Designing ophthalmology services Part 3: How do we address the queues post-COVID-19?

BY MARTIN DOWDALL, CASPAR GIBBON AND KATE SILVESTER

There is going to be enormous demand on ophthalmology services as they start to welcome patients back. The authors explain how modelling can help make the most of the available resources.

If planning the resources required to meet the new and follow-up demand under normal circumstances isn't testing enough, ophthalmology services now face the challenge of restoring their services when the COVID-19 lockdown is lifted.

Wouldn't it be wonderful if we had 'a dynamic map', a visual 'model' of our service showing the resources in the system and the patient flows between them? Many other industries (e.g. transport, power generation) make their invisible system's past performance visible (retrospective) and use (predictive) models to plan for different future scenarios [2].

Making the complex adaptive behaviour of the whole ophthalmology service visible is a tough ask – especially for beginners. So where do we start?

Mapping the system

A high-level, macro-system map of the ophthalmic service is key to establishing a common understanding for the stakeholders within it.

Figure 1 shows the key resources in the system (boxes) and the boundary of the hospital's ophthalmology service is shown in blue. The 'No Entry' signs indicate that flow into this resource was closed during the COVID-19 lockdown. Some of this demand was diverted into the red 'emergency / triage sub-system'. The queues generated during the COVID-19 lockdown are shown in red letters and those present pre-COVID are shown in black.

When the COVID lockdown is released, we need to deal with the following:

- **a** is the queue of new disease in the community that was unable to access a GP or optometrist during the COVID lockdown and unable or chose not to attend A&E or Eye-Cas. These patients may be still waiting at home with unresolved symptoms and may access their GP / optometry services when these re-open.
- **b** is the work-in-progress (WIP) that was seen in EyeCAS during lockdown and needs to be followed-up in specialist clinics.
- **c** is the new demand, i.e. the 'normal' incidence of new ophthalmic presentations coming in via GPs and optometrists once these up-stream services are re-established.
- **d** is the established queue of new patients referred by their GPs and optometrists prior to the COVID lockdown but not yet seen in the clinics.
- **e** is the already established queue of follow-up (FU) patients in our service prior to COVID-19. These patients, with serious chronic pathology, will be nearing or will have passed their due dates

f, **g** and **h** are queues for interventions established before the COVID lockdown.

Before we start trying to predict what will happen, we need to understand the past performance of our system.

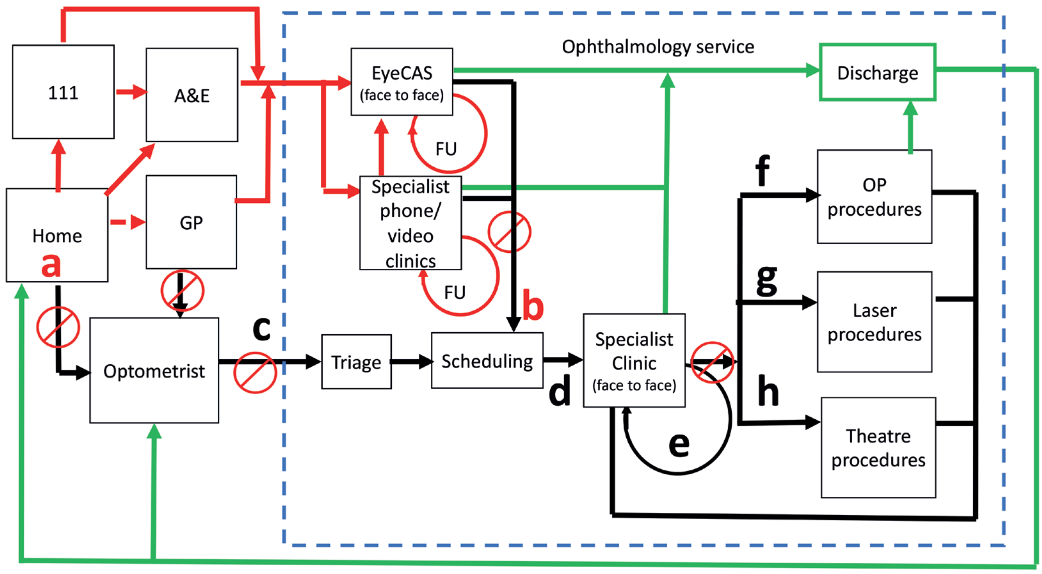


Figure 1: Systems map of an ophthalmic service to show patient flow during the COVID-19 lockdown.

