Emergency Medicine Programme

Implementation Guide 8:

Matching Demand and Capacity in the ED

January 2013

Introduction

This is a guide for Emergency Department (ED) and hospital operational management teams to help them understand how demand and capacity factors can be analysed and managed to deliver better patient care.

The presentation explains:

1. The language of demand capacity management
2. How to estimate ED system demand and capacity
3. The effects of variability on ED processes
4. How process bottlenecks impact ED performance
7. Demand capacity analysis can be used to manage EDs

This guide is based on best current evidence and international literature on ED demand capacity matching. It supplements the Emergency Medicine Programme Report 2012. Further information on demand capacity management in the ED setting is available in the references at the end of this presentation.

Definitions - demand

EMERGENCY MEDICINE

ED demand is: - patients;

- the rate at which they arrive;
- their care time;
- the care resources that each patient needs.[1]

Key Features of ED demand

- **ED demand is variable but predictable**
  The precise arrival time of the next patient is unknown but data analysis over time enables us to predict that on "average" an ED has "X" number of arrivals within a particular hour.
- **ED demand has to be met when it occurs**
  Patients must be seen on arrival or within standard MTS time frames; emergency care is not usually deferred.

ED Demand – other key features

EMERGENCY MEDICINE

- **Some demand may be hidden or unrecognised**
  Unmet demand may emerge when process capacity is increased
  (e.g. patients who previously left before completion of treatment wait to be seen as ED processes improve).

- **New demand may be created**
  Improvements in the quality of care may increase ED demand;
  New treatments (e.g. stroke reperfusion) may also increase demand;
  Increased demand in the ED may reduce demand elsewhere in the health system e.g. ED care pathways reducing demand for in-patient admissions.

- **Efforts to divert ED demand are unlikely to be effective**
  Experience internationally has demonstrated little or no reduction in ED demand from initiatives to divert ED attendances to other care settings.

Definitions – variability

Two kinds of variability are recognised:[2]

Natural or random variability

- No two patients or clinicians are the same;
- Patients do not arrive at exactly the same time each day;

Natural variability cannot be eliminated and must be accommodated in ED processes.

Non-random variability

This is introduced to a system through individual behaviours and priorities.[3]

e.g. One senior clinician may like to run an ED shift in a different way to colleagues. This may cause problems that affect patient care or require increased resources. If so, it should be eliminated through standardising work i.e. finding an agreed best way to do things and ongoing process improvement.

Non-random or artificial variability should be minimised and eliminated where possible to optimise process quality and efficiency.

Definitions - capacity

Capacity is the maximum level of value-added activity that a process can achieve under normal operating conditions over a period of time.[4]

ED examples include the number of patients a nurse can triage or a doctor can assess in an hour.

The capacity of the ED is largely a function of:

- the processes,
- staffing levels;
- physical space;
- equipment capacity. [1]

Processes exert a big effect on ED capacity particularly when constraints or bottle-necks occur. [1]

Definitions – processes and bottlenecks

A process is a series of actions taken to achieve a result.

A constraint[5] (or bottle-neck) is a hindrance to a process or production, a delay in one process step in a process reducing the efficiency of the next step. e.g. waiting for test results to be available to safely discharge a patient

Constraints can include:

- Equipment e.g. access to telemetry cubicles;
- People e.g. a shortage of EM Middle Grade Doctors;
- Policies, protocols and procedures e.g. not having appropriate referral protocols.

ED constraints can be imposed by internal and/or external factors.

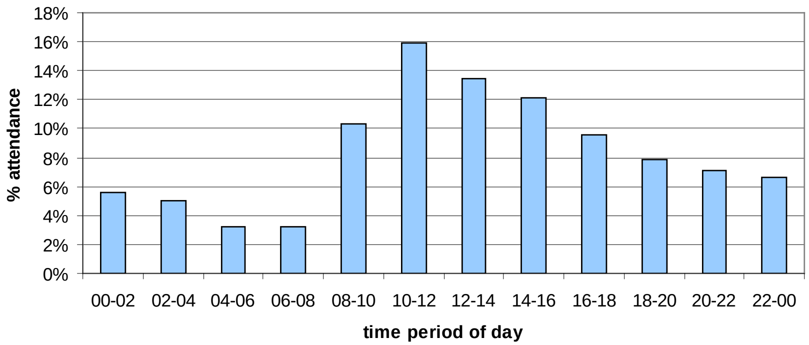
Understanding and managing constraints enables process improvement, as outlined later in this presentation.

