

## QUALITY IMPROVEMENT Systems engineering case study: Improving patient safety and system productivity by reducing lead-times for blood tests

Author: Kate Silvester<sup>A</sup>

### ABSTRACT

Local case reports highlighted safety issues arising from the delays for inpatient blood tests. A multidisciplinary team used systems engineering techniques to understand the dynamics of the end-to-end process and how policies, which benefited the efficiency of individuals and departments, compromised the safety and effectiveness of the whole hospital system. Upstream services were synchronised so that samples arrived in the lab earlier in the day. The laboratory process was then redesigned to meet the variations in hourly workload. Average laboratory lead-times for inpatients reduced from 1 h 58 min to 1 h 29 min with less variation predictability of lead-times. Most results were available by 10.30am rather than 12.00pm and clinical decisions were based on *today's* test. Removing the batching and prioritisation policies did not compromise the service for accident and emergency, general practice or outpatient and freed resources to address further delays up and downstream.

**KEYWORDS:** System, engineering, constraint, blood tests, improvement

### Introduction

Clarkson and Dodds summarised the development of the quality improvement in the NHS.<sup>1,2</sup> In order to improve the impact of the changes at macro-system level, ie hospital and health economy, they propose that system engineering, as applied to other complex systems eg telecommunications and global supply chains, should be used to diagnose and focus on the constraints within the macro-system and to coordinate and synchronise the improvements by the individuals and departments concerned.

This paper recounts how 'hard' systems engineering (involving manpower and machines) was used to bridge the gap between the current and required performance of the process for blood tests. In improving the previously unrecognised delays for blood tests, we uncovered the 'soft' system or 'simple rules' that govern the individuals and departments and these need to be understood and changed too.

Author: <sup>A</sup>clinical systems engineer, Kate Silvester Ltd, Warwick, UK

### Problem description and method

A high-level map of the reported patients' journeys through the hospital and observations of seven ward rounds revealed that many decisions made on ward rounds were based on blood tests taken up to 24 hours previously. Not only was this previously unrecognised constraint seen as a major safety issue (one patient died waiting for a gentamicin level and another required transfusion following a consultant's decision to discharge), the clinicians and managers recognised that these delays compromised the whole healthcare system as patients waiting in hospital beds for clinical decisions would delay new emergency and elective patients getting to their specialist wards.

A consultant physician led a multidisciplinary team of a porter, a junior doctor, a ward-based nurse, a phlebotomist, the laboratory manager, a laboratory technician and a laboratory scientist. The team agreed to meet weekly and follow Lean's A3 problem solving process.<sup>3</sup>

The team mapped the process from requesting the correct blood test through to acting on the result. The team members had not previously recognised the impact of their individual and departments' work-schedules on the end-to-end process. The reasons for the delays were then explored and coordinated interventions tested. An initial demand and workload analysis were performed for the laboratory part of the process and continued after workflow changes.

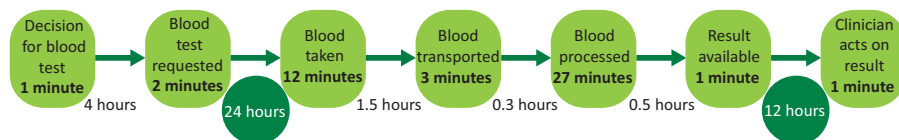
### Results and analysis

The high-level results of the process mapping and analysis for an inpatient (IP) are shown in Fig 1 and Table 1.

A cycle time is the time from a resource starting a task to being ready to start the same task for the next patient. The sum of average cycle times for each task in the end-to-end process showed that, in theory, the minimum lead-time was 47 minutes. However, the invisible delays between the tasks in the process caused by the working schedules and policies meant that the lead-time could be over 40 hours. This explained why many clinical decisions were incorrect. Clinicians assumed that they were making decisions on *today's* test, whereas their decisions were based on, at best, a blood sample taken the previous day.

The team realised that the main delays were not in the laboratory as they initially thought, and that their own workarounds didn't help and wasted both resource and time.

**Fig 1. The end-to-end blood test process.** The average cycle times taken to perform each task and the delays between the tasks were measured by the team.



They had to design an end-to-end system to ensure that results for all IP blood samples requested on the day of the ward rounds were available before the end of the ward rounds at 12.00pm. This also should not compromise time to reporting for patients in accident and emergency (A&E), outpatients (OP) and general practice (GP), important additional system elements.

The laboratory staff retrieved the data for all blood requests that included urea and electrolytes. Fig 2 shows the demand into the lab before and after the changes to the upstream ward, phlebotomy and portering processes as shown in Table 1.