Measuring patient flow

Each resource has a flow-in (demand / time) and a flow-out (activity / time) and, at any point in time, the cumulative difference between the two is work-in-progress (WIP), otherwise known as a queue. We want to show how these changed over time and the best way of showing this is as a time-series chart as in Figure 3.

To do this clinicians, managers and data analysts must address two key issues:

1. Ensure that every patient in the system concerned has two dates:
   - a demand date (request date) and an activity date (date seen) and, for those that haven't been seen yet, a due date
   - the interval between the demand date and the activity date is the patient's lead-time.
2. Ensure their data query captures all the patients in the demand, activity and WIP in their reporting window – *and here lies the underlying fault in the majority of NHS performance reports.*

Creating a time-series chart to show the patient flow – the Vitals Chart®

To understand the variation in patient flow through their service, a team requests a report to show the weekly demand, activity and WIP for the previous 32 weeks, starting at t1 and ending at t2. This is called the reporting window. So, how do we capture all the patients that were flowing through the system during this period?

The relationship between each patient's demand and activity (or due date) and the reporting period is shown in Figure 2 below.

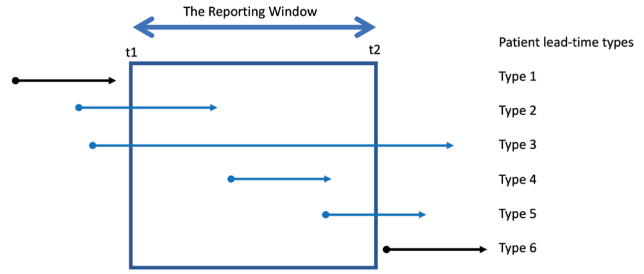


Figure 2: The relationship between the reporting window and each individual patient's lead-time starting on their demand date (dot) and ending on their activity or due date (arrow head).

- **Type 1** patients were referred and seen before the start of the reporting window (t1). We don't need to capture them as no part of their lead-time falls within our reporting window.
- **Type 2** patients were referred before the start of the reporting window (t1) and seen in the reporting window. We do need to capture them, as they are in WIP at t1 and then become activity.
- **Type 3** patients were referred before the start of the reporting window (t1) but have a due-date after the end of the reporting window (t2). These are the patients in WIP at t1 with chronic disease whose due-date may well have passed due to the COVID lockdown and we don't want to lose them.
- **Type 4** patients have both their demand and activity date within our reporting window and we want to capture them.
- **Type 5** are patients for whom we received the request within the reporting period but we haven't seen them yet. They are in WIP at t2 and we need to capture them too.
- **Type 6** are patients who will be referred and seen after our reporting window so we wouldn't expect to capture them.

'A trap-for-heffalumps' [3]

The majority of data requests in the NHS are driven by financial reporting and are along the lines, 'Please may I have a report showing all the demand and activity for the period between t1 and t2'.

One common data error made in servicing such a request only captures the patients with lead-time types 2, 4 and 5 since their dots (demand) and arrow heads (activity) fall within the reporting window in Figure 2. This type of data request captures the demand and activity, which is what the financial reports require but, in order to plot patient flow over the period of the reporting window, we also need to know the total number of patients in the system during that period of time, i.e. the WIP at t1 and t2. In this case we have missed the lead-time type 3 patients – the very patients the clinicians and operational managers need to know about!

The left-hand chart in Figure 3 below shows the impact on the Vitals Chart® of this type of data error.