Understanding ED Demand

Understanding your ED demand

ED demand can be analysed on many levels including:

- Overall new patient attendance demand by day of week, hour of arrival;
- Identifying predictable surge patterns in ED demand;
- Demand for patient care streams e.g. ambulatory care; RAT; initial ECG etc.;
- Numbers of review clinic attendances;
- Patterns of Clinical Decision Unit admissions.



ED demand analysis, such as the pattern of hourly arrivals (illustrated), enables staffing rosters to be matched to demand patterns e.g. ensuring additional staff are on duty from 8am onwards in this example.

Data Analysis

- ED Information Systems (EDIS) provide high quality data for demand analysis.
- An EDIS is not absolutely necessary. Simple sampling exercises provide basic data to understand ED demand[6], though this sampling exercises can be time-consuming and produce less reliable data.
- Good EDIS data and high quality simulation have been used to analyse and manage process demand and capacity in an Irish ED.
- The EMP advises that EDs identify where their process bottlenecks occur and undertake simple sampling exercises (or EDIS-based data reports) to better understand demand at these process steps.

Demand Exercise

Example scenario:	An EMP implementation team finds that queues for triage occur for walk-in patients and uses demand capacity management to resolve this problem.
Step 1:	Determine how many patients attend each hour. Run an EDIS report to identify how many patients on average attended each hour based on a previous month (or year's) data. If no data report is available, ask staff to note the time of arrival of each patient for a sample period from a few hours before the queues are usually seen until after they diminish, for one or two days (less reliable, but nonetheless useful data).

Demand Exercise

Time interval	Average patient arrivals
08:00 – 09:00	4
09:00 – 10:00	8
10:00 – 11:00	12
11:00 – 12:00	15
12:00 – 13:00	17
13:00 - 14:00	16
14:00 – 15:00	15
15:00 – 16:00	9
16:00 – 17:00	7

Analysis of this fictional example shows that 12 patients arrive between 10:00 and 11:00.

Clearly, this data does not account for natural variability in patient arrival data; patients don't attend at regular intervals during the hour. This 'simplified' data will nonetheless allow us to plan to meet triage service demand.

Demand Exercise: cycle time

EMERGENCY MEDICINE

Step 2. Estimate Triage Cycle Time:

Sample how long it takes a Triage Nurse from the end of one patient's triage to the end of the next patient's triage; this includes the time taken to call in the patient and move them on at the end of triage. This time may be available on EDIS or obtained through observing a small sample of triage cases. Example result: Average cycle time = 6 minutes.

Step 3. Compare Triage Cycle Time to Triage Demand

Example result: A single triage nurse has maximum capacity for 10 triage cycles each hour, working non-stop. Queues are inevitable when 12 patients arrive in an hour (6 minutes X 12 patients = 72 minutes work) with only one nurse providing triage.

Demand Exercise: server utilisation

Step 4. Allow for process variability:

Queues will also result if only 10 patients attend per hour, for triage with an average cycle time of 6 minutes because there will be variability in the time taken to triage (i.e. some patients will take longer than 6 minutes) and patients do not arrive at regular intervals during the hour. The Triage nurse can be considered a “server’ in processing terms.

Queuing theory tells us that pushing a system close to 100% server utilisation will inevitably cause queues.

More specifically, small changes in demand above 85% server utilisation result in queue formation due to the effects of random variation.

The more reliable a process and the smaller the variation that applies, the higher the server capacity that can be tolerated without queues (e.g. 90% in some highly reliable processes).

