

For further information see  
[www.modern.nhs.uk](http://www.modern.nhs.uk)

### Acknowledgements

The development of this guide for Improvement Leaders has been a truly collaborative process. We would like to thank everyone who has contributed by sharing his or her experiences, knowledge and case studies.

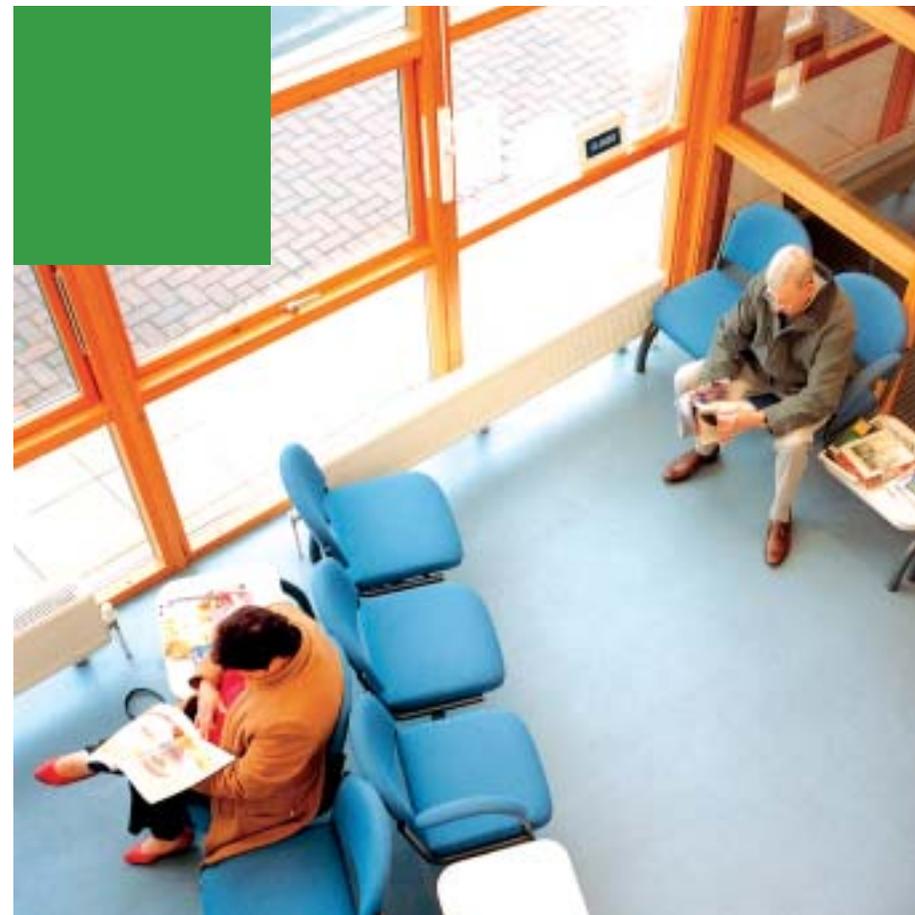
Design team: Kim Ashall, Sue Beckman, Helen Bevan, Ceri Brown, Kevin Cottrell, Jim Easton, Richard Green, Judy Hargadon, Sharon Jeffrey, Kam Kalirai, Ruth Kennedy, Susan Long, Mike McBride, Steve O'Neill, Jean Penny, Neil Riley, Kate Silvester, Paul Zollinger Read.

Our thanks to Darent Valley Hospital and Gravesend Medical Centre, Kent for their cooperation with the photography.

© 2002 Crown Copyright

Designed by Redhouse Lane, 020 7291 4646  
Published by Ancient House Printing Group,  
Ipswich, 01473 232777

### Improvement Leaders' Guide to Matching capacity and demand



I am pleased to present this guide – one of a series for Improvement Leaders in the NHS.

A key objective for all of us in the NHS, whatever our role, is to continually look for ways to improve the experience and care of patients. Many improvements have been achieved already, whether as part of a large national programme, or on a much smaller scale, through the commitment of a small team of healthcare staff. Everyone involved in such projects has gained so much knowledge about initiating and sustaining these improvements.

The aim of this set of guides is to gather this knowledge into a summary of current thinking. We've put them together in response to a huge demand for tools and techniques to support improvement in patient care. All the guides include useful, practical advice that can be applied in healthcare settings. Written by experienced healthcare staff, they are aimed at all Improvement Leaders in the NHS – by which we mean everyone who wants to improve the care and experience of patients, whether a manager of a small team in general practice, a director of modernisation for a large Trust, or the clinical leader of a team of doctors, nurses or therapists.

As I said, the guides are based on current learning and thinking – but this is constantly changing. If you've found this printed version useful, keep checking the website on [www.modern.nhs.uk/improvementguides](http://www.modern.nhs.uk/improvementguides). Here the guides will be regularly updated as we learn more and have new things to share.

Improvement and modernisation is really just beginning. It's an exciting time, and a great opportunity to develop and share new skills and understanding so we can truly make things better for our patients.