The peak demand for the ward samples now arrives in the laboratory earlier between 8.00am and 9.30am.

The mean and variation in demand, based on total laboratory data (A&E, IP, OP and GP samples) after the upstream changes were made to the IP process is shown in Fig 3.

To ensure that the lab could provide the results within an hour of receiving the samples on 95% of days, we planned the lab flow capacity to meet the average plus two standard deviations of the

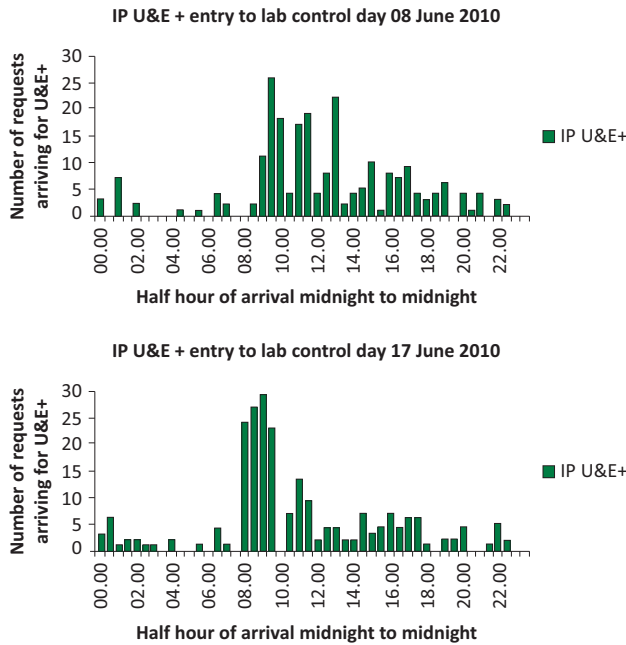
half hourly demand. The period of highest demand is 2.00pm ie 100 samples / 30 minutes, or one sample every 18 seconds and this is when samples from GP and OP clinics start to arrive. In the morning, the rate of demand is half the afternoon rate ie one sample every 36 seconds.

The effective cycle time in seconds is calculated for each step in the laboratory process and revealed two flow constraints (Table 2). First, the sorting machine (track) can't meet the afternoon demand of one sample every 18 seconds. Its main purpose is to sort the big batches from the centrifuges so that the samples can be put in order of their priority into the two analysers. It also receives and caps the processed samples returning from the analysers prior to storage. The immediate solution was to free up the sorting machine by scheduling the capping-off to after 8.00pm and delivering samples directly to the analysers. The second constraint was centrifugation. The best effective cycle time that the three centrifuges in parallel could achieve is 6.25 seconds, and therefore this is not a resource constraint. The

**Table 1. Reason for delays, consequences of work-arounds and the interventions required**

Delay	Reason	Work around	Intervention required
4 hours delay to making the request.	Not enough time to complete blood forms on the ward round, so forms completed after the 4-hour ward round.	Junior doctor leaves ward round to complete form (and take blood) missing vital decisions regarding patient care. The alternative is to request tests the day before – many of which may not be correct or necessary on the following day.	Forms available on the trolley and completed by any doctor on the ward round (including the consultant when necessary). Forms dated with the required date of the test.
24 hours delay to blood taken.	The phlebotomists' rounds happened while the ward round happened. So, any blood test requests made after the ward round were taken 24 hours later.	Doctor taking blood takes 12 minutes: 10 minutes to find the kit and gloves with many interruptions. So, doctor asks phlebotomists to take 'urgent' samples.	Phlebotomists start at 7.30am rather than 8.00am to get the samples to the lab. Every bay on every ward has a fully stocked trolley for taking blood and putting up drips.
1.5 hours delay to transport.	The phlebotomists (and other clinical staff) left specimens in the ward out-tray for the porters rounds which occurred every 1.5 hours.	Doctors and nurses phone the porters to take urgent samples to the lab. Porters prioritise these calls.	Between 7.30 to 9.00am two porters 'chase' the six phlebotomists from ward to lab and back to next ward so that smaller batches are delivered continuously to the lab.
20 minutes delay to processing.	Big batches of work arriving in the lab to be registered.	Lab technicians prioritise 'urgent' samples from A&E over ward, OP and GP samples.	Six technicians continuously register samples from all sources as they arrive.
Half hour delay to verification and release.	Lab scientist looks at the results every half hour before releasing a batch electronically onto the clinical info system.	Lab staff stop work to search for 'urgent' requests and results.	One lab scientist continuously reads, verifies and releases results from 8.00am to 12.00pm.
12 hours delay to action.	Results released in the afternoon are acted on the next morning.	Night staff and out-of-hours GP called urgently to deal with unexpected and dangerously abnormal results for patients they don't know.	All samples and results cleared and released by 4.00pm so requesting doctors can act on the result.

