 - 5)
- the round they are on (1 up to round 10) (column 1)
- the number they threw on the die (A)
- the number of widgets they moved to the next player (B)
- they are contracted to move 3.5 i.e. the average between 1 and 6 on the die (C)
- the difference between what they moved and what they were contracted to move (B – C)
- the cumulative performance for each round (week) (D + previous E)
- plot the cumulative score for each round (week) on the graph

After 10 rounds

- ask all players to hold up their score sheets so that the whole table can see the cumulative performance of the system. In general the cumulative performance (graph) should get worse (more negative) for each round and the further downstream in the process you get
- on the flip chart draw up a table with each team's name (e.g. table A, table B etc.) and ask the teams to call out the number of widgets (patients) they actually processed at the end of 10 rounds
- congratulate the team that scores the most and ask them what was different about their system when compared to the other teams. The answer should be nothing – their score due to luck.