David Fillingham, *Director NHS Modernisation Agency*  
April 2002

## The Improvement Leaders' Guides

Collectively the Improvement Leaders' Guides form a set of principles for creating the best conditions for improvement in healthcare. The greatest benefit is when they are used to support a programme of training in improvement techniques.

### Where should I start?

The seven guides are not sequential and ideally you should read them all at an early stage in your improvement project, to be aware of the tools and techniques in all the guides. However there are some things we would suggest you should do first, as you develop your plan based on local needs and experience.

### Each guide includes

- some background information on the topic
- some activities which you, as an Improvement Leader, may find useful to help the teams you work with understand the basic principles
- questions that are frequently asked about the topic and suggested ways to answer them
- guidance on where to go for more information. Sources include the excellent toolkits that have been produced to support improvement programmes in specific services, such as Cancer, Critical Care, Mental Health and Clinical Governance. Useful books, papers and websites are also listed

## What's in each guide?

### Improvement Leaders' Guide to...

### What the guide has to offer an Improvement Leader

<p>Process mapping, analysis and redesign  <a href="http://www.modern.nhs.uk/improvementguides/process">www.modern.nhs.uk/improvementguides/process</a></p>	<p>This is definitely the place to start. This guide offers help in the use of the 'Model for Improvement'. This is a framework for setting aims, identifying the possible changes and beginning to think about measures that will show that your changes have made an improvement. Then there is the vital first stage of mapping your chosen patient process and analysing it to really understand what is happening.</p>
<p>Measurement for improvement  <a href="http://www.modern.nhs.uk/improvementguides/measurement">www.modern.nhs.uk/improvementguides/measurement</a></p>	<p>Question: how do we know a change is an improvement? Answer: by measuring the impact of the changes. This guide should also be considered very early on in an improvement project and gives valuable advice on what and how to measure for improvement and how to present the data to interested parties.</p>
<p>Matching capacity and demand  <a href="http://www.modern.nhs.uk/improvementguides/capacity">www.modern.nhs.uk/improvementguides/capacity</a></p>	<p>In our experience the process of truly matching capacity and demand has led to some of the most exciting changes in a healthcare process. This guide explains the most effective ways to understand the capacity and demand of a service and the 'bottlenecks' in the system which often cause patients to wait. It goes on to suggest ideas to reduce or eliminate these queues and waiting lists for patients. It is vital that process mapping and analysis is done prior to using this guide.</p>
<p>Involving patients and carers*  <a href="http://www.modern.nhs.uk/improvementguides/patients">www.modern.nhs.uk/improvementguides/patients</a></p>	<p>Everything we do should be focused on patients and their carers. They must be involved in our improvement programmes and projects from the very beginning. We are able to offer advice based on current thinking and experience of how to involve patients and carers in the most effective way, with warnings of possible pitfalls.</p>
<p>Managing the human dimensions of change*  <a href="http://www.modern.nhs.uk/improvementguides/human">www.modern.nhs.uk/improvementguides/human</a></p>	<p>Some of us take to the idea of change more easily than others. Some like to develop ideas through 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. This guide gives ideas of how to ensure the best possible outcome when working with different people.</p>
<p>Sustainability and spread*  <a href="http://www.modern.nhs.uk/improvementguides/sustainability">www.modern.nhs.uk/improvementguides/sustainability</a></p>	<p>It is fundamentally important that after making improvements they are sustained and built upon. This is a real challenge to anyone involved in improvement projects. It is also important that we share our learning and ideas with other areas of healthcare so that the maximum number of patients can benefit. This guide suggests methods and principles based on experience from healthcare both in England and beyond for sustaining and spreading improvement ideas.</p>
<p>Setting up a collaborative programme*  <a href="http://www.modern.nhs.uk/improvementguides/collaborative">www.modern.nhs.uk/improvementguides/collaborative</a></p>	<p>Experience has shown that working collaboratively produces the best environment for creating and sharing improvement ideas. Use this guide when a group of healthcare staff want to work in a different way, to innovate and test new models of delivering care, to dramatically improve the service for a group of patients and to create learning for their own organisation and the whole of the NHS.</p>

\* Available July 2002

### A few additional thoughts about the guides

#### The guides are based on current thinking and experience.

Be aware that this is constantly changing. Check updates on the Improvement Leaders Guides website, [www.modern.nhs.uk/improvementguides](http://www.modern.nhs.uk/improvementguides) which will be updated often as we test out and learn from new techniques.

#### Be aware of your own experience.

If this field is totally new to you, plan how you can find out more through further reading or development courses. If you are more familiar with leading service improvements, can you share your experiences and knowledge with others in your healthcare community and the wider NHS?

#### Make contact with others who have improvement skills.

Many people in healthcare have had training in the improvement skills contained in these guides. Their training will most likely have been for a particular service such as primary care, dermatology or cancer service. Make contact with them to form a health community improvement network to support and learn from each other.

#### Try it for yourself.

These guides don't represent the only way to do things, but they provide a good starting point. Create your own case studies and then share your experiences.

#### Take the thinking forward.

The website will be a dynamic medium. Please contribute to the discussion if you can. We would welcome and value your experience

#### Have fun.

Many have said that leading an improvement project has been one of the most enjoyable and fulfilling roles of their careers!