Triage Exercise: Server Utilisation

EMERGENCY MEDICINE

Time interval	Average patient arrivals	Number of nurses	Capacity (patients/hour)	Server Utilisation
08:00 – 09:00	4	1	10	40%
09:00 – 10:00	8	1	10	80%
10:00 – 11:00	12	1	10	120%
11:00 – 12:00	15	2	20	75%
12:00 – 13:00	17	2	20	85%
13:00 - 14:00	16	2	20	80%
14:00 – 15:00	15	2	20	75%
15:00 – 16:00	9	2	20	45%
16:00 – 17:00	7	1	10	70%

Improving the capacity by introducing a second Triage “server” to maintain the process is at >85% server capacity will reduce and prevent queues. When should this second server be introduced, based on the above data?

Demand Exercise: server utilisation

EMERGENCY MEDICINE

Step 5. Optimise server utilisation:

Example answer: Queues will have occurred during the hour from 10:00 onwards, so the two Triage Nurses will be playing catch-up from 11:00, as well as dealing with ongoing demand. Two Triage Nurses would result in 60% server utilisation from 10:00, but queues would be less likely.

Resource constraints may mean that a second Triage Nurse is not be available or may need to be shared from another care stream.

Can this other care stream be made more efficient to free-up nursing resource to help with triage?

Consider what strategies may be used to find the most cost-effective way to respond to this increased Triage demand.

Responding to Demand

EMERGENCY MEDICINE

1. **Reduce the cycle time**:
   Could triage be done quicker i.e. the cycle time reduced? Is all equipment to hand? Are non-essential tasks being done at Triage?

2. **Reduce the non-random variation:**
   Streamline the triage process and ensure staff have appropriate training.

3. **Reduce demand:**
   Sending patients from the walk-in triage stream to the ambulance patient triage stream, if there is spare capacity in that stream. Demand can be reduced for other processes by selecting the most appropriate patients for that stream e.g. identifying patients who really need an ECG.

4. **Increase process capacity:**
   Allocating a second Triage Nurse.

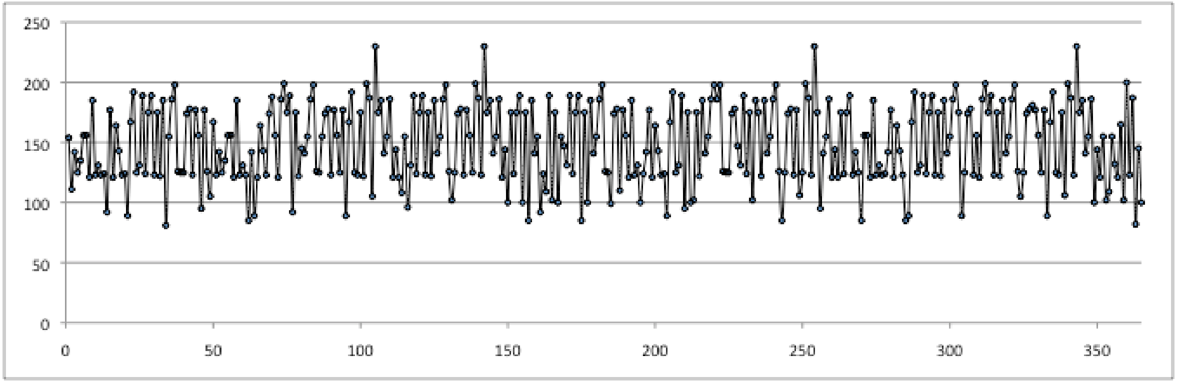
Understanding Variability in ED Demand

ED Demand Variability

- Even highly variable processes show patterns if analysed over sufficient time periods. ED new attendance demand is variable but usually occurs within predictable limits.

- Run charts[7] can be used to display process data or to track the impact of interventions or events.

- A "run-chart" of ED attendances over a year demonstrates the variability in ED attendance data. Basing capacity on the average of 148 daily attendances would underestimate demand and result in queues on many days.