To capture all the demand and activity (or due-dates) and WIP for patients with lead-time types 2, 3, 4 and 5 in Figure 2, the correct data query has to be written, 'Please may I have the demand dates for all patients referred on or before t2 and who have an activity date (or due-date) on or after t1'. Following this correct but counterintuitive data query in Figure 2 generates the correct chart on the right in Figure 3.

It is essential that clinicians, managers and data analysts understand this counterintuitive way of requesting their financial and operational data [4,5,6].

Since the real WIP is hidden from view in current performance reports, booking staff are constantly surprised by patients who appear 'out-of-nowhere' (particularly follow-ups) and, since demand and WIP are often confused, this reinforces the team's belief that the 'demand outstrips their capacity' and requires short-term overbooking and waiting list initiatives to get these patients seen on time.

Now let's imagine a world in which the clinicians, operational managers and data analysts have learned to create and update the weekly Vitals Chart® for each of the resources on their system map in Figure 1. What impact would that have on their understanding of the flow in their system and the impact of any potential changes?

The team can clearly relate to the causes of variation in demand and activity in the top chart of Figure 4, i.e. holidays and COVID lockdown. The crucial indicator of the stability of the system is the WIP. It varies a bit but it is stable over time.

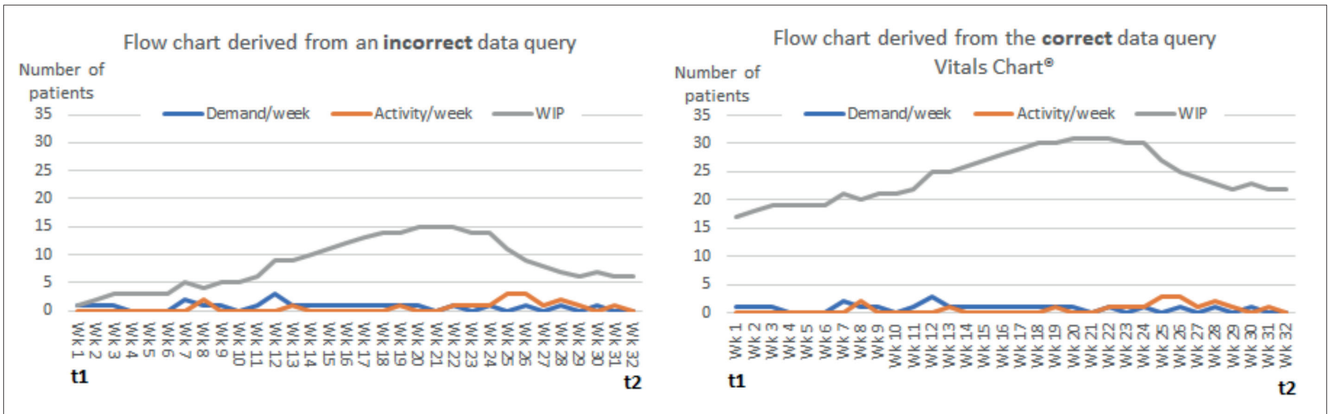


Figure 3: Incorrect (left) and Correct Vitals Chart® (right) generated from the same data set.

Modelling

Unless these Vitals Chart® data are accurate, then any subsequent modelling to predict the outcomes of scenarios, however sophisticated, will be flawed.

1. Modelling a stable system: Little's Law

If the WIP is stable over time, then the average flow in = average flow out (λ), and Little's Law can be used to predict the performance of this system because average WIP = Lead-Time x λ [7,8].

In this example, there are on average 600 patients waiting for their first appointment in the cataract clinic and the average flow is 15 patients / week. We would expect the average lead-time for referral to first assessment to be 600/15 = 40 weeks = 280 days.

The bottom chart in Figure 4 shows the lead-times plotted for consecutive patients by date seen in the clinic and shows that the lead-times are variable but there is a peak of patients seen at around 280 days. Little's Law is a valuable way of verifying the data is consistent before using it to predict the outcomes for future scenarios.

Since the COVID lockdown, there has been no demand or activity, so the WIP has stayed the same but the individual patients' lead-times will be increasing by a week, every week.