#### Let us know what you think of the guides.

We want your comments and thoughts about the Improvement Leaders' Guides. Our aim is to keep improving them so let us know what you think.

- how can we improve the guides? Is there anything we have left out?
- have you found them useful? If so which guide in particular and which section?
- how have you used them? Can you tell us any stories?
- if there were to be other guides, what topics should they be on?
- have you visited the web site? How can we improve it?
- is there any thing else you would like to tell us about the Improvement Leaders' Guides

Email us now on [improvementguides@npat.nhs.uk](mailto:improvementguides@npat.nhs.uk)

## Improvement Leaders' Guide to matching capacity and demand



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

*Eliyahu Goldrat*

## Contents

- 1 Introduction
- 2 Mapping the patient's journey
- 3 Analysing the patient's journey
- 4 Identifying the bottlenecks
- 5 Measuring demand, capacity, backlog and activity at the bottleneck
- 6 Redesign to match demand and capacity
- 7 Beware the dangers of variation
- 8 Activities to support matching capacity and demand
- 9 Frequently asked questions and answers
- 10 Useful reading for more information and ideas
- 11 Glossary of terms

## 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 in various projects that our valuable resources are not always best used 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 it is not the place to start.

- first you must map and analyse the process to really understand what happens to the patient
- then test out and implement changes that reduce the number of hand-offs and the number of non value adding steps across the whole process
- now look very carefully at the process map and identify those stages 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 understanding the capacity and demand problems
- begin to test and implement the relevant change ideas as a result of what the measurement shows you
- then create templates of the processes, begin to schedule those templates and watch the whole process improve.

## 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' Guide to Mapping the Patient Process [www.modern.nhs.uk/improvementguides/process](http://www.modern.nhs.uk/improvementguides/process) and The Improvement Leaders' Guide to Measuring for Improvement. [www.modern.nhs.uk/improvementguides/measurement](http://www.modern.nhs.uk/improvementguides/measurement) 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, what you need to do to map the patient process is:

- define and agree which group of patients is to be mapped
- 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 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. However there is usually something that is the actual cause of the bottleneck and this is the constraint. This is usually a skill or piece of equipment.

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

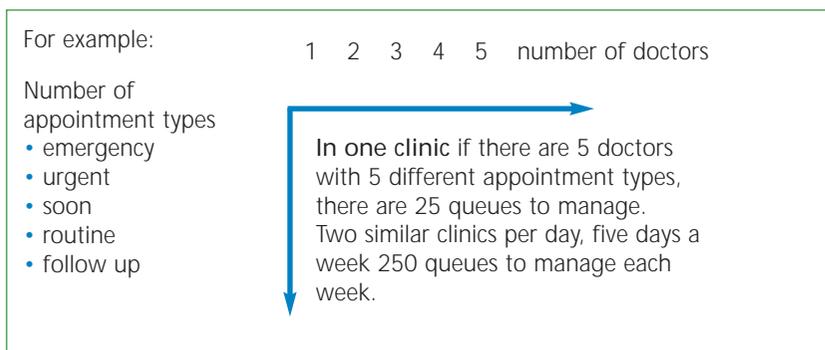
- get an appointment at their surgery (bottleneck). The constraint could be the availability of the skill of the GP.
- get their MRI appointment (bottleneck). The constraint may initially be thought to be the MRI scanner (equipment), but the scanner is there 24 hours each day. Again the constraint is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

### 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, or your parent or child, are 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?
- estimate the number of queues (groups of people waiting) at the bottleneck and the amount of time and effort required managing those queues, as in the diagram below
- look to see if work or patients are 'batched'. This is when the work accumulates for hours or even days before it is considered to be enough to bother attending to. For example, reporting a whole week's X-rays in one go, or allocating appointments for a whole week's referral letters at

Number of patient queues being managed



one time, rather than dealing with them as they come in

- if it is someone with 'expert' skills causing the bottleneck, look to see what the expert is doing. Is it 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.



#### Case study

##### Endoscopy Service in the West Midlands

This Endoscopy project started with long and variable waits for patients and knowledge that they were a key bottleneck in the journey for cancer patients. The team mapped the endoscopy journey to understand where the problem areas were for patients, and created what they wanted the ideal journey to be. 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
<b>Flexi. Sig.</b>										
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	
<b>Colonoscopy</b>										
Urgent	x	x	x	x				x	x	
Soon	x	x	x	x				x	x	
Routine	x	x	x	x				x	x	
<b>O.G.D.</b>										
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	
<b>ERCP</b>										x