Statistical Process Control Charts

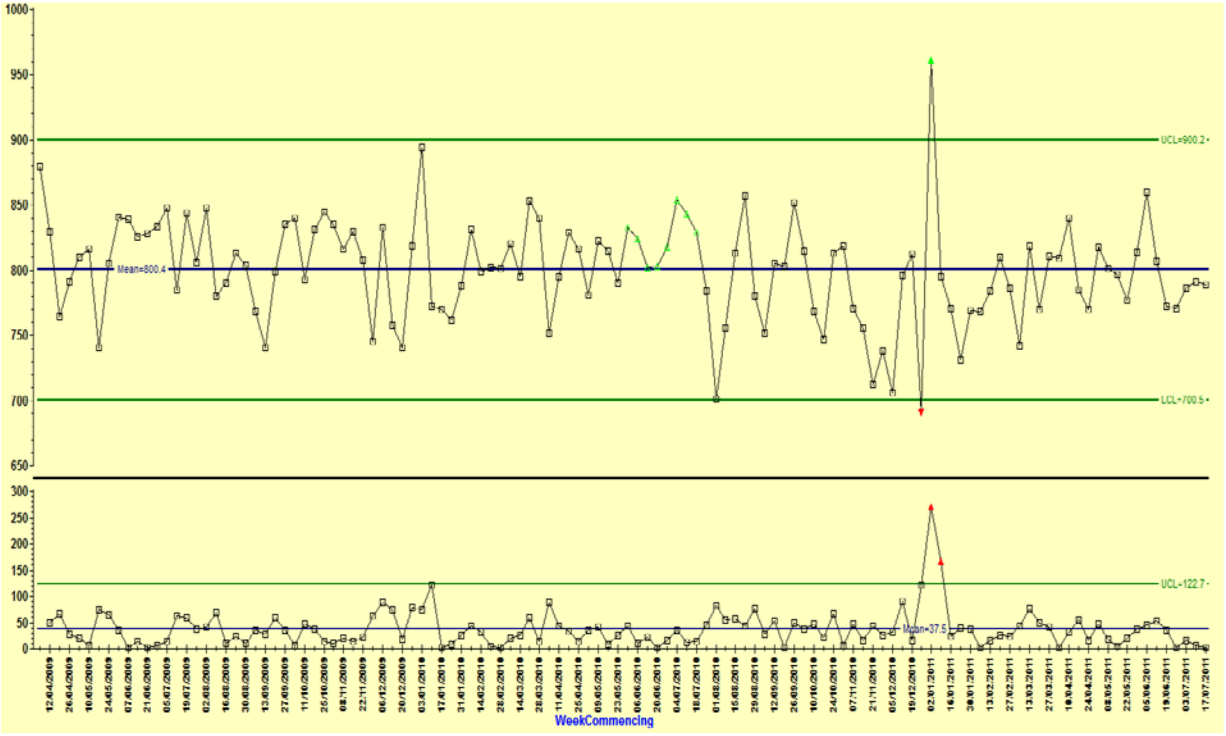
- A Shewart chart or Statistical Process Control (SPC) charts can be used to analyse variation. It is an extension of a run chart that indicates a central tendency of the data and upper and lower control limits (usually the mean and 2 standard deviations).[1,8]

- SPCs demonstrate that although patterns are variable they usually fall within predictable limits, unless some exceptional event occurs.

- SPC charts are commonly used to analyse ED attendance patterns and demand for in-patient beds.

Statistical Process Control Example

EMERGENCY MEDICINE



Displaying and Analysing ED Data

EMERGENCY MEDICINE

- Charts are often the most useful way to display and share ED process data. These can include bar charts, frequency plots, pie charts, run charts[7] or more complex SPC charts.[8] Data tables can be helpful too.

- When reviewing charts from different EDs or sources, check the scales are comparable or misinterpretation is likely.[1]

- Ensure that reporting periods are similar. Comparing calendar months will result in different data patterns versus comparing standard 28 day periods. Standard 28 day periods are preferable for tracking ED attendance patterns as each has the same number of weekend days.

- Adequate data quality and standardised data definitions are essential to producing high quality analyses on which capacity planning and other decisions can be based.