A&E = accident and emergency; GP = general practitioner; OP = outpatient



**Fig 2. The daily pattern of arrivals into the lab before and after the interventions upstream.** IP = inpatient; U&E = urea and electrolytes.

problem here was the queue-batch-queue policy that creates a time-trap for the samples. To cope with the increasing demand, the lab had replaced several smaller centrifuges with the three larger ones and the policy was to wait until each centrifuge was full of its maximum load of 32 samples before starting it. This meant that in the morning, the first sample of the batch of 32 could wait nearly 20 minutes for the centrifuge to start, and the afternoon it could wait nearly 10 minutes. To mitigate the impact of these delays, one of the three centrifuges was reserved for the urgent samples from A&E and this machine was run irregularly and intermittently according to unclear rules. So, the rest of the samples now competed for the remaining two centrifuges and urgent samples from the A&E and IP wards taking priority over OP and GP samples. This time-trap at the centrifuges could be reduced by starting each of the three centrifuges regularly at

3.33 minute intervals with however-many samples were waiting. It also became apparent to the lab team that they could free up some of their resource in the morning to join the phlebotomy and portering team on the wards, thus improving the flow into the lab without compromising the flow through the lab (Fig 4).

Samples are grouped by source (IP, A&E, GP and OP) and ordered along the x-axis by their authorisation date and time (result available to clinicians). The mean is shown as a green line and the red lines show the range of variation either side of the mean (mean  $\pm 3$  sigma).

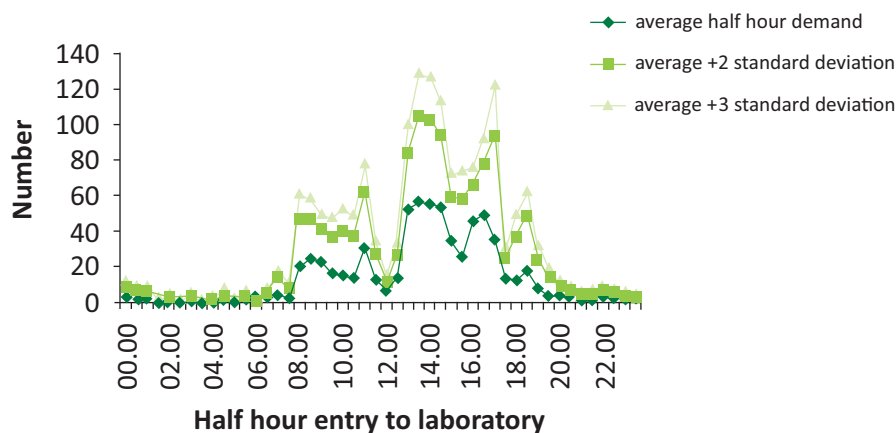
The white dots are for blood results that have only been released the day after the sample was received. The raw data reveals that these samples arrived in the lab during the night shift (8.00pm to 8.00am) and were not deemed urgent by the clinical or lab staff or required transfer for unusual or more complex testing at the central laboratory 14 miles away (eg gentamicin). These results are clearly from a different system (outside mean +3 sigma) and have been excluded from the statistical analysis for those samples arriving into the lab during the normal working hours 8.00am to 8.00pm.

The summary tables show that the mean and sigma (variation) in laboratory cycle times for IP samples reduced indicating that the IP system was more responsive and predictable, with negligible impact on the samples from A&E and GP, but the mean and variation for lead-times for those from OP also improved.

Though improved, there is still variation in flow through the blood test system that needs to be addressed. The red dots indicate 'special causes' ie statistically significant changes in the pattern of behaviour of the blood testing process demonstrating that it is not yet stable and resilient to the numerous other known and unknown causes of variation. For example, lab flow remains compromised at lunchtime, when the bulk of GP and OP samples arrive at 2.00pm and when the laboratory scientist authorising the results can only just cope with the peak flow rate in the afternoon.

### Discussion

As in many hospitals and healthcare systems, all the individuals and departments involved the blood test process had made changes to improve departmental efficiency and unit cost (cost/activity). Creating a multidisciplinary team to map the end-to-end process revealed the previously hidden delays for bloods tests and the impact on patient safety. For the first time,



**Fig 3. The variation in total demand into the laboratory.** Half hourly demand profile entry to laboratory for 7 days; June to August 2010.

**Table 2. The laboratory cycle times and flow capacity (manpower and machines) available**

Laboratory process step	Average cycle time: time from the start of one specimen to the start of the next specimen at the same process step.	Batch quantity	Capacity available	Effective cycle time in seconds
Sorting	5 seconds	1	One person	5
Log-in	90 seconds	1	Six people and six computers	15
Centrifuge	10 minutes	32	One person managing three machines (each with a max of 32 samples)	6.25
Sorting machine (track)	20 minutes	32	One person and one machine	37.5
Analyser	30 seconds	1	One person and two machines	15
Verification	20 seconds	Policy of only checking results every 30 minutes	One person and two computers	20

staff involved in the day-to-day, detailed work, both inside and outside the laboratory, could ‘see’ the bigger system in which they worked and relate the delays to their own relatives’ experiences as patients of this system. The team’s manual measures of the end-to-end process for two patients provided the emotional energy for a more ‘joined-up’ approach (see problem description). All staff appreciated the rational and blame free systems engineering approach to finding and focusing on the main constraints in the system rather than the ‘ad hoc’, piecemeal approach to improvement. Plotting individual

patient data eg lead-times on control charts to show the flow dynamic was familiar to the laboratory staff using Shewhart’s charts to monitor performance of laboratory equipment.<sup>4</sup> This statistical technique was new but comprehensible to the clinical staff who use time series data to monitor every patient’s physiology. However, it was revolutionary to the managers who only received monthly reports in which individual patient data are compressed to averages or to the percent achieving arbitrary lead-time targets and then compared to other laboratories in different healthcare systems.

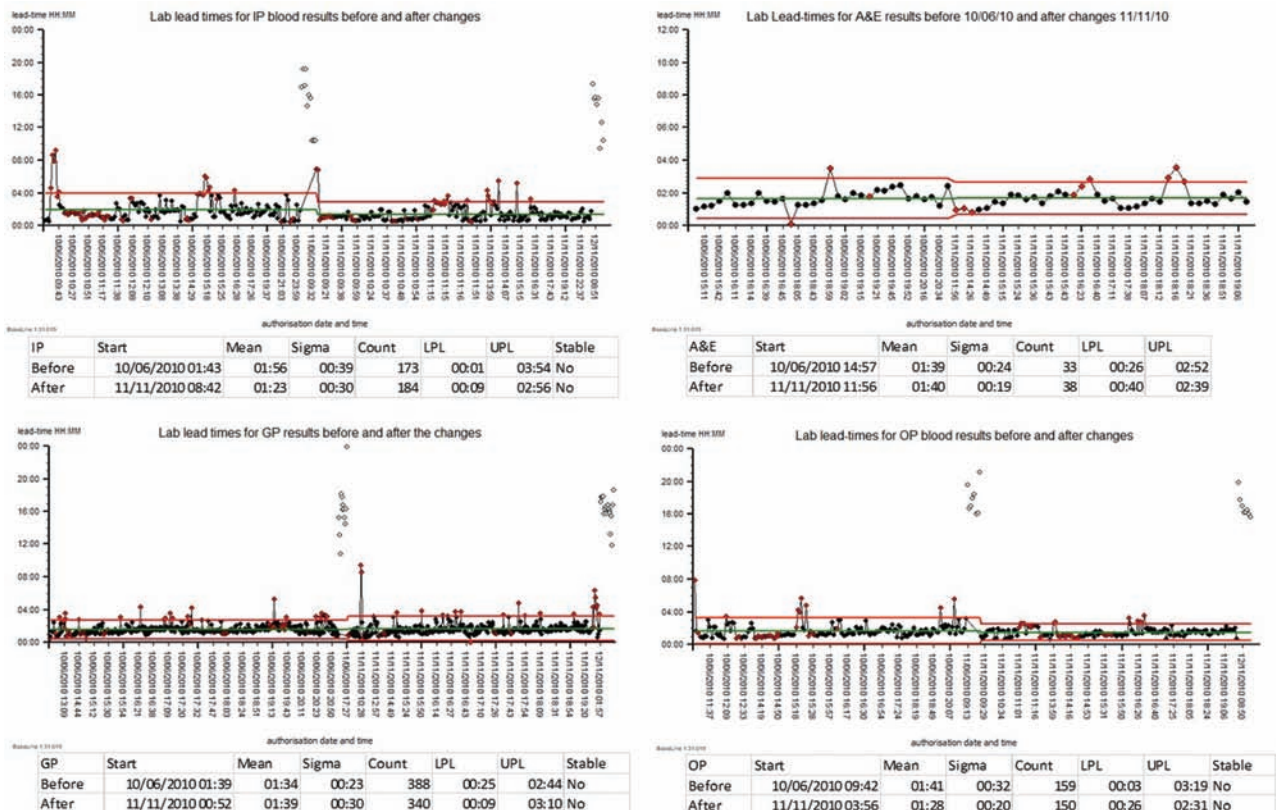


Fig 4. Laboratory lead-times HH:MM for blood chemistry requests before and after the changes.

For the first time, all the team members could see how their pattern of daily demand and workload could be made visible and analysed to ensure they had the right amount of resource time available to meet the variations in hourly demand throughout the day. This is routine work for systems engineers in industry and provides the missing link between improvement and design. Many improvements in healthcare are driven by what individuals and departments believe is possible (improvement), not what the end customer (ie patient) of the over-all system requires (design).

The multidisciplinary systems engineering approach also uncovered the 'soft' system ie the rules or policies that governed the way individuals and departments worked. The original change from smaller to larger centrifuges and simple rules, such as 'Keep all resources 100% utilised', illustrate the underlying 'economies of scale' thinking that aims to optimise the efficiency (unit cost) of every resource in the system. However, the resulting batches and prioritisation compromise the flow and yield (productivity) of the whole system. Once this is understood, these 'rules' can be challenged and adjusted to ensure that the overall system's purpose is achieved (right care; on time, every time; in full for every patient). Systems engineers have learned to balance each resource's efficiency to optimise the macro-system's productivity. This 'economy of flow' thinking is described in the *Theory of Constraints* and is familiar to chemists.<sup>5</sup> Somewhere in any system is the constraint – the rate limiting step that governs the flow through the system. First, we need to ensure that the constraint can meet the maximum rate of demand and then there is no point of any other resource working faster than the constraint – they will just create queues of work that waste resources in storage and prioritisation. Instead, spare resource can be freed-up to improve flow at the constraint (if necessary) and used to reduce variation thus making the process more predictable.

Just as the discipline of medical practice is generalisable from one patient to another, so is the discipline of system engineering but the findings and interventions are system specific. This highlights the dangers of policies that aim to spread changes from one system to another.

There were limitations within our system too. The source of the sample was recorded, and this showed that the intensive treatment unit patients' samples, taken by the clinical staff,

arrived in the lab before 8.00am. The phlebotomy/portering team tried to match their schedule to the ward rounds but this wasn't possible as a) the ward rounds were consultant and not ward specific and b) patients not due a consultant review could still require daily review of their blood results by other clinical staff.

Only the lead-times within the laboratory could be shown as the times the sample are requested and taken are rarely recorded on the paper-based system. The team did manage to track, manually, the total lead-time from request to result being available for two patients and were shocked by the results (see problem description and method).

Considering the impact on patient care from the local case reports, the team wanted to demonstrate the impact of improved lead-times for blood tests on hospital mortality (HSMR) and efficiency (number IP episodes / number of beds). However, the delays for IP blood tests was only one of the many constraints in the hospital system that were addressed and the reduction in HSMR from 121 to 98 and the 11% improvement in IP efficiency couldn't be attributed to this change alone.<sup>6</sup> ■

## References

- 1 Clarkson J, Dean Ward J, Komashie A, Bashford T. A systems approach to healthcare: from thinking to practice. *Future Healthcare Journal* 2018;5:151–5.
- 2 Dodds S. Systems engineering in healthcare – a personal UK perspective. *Future Healthcare Journal* 2018;5:160–3.
- 3 Shook J. *Managing to learn: Using the A3 management process to solve problems, gain agreement, mentor, and lead*. Lean Enterprise Institute, 2008.
- 4 Tennant G. *Six sigma: SPC and TQM in manufacturing and services*. Gower Publishing, 2001:6.
- 5 Goldratt E. *Theory of Constraints*. North River Press, 1990.
- 6 Silvester K, Balfour J, Blackley J et al. *Unblocking a hospital in gridlock*. The Health Foundation, 2013.

**Address for correspondence: Dr Kate Silvester, Kate Silvester Ltd, 33 Dorsington Close, Hatton Park, Warwick, Warwickshire CV35 7TH, UK.  
Email: [kate@katesilvester.co.uk](mailto:kate@katesilvester.co.uk)**