73 queues to manage in endoscopy, West Midlands

## 4. Identifying the bottlenecks

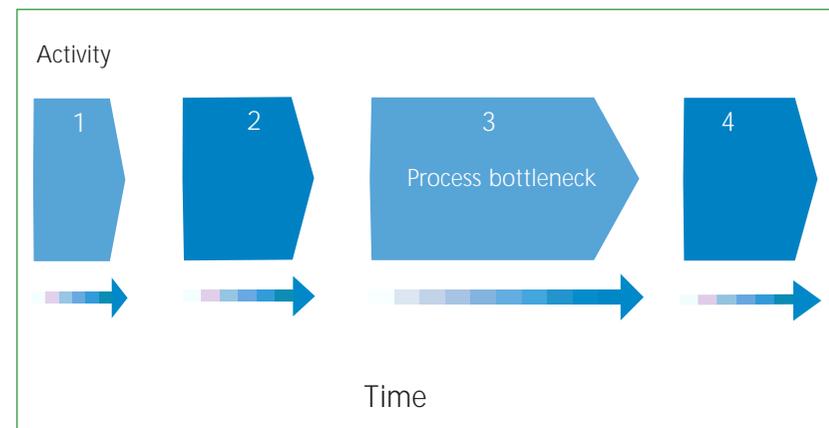
### 4.1 Concentrate on the bottlenecks

- 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: 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**
  - possible responses might be:
    - the consultant does not have time to see all her patients in clinic.
    - she has to see everyone who attends (including first visit assessments and follow-up patients).
    - it is what she has always 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

### 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.

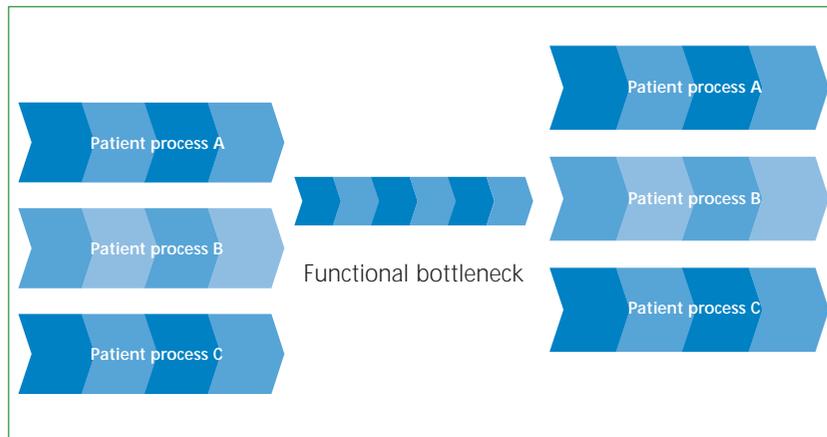
### Process bottleneck



**Process bottlenecks** are that stage in a process that takes the longest time to complete.

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 bottleneck



**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.

- one process, such as ENT surgery, might share a function such as imaging with other processes, such as orthopaedic surgery, and medicine
- a surgeon may be called to theatre when he should be in outpatients
- a GP has to go out on an emergency call when she should be in surgery

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 to flow in another.

## 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

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

- how to measure 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 a consultation time of 30 minutes each = 600 minutes (10 hours) of demand each day.

Make sure all demand is measured:

- all requests that come in by letter, phone call, fax, email, etc
- consider hidden demand as well – 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

Capacity is the resources available to do the work at the bottleneck step.

- how to measure 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.

### 5.3 Backlog

Backlog is the previous demand that has not yet been dealt with, showing itself as a queue or waiting list.

- how to measure 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 (50 hours) backlog

### 5.4 Activity

Activity is all the work done at the bottleneck step.

- how to measure 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 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, or specialised pieces of equipment, or test results

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

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)
<b>Total demand in minutes</b>	<b>1,020 minutes (17 hours)</b>

#### 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)
<b>Total capacity in minutes</b>	<b>960 minutes (16 hours)</b>

In this example, demand exceeds capacity by 60 minutes (1 hour) each week

## 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 the capacity, demand, backlog and activity issues wherever there are queues or waiting lists.

### Golden rule:

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

- who receives requests and what they do with them
- any patterns of 'did not attend' (DNAs)
- look carefully at the backlog patterns. If the backlog numbers remain constant over time, then demand and capacity are equal. However, the waiting times might not be.

### 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 the demand patterns for:

- variation in referral protocols and referral thresholds
- daily, weekly, monthly, and seasonal variations
- types of requests and who is making the request

### 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 projects 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 eg MRI 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:

- arrange pre-assessment of patients, checking fitness to proceed (prior to surgery)

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

- this is usually done by sharing skills in the team eg development of a nurse specialist role
- consider if the patient can provide their own care eg 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

Distribute the work amongst the clinical team:

- so that everyone works to their highest level of skill and expertise
- so that each has a similar waiting list/queue



#### 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.

### 6.3 Resolve capacity problems at the appropriate point in the system

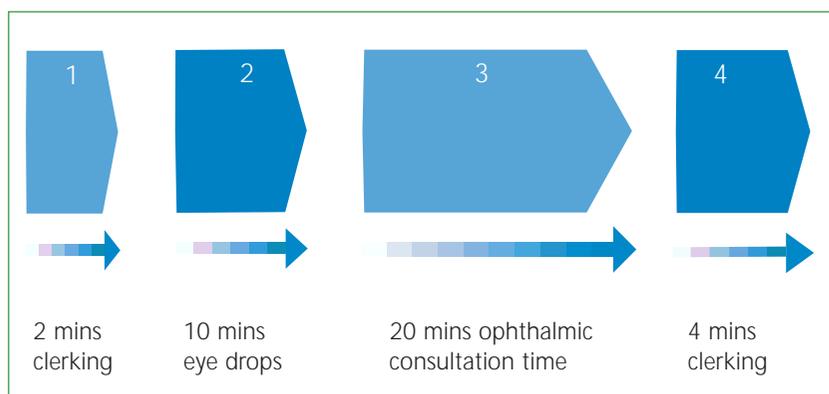
Often changes are introduced in the wrong place. Focusing on and speeding up the beginning of the patient journey will cause a build-up of patients further along the process at the site of the constraint.

- 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 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
  - providing feedback on inappropriate referrals

#### Ophthalmic process



### 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, 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
- reduce/eliminate batching – do work when it arrives rather than waiting to deal with a whole set of similar tasks at the same time
  - report X-rays as they are done or make appointments as they come in

rather than piling them 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 eg blood tests, ECGS etc, can be done at the same time and results can be transported easily 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
  - 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
  - one non-healthcare example of a pull system is the use of chevrons on motorways. The rule is to keep two chevrons between you and the car

in front. Therefore all cars go at the rate of the slowest car but if that car speeds up so will the whole system

- 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. The challenge is to develop pull systems working at the rate of the bottleneck

### 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 far in the past, they not only cannot respond to today's requests but also are actually 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 plan must include:

- an understanding of the true extent of the backlog
- 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

- commitment by senior clinical and managerial leaders, as this needs commitment across departmental and organisational boundaries

The overall goal is to reduce demand appropriately, effectively and permanently, and increase capacity appropriately. 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

#### Golden rule:

Every time demand exceeds capacity, you carry forward the excess demand as backlog. But you cannot carry unused capacity forward.

### 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 be ‘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.



#### 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 reducing waiting time.

**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
- make pauses in the schedule to catch up
- plan capacity at 80–85% of the fluctuation in demand. This ensures that queues and waiting lists never 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 contingency plans 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/sessions to compensate for those lost on 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 in the useful reading section on the website, [www.modern.nhs.uk/improvementguides/reading](http://www.modern.nhs.uk/improvementguides/reading) including lots of ideas and examples where teams have made improvements in a wide range of healthcare settings.



**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. Beware the dangers of variation

### 7.1 Carve out and segmentation

Often we have dealt with healthcare problems by prioritising ('ring fencing') or carving out the time of an expert, or the time of specialised equipment; or by keeping resources or facilities only for one particular group of patients. Accurate measuring of the backlog/waiting time for other groups of patients has shown that carving out capacity significantly increases the variation in waiting times and creates a very difficult system to manage effectively.

**The difference between carve out and segmentation**

	Segmentation	Carve out
Objective	<ul style="list-style-type: none"> <li>• to improve the flow for all patients</li> </ul>	<ul style="list-style-type: none"> <li>• to improve the flow for a specific group of patients at one bottleneck</li> </ul>
Principles	<ul style="list-style-type: none"> <li>• looks at the whole patient process</li> <li>• groups patients with similar processes</li> <li>• keeps the flow through the process bottleneck constant</li> <li>• matches demand to capacity along the process</li> </ul>	<ul style="list-style-type: none"> <li>• looks at one bottleneck eg CT, theatre, outpatients</li> <li>• prioritises the queue irrespective of the patient need</li> <li>• interrupts the flow of patients and keeps them waiting at all steps in the process</li> </ul>
Effect on waiting time	<ul style="list-style-type: none"> <li>• reduces/eliminates waiting for patients</li> </ul>	<ul style="list-style-type: none"> <li>• makes the waiting time worse</li> </ul>
Effect on other patient groups	<ul style="list-style-type: none"> <li>• none</li> </ul>	<ul style="list-style-type: none"> <li>• other patients, eg with non urgent chronic diseases, have a much longer wait</li> </ul>

	Segmentation	Carve out
examples	<ul style="list-style-type: none"> <li>opticians diagnose, prepare and add cataract patients directly onto the surgical list. This removes them from other ophthalmic patients in the outpatient clinic</li> </ul>	<ul style="list-style-type: none"> <li>reserving slots on a CT scanner for certain groups of patients eg neurology</li> <li>ring fencing beds for certain groups of patients</li> <li>reserving clinic/surgery slots for specific groups of patients eg urgent, soon, routine</li> <li>creating different queues for different consultants even when process is the same eg endoscopy</li> </ul>
More examples – non healthcare	<ul style="list-style-type: none"> <li>cash dispensers – customers totally removed from other customers at the counter in the bank</li> </ul>	<ul style="list-style-type: none"> <li>bus lanes at peak times on busy roads</li> <li>parking for mother and babies in supermarket car parks</li> <li>first and second class post</li> </ul>
Summary definition	<ul style="list-style-type: none"> <li><b>segmentation</b> is when the separation of the process of care along the whole pathway for one group of patients is <b>not at the expense</b> of other groups of patients</li> </ul>	<ul style="list-style-type: none"> <li><b>carve out</b> is when the flow of one group of patients is improved <b>at one bottleneck at the expense</b> of another group of patients</li> </ul>

**Golden rule:**

Avoid making great improvements for one group of patients if it is to the detriment of another group of patients.

### 7.2 Reducing variation

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 multi-disciplinary 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 plan. 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.

## 8. Activities to support matching capacity and demand

Before organising any activity, consider the following:

- who is the audience?
- what is their prior knowledge?
- 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 through 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.modern.nhs.uk/improvementguides/human](http://www.modern.nhs.uk/improvementguides/human)

### 8.1 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 capacity but is it just an advertising ploy?
Wards based on patient dependency, rather than specialty	Segmentation
Nurse led diabetic clinics in General Practice	Segmentation
<b>Think of some examples of your own</b>	

## 8.2 Dice game

### Objective

- demonstrates the effect of variation in a clinical 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 'widjits' eg 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 widjits on the left of player 1 and explain that the widjits should be passed to the next player to his right

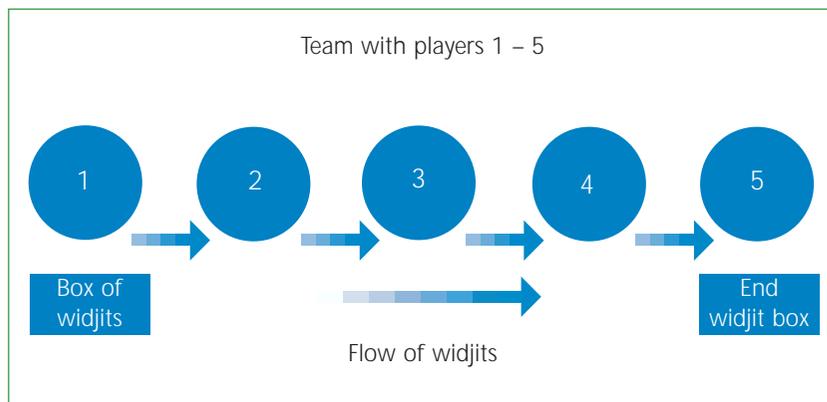
- explain that the players represent a healthcare process (eg GP to outpatients to x-ray to theatre to rehabilitation) and they have been contracted to process 35 patients at the end of 10 weeks. This is based on the fact that each 'department' can throw anything from a 1 to a 6, and on average, will 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 widjits (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 (end widjit box), you can start round 2
- the game lasts for 10 rounds – each person will have the opportunity of throwing the die 10 times

- 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 (eg table A, table B etc) and ask the teams to call out the number of widjits (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 was just luck.

### Setting up the dice game



### Completing the score sheet

Each person records on the scoresheet (see page 34):

- 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 widjits they moved to the next player (B)
- they are contracted to move 3.5 (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)

Adapted from 'The Goal' by Goldratt & Cox.

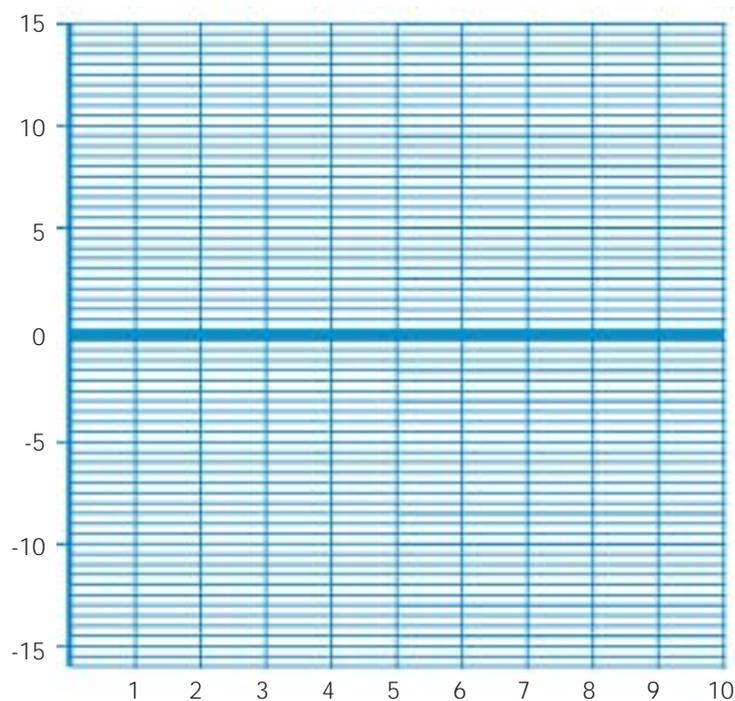
The Dice Game Score Sheet: example

Position in process	Round number										
Round	Example	1	2	3	4	5	6	7	8	9	10
A Capacity (The number that you threw)	6										
B Activity (The number that you move)	3										
C Contract (The number you should have moved to meet contract)	3.5	3.5	3.5	3.5	3.5	3.5	3.5	3.5	3.5	3.5	3.5
D Difference between Contract and Activity (B minus C)	-0.5										
E Cumulative performance to plot (D plus previous E)	-0.5										

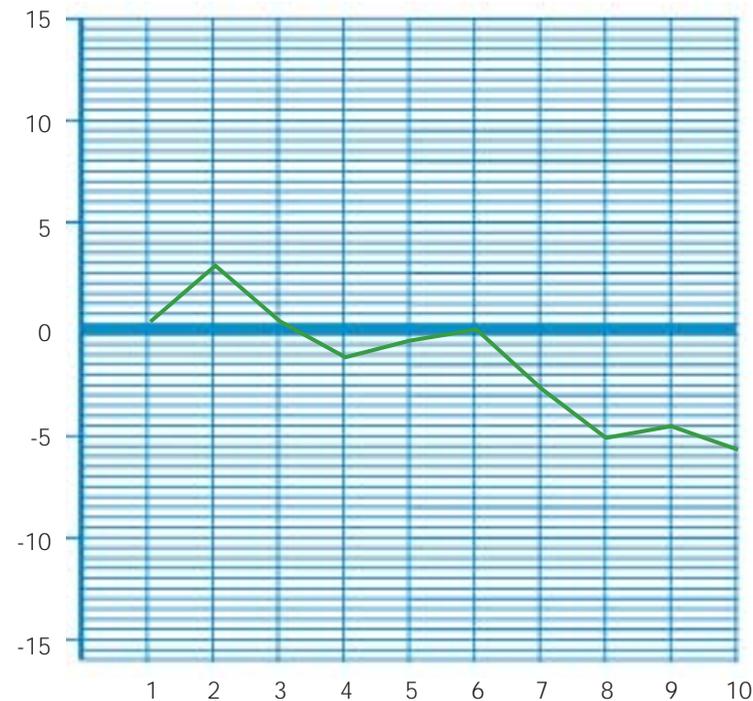
At the end of round 10 each player should have a score sheet looking something like this:

Position in process: 1	Round number										
Round	Example	1	2	3	4	5	6	7	8	9	10
A Capacity (The number that you threw)	6	4	6	1	2	4	4	1	1	4	3
B Activity (The number that you move)	3	4	6	1	2	4	4	1	1	4	3
C Contract (The number you should have moved to meet contract)	3.5	3.5	3.5	3.5	3.5	3.5	3.5	3.5	3.5	3.5	3.5
D Difference between Contract and Activity (B minus C)	-0.5	0.5	2.5	-2.5	-1.5	0.5	0.5	-2.5	-2.5	0.5	-0.5
E Cumulative performance (D plus previous E)	-0.5	0.5	3	0.5	-1	-0.5	0	-2.5	-5	-4.5	-5.5

Plot of cumulative performance



Plot of cumulative performance



## Discussion

### What is going on?

- if player 1 rolled a 5, they would pass 5 patients on and record that they passed 1.5 more than the average activity required of them
  - if player 2 rolled a 4, they would pass on 4 and record that they passed 0.5 more than the average activity required of them. They would have a queue (backlog of 1)
  - if player 3 rolled a 2, they would only pass 2 along. However they delivered 1.5 less than the average activity required of them. They also had a queue of 2 (backlog)
- as the game progresses the players downstream of position 1 notice that there are rounds when either:
- they have more patients waiting than they have capacity (throw on the die) to pass on and have a backlog
  - or they have fewer patients to pass on than they threw on the die – they are starved of work.

### Question

How could the teams improve their systems performance so that it always meets their target of 35 patients at the end of 10 weeks?

### Answer

The most common strategy is that each player would always have a pile (queue) of patients waiting so that if they throw a big number then they will always have patients to pass along. But this is not good for our patients

### Question

What is the real problem for the team?

### Answer

The real problem is the assumption that the capacity of the system is the average for all the departments in the process. Since the effective capacity of each station in the chain varies between 1 and 6, the average effective capacity is 3.5. However this isn't the case.

### Question

What is the chance that the team will process at least 1 patient?

### Answer

100%. Each position is bound to throw 1 to 6 and will therefore be able to pass one patient along from position 1 to position 6

### Question

What is the chance that the team will process 6 patients in each round?

### Answer

The chance of producing at least 6 patients =  $\frac{1^5}{6^5} = 0.013\%$

### Question

So what is the chance of the team producing 3.5 patients in each round?

### Answer

About 9%. This system has a 9% chance of meeting its target!

Chance of at least 3 patients  
=  $\frac{4^5}{6^5} = 13\%$

Chance of at least 4 patients  
=  $\frac{3^5}{6^5} = 3\%$

### Question

Is this why the contracting process is difficult to manage?

### Answer

GPs are contracted to do X activity, the outpatient departments are contracted for Y activity, and Z inpatient activity. But no one understands the real flow through each clinical process.

### Question

What can we do?

### Answer

You have a choice of three strategies

**Strategy 1** Measure the variability of the capacity in all the departments continuously and build a computer model for predicting the amount of backlog that each department needs to hold in order to optimise the performance of their process?

**Comment:** Don't even think about it!

**Strategy 2** Increase capacity – more dice.

**Comment:** Doubling the capacity with the same amount of variation will not help.

**Strategy 3** Reduce the variation – load the dice so they always throw a 3 and 4 only.

**Comment:** This is the best solution. Find out why the capacity varies so much and stop it. 80 to 90% of the variability is within the system's control.

### Further questions to stimulate the debate

- what causes the capacity to vary?
- how would the team set about stopping it?
- how should the commissioning process work?
- what would be the appropriate measures of performance?
- which do we currently do in the NHS?