Accommodating Variability in the ED

EMERGENCY MEDICINE

- ED processes usually have fixed capacity to meet variable demand e.g. a fixed number of staff on shift.

- ED processes need to have enough routine capacity to accommodate usual levels of demand variability as part of normal business.[2] Without adequate baseline process capacity, systems (and staff) will become overstretched with increased risk of process failure causing adverse outcomes for patients and staff e.g. serious safety events.

- Few ED processes have excess capacity to deal with all potential surges in demand and even highly performing systems may experience occasional queues.

- ED cost analysis models often fail to account for the costs associated with process failures and inefficiency, nor do they consider the gains made from education, training and quality improvement that can be achieved during times when staff are less busy[2] e.g. when early morning staffing levels exceed service demand but are necessary for staff training.

Using Demand Capacity Analysis to Better Manage your ED

The Pathway of Care

The previous exercise looked at one step in a patient's pathway of care.

ED management involves multiple co-delivered processes and pathways of care.

The efficient provision of multiple processes and pathways requires understanding of:

1. Constraints and bottlenecks
2. The use of streaming
3. How to manage capacity across a system

Theory of Constraints - overview

- The theory of constraints was developed to make manageable processes more effective in achieving an organisation's goal.[5]

- It states that time lost at bottlenecks in one process is lost to the entire system.

- Time saved by improving efficiency at a step in a process that is not a bottleneck, will not improve the overall system performance.

- Bottlenecks may jump to different stages in highly variable processes.

Removing Constraints

To remove process constraints :

- Identify that there is a problem in a particular process;
- Determine how it can be resolved (i.e. analyse causes and solutions);
- Align your team, processes and organisation to tackle this bottle-neck;
  (using EMP Process Improvement methods)
- Make other major changes necessary to sort out the problem;
- If a new bottle-neck is identified tackle that and continue to improve; avoid inertia.

Theory of Constraints in the ED

The Theory of Constraints[5] applied to the ED setting suggests that:

- We need to analyse the patient journey through the ED to identify where true bottlenecks occur. Improvement efforts should be focused at first on the most severe bottlenecks.

- Working on non-bottlenecks will not yield big results e.g. reducing the time from disposition decision to discharge will not have a major impact on Total ED Time if most delays occur from triage to time seen by a treating clinician.

- Constant monitoring and analysis is necessary to sustain improvement because there is always at least one constraint in a process, solving one problem may throw up another and new constraints can occur e.g. once turn around times for laboratory test results are reduced, the next bottleneck may be the time from registration to when blood tests are taken.

Streaming

EMERGENCY MEDICINE

- Streaming is the separation of patients into groups undergoing similar processes or cycle times and processing the groups in parallel to increase the efficiency of the processes.[1]

- Streaming reduces queues by having two (or more) process outflows, with some patients avoiding particular process bottlenecks e.g. having two triage streams, one for walk-ins and one for ambulance arrivals or patients with limb injuries being streamed from triage to ANP care, thereby avoiding any delay to be seen by a doctor.

- Demand analysis for a care stream should be undertaken to determine if any additional resource identified for streaming is justified.

Streaming: demand capacity analysis

EMERGENCY MEDICINE

- Assessing the demand for a particular care stream can be undertaken by sampling how many patients would be suitable for this service across a 24 hour period.

- There should be enough patients to keep a dedicated team appropriately busy during the hours the stream operates, because if this team is left unoccupied and is not redeployed within the ED, their time will be lost to the system (termed carve-out).

- We may tolerate carve-out where the inefficiency is off-set by the quality gains e.g. allocating senior staff to resuscitation cases even though there are long queues for other work-streams.

Managing Capacity across a System

EMERGENCY MEDICINE

- All EDs function with limited resources.
- Demand capacity analysis can indicate where major capacity short-falls occur and these may be addressed through:
    - Improving process efficiency to optimise throughput;
    - Re-distributing resource from other work-streams (and managing loss of resource from donor work-stream);
    - Deploying new resources, if proven necessary.
- Process improvement is likely to yield the most cost-effective solutions for demand capacity management in the ED setting.
- Analysis of historical data enables service planners to estimate and meet likely shifts in demand due to service changes.