## 9. Frequently asked questions and answers

### Question

How do we start?

### Answer

Don't forget to start in a small way perhaps in one clinic or on one theatre session. Choose an area where there is support from a senior clinician and manager. Map the process first and identify the bottlenecks. Involve everyone in the process mapping so that they all understand what is going on. Help the team to test and implement other smaller changes as well as looking for major changes. Remember to consider the people aspects – don't jump straight into capacity and demand, which can seem quite threatening.

### Question

How can I make sure all the staff measure the same thing?

### Answer

Create a proforma. Have a standard paper or electronic form readily available to all staff and advise them how to complete the measurements before they start. Perhaps have a practice for a day or so before the real measurement begins.

### Question

How will all this really help us?

### Answer

The combination of process mapping and analysis and matching capacity and demand will show when changes introduced begin to make improvements – a great boost for the team and other interested people. It will also help to show other areas for improvement and provide wonderful support in business cases for extra resources as you can show that all other options have been considered and tested.

### Question

What is the difference between carve out and segmentation?

### Answer

Refer to section 7.1. Look around you for your own examples. It is not easy to understand at first so keep talking to your colleagues and other Improvement Leaders.

### Question

This all feels a bit remote from the patient. How can we make sure that we involve patients and their carers?

### Answer

This is so important that there is an Improvement Leaders' Guide dedicated to this topic. Everything we do should be focused on patients and their carers – so they must be involved in our improvement programmes and projects from the very beginning. We are able to offer advice based on current

thinking and experience of how to involve patients and carers in the most effective way, with warnings of possible pitfalls. See [www.modern.nhs.uk/improvementguides/patients](http://www.modern.nhs.uk/improvementguides/patients)

### Question

Help, I'm stuck!

### Answer

As an Improvement Leader you should be ready to advise, coach or actually support. If you need help yourself, talk things through with your colleagues in your healthcare organisation, contact anyone involved in any of the Modernisation Agency programmes in your health community or visit the Modernisation Agency web site.

[www.modern.nhs.uk](http://www.modern.nhs.uk)

There are a number of ways to contact others who will be able to help.

### Question

Is there one book you would recommend that would help me to understand capacity and demand a bit more?

### Answer

There are lots of good books written but if you want to start with one that is easy to read try this one *Goldratt, E., Cox J., (1993), The Goal, Gower, Aldershot.*

### Golden rule:

'Those who say that it cannot be done should not interrupt those doing it.'

Old Chinese proverb

## 10. Useful reading for more information and ideas

Much has been written about improvement and change. So much so, in fact, that it is very easy to get overwhelmed by all the material. We have, therefore, gathered together the things that we think you might find most useful. We would like to guide you in three directions:

### 1. Toolkits

Produced by national and regional programmes for staff addressing the issues for one particular aspect of care. This can range from general workforce planning issues to addressing the problems of a particular service eg endoscopy, orthopaedics. They are written for clinical staff in the specific service and will give you many more change ideas, lots of case studies, national contact names and information on how to access up to date activity in that particular area or service.

**Use:** when you have identified a problem associated with a particular service.

### 2. Books, papers and articles

Produced by international experts in their field addressing the science and theory behind many of the tried and tested tools and techniques in the guides.

**Use:** when you want a deeper understanding in any of the topics.

### 3. Websites

Time is precious and the World Wide Web is vast. Therefore we want to guide you to a few selected websites designed to extend your knowledge and thinking on improvement thinking.

**Use:** when you want to extend your general knowledge and gain access to improvement thinking around the world.

So visit the Improvement Leaders' Guide website for the useful reading section [www.modern.nhs.uk/improvementguides/reading](http://www.modern.nhs.uk/improvementguides/reading). This will be continuously updated as new editions are published and you tell us what you find useful.

## 11. Glossary of terms

Some of the words used in improvement have been defined. Use them carefully.

<b>Activity</b>	All the work done. This does not necessarily reflect capacity or demand, as the activity in June may well include demand carried over from May, April, or even March
<b>Backlog</b>	Previous demand that has not yet been dealt with, showing itself as a queue or a waiting list
<b>Batching</b>	Piling up a type of work as it comes in until a later time when all this type of work is done together
<b>Bottleneck</b>	Part of the system where patient flow is obstructed, causing waits and delays
<b>Capacity</b>	Resources available to do work. For example, the number of pieces of equipment available multiplied by the hours of staff time available to run it
<b>Constraint</b>	The actual cause of the bottleneck. Usually a necessary skill or piece of equipment <i>[NB Goldratt uses constraint to mean the same as bottleneck, but recognises that there are different types of constraints]</i>
<b>Demand</b>	All the requests/referrals coming in from all sources
<b>Hand-off</b>	When the patient is passed on from one healthcare professional to another
<b>Parallel processes</b>	Different activities that take place in the same time period
<b>Queue</b>	Work waiting to be done at a given point. For example, patients waiting to be seen in the clinic or people on a waiting list to come in to hospital for surgery.
<b>Scope</b>	A definition of the boundaries of the area under examination. For example, the beginning and end points of the stage of the patient journey under review