Flexing Capacity

- Flexible resource allocation across work-streams enables closer demand capacity matching. ED examples include:
  - Nurses working across shared care areas, helping colleagues cope with unexpected demand surges e.g. a high-dependency patient.
  - Experienced ED doctors who can re-deploy across all work-streams according to demand. Senior doctors are also likely to have shorter patient assessment cycle times compared to doctors in training, resulting in increased process efficiency.
- A constant balance must be achieved between competing streams for flexible resources through good monitoring and management. Flexing capacity will not compensate for inadequate overall system resources.

Demand Capacity Matching in the ED
Pitfalls to Avoid

EMERGENCY MEDICINE

- Lack of a clearly identified goal for the system, so that processes are not understood in the context of the overarching goal and are not continuously improved to achieve the system's goal;
- Using poor ICT as an excuse not to start to understand and manage ED demand;
- Failing to accommodate natural variability and/or minimise artificial variability in ED processes;
- Wasting time trying and resources tackling non-bottlenecks;[5]
- Pushing too hard with inadequate capacity to match demand, leading to systems failure and potential harm to both staff and patients in the ED setting.[1]

Matching Demand Capacity

Who needs to know about demand capacity management in the ED?

- ED Clinical Operational Groups (EDCOG). Team awareness of ED demand should be increased through displaying and discussing ED data. Non-random variability should be minimised through use of protocols, SOPs etc.
- Hospital management team
- Hospital Group Unscheduled Care Network

Why? To improve quality, timeliness and value in ED patient care.

Where? The EMP recommends that demand capacity analysis is an integral component of ED COG meetings and should occur in all Emergency Care Networks.

How? Start now using a basic sampling approach and available data. Roll-out of EDIS and EMP implementation will support this work.

Key Learning Points

EMERGENCY MEDICINE

- ED teams can use demand and capacity analysis to manage and plan services.
- ED demand is variable but predictable.
- ED processes must accommodate random or natural variation but non-random artificial variation should be minimised.
- Server utilisation of > 85% is likely to result in queues.
- Improvement efforts should focus on true bottlenecks.
- Improving process efficiency is likely to provide the most cost-effective solutions for demand capacity management in the ED setting.

References

1. Higginson I et al. Demand and capacity planning in the emergency department: how to do it. Emerg Med J 2010.
2. Koelling CP, Eitel D, Maapatra S, Messner K et al. Value Stream Mapping the Emergency Department. Available at http://www.iienet.org/uploadedFiles/SHS/Resource_Library/details /180.pdf. Accessed 28th March 2011.
3. Institute for Healthcare Improvement Cambridge Massachusetts Innovation Series 2003. Optimizing Patient Flow Moving Patients Smoothly Through Acute Care Settings. Available at http://www.ihi.org/IHI/Results/WhitePapers/OptimizingPatientFlowMovingPatientsSmoothlyThroughAcuteCareSettings.htm. Accessed 21st March 2011.
4. Slack N, Chambers S, Johnston R, et al. Operations management. Harlow, UK: Pearson Education Ltd, 2004 in Higginson I et al. Demand and capacity planning in the emergency department: how to do it. Emerg Med J 2010.
5. Goldratt E, Cox J. The Goal A Process of Ongoing Improvement. 3rd Revised Edition 2004. North River Press Publishing Corporation, Great Barrington MA01230.
6. Jody Crane, Chuck Noon. The Definitive Guide to Emergency Department Operational Improvement 2011. Productivity Press, Taylor & Francis Group. New York NY 10017.
7. Perla RJ, Provost LP, Murray SK. The run chart: a simple analytical tool for learning from variation in healthcare processes.BMJ Qual Saf 2011;20:46-51
8. The Improvement Guide: A Practical Approach to Enhancing Organizational Performance. 2nd ed. The Jossey-Bass & Management Series, San Francisco Ca: Jossey-Bass Publishers; 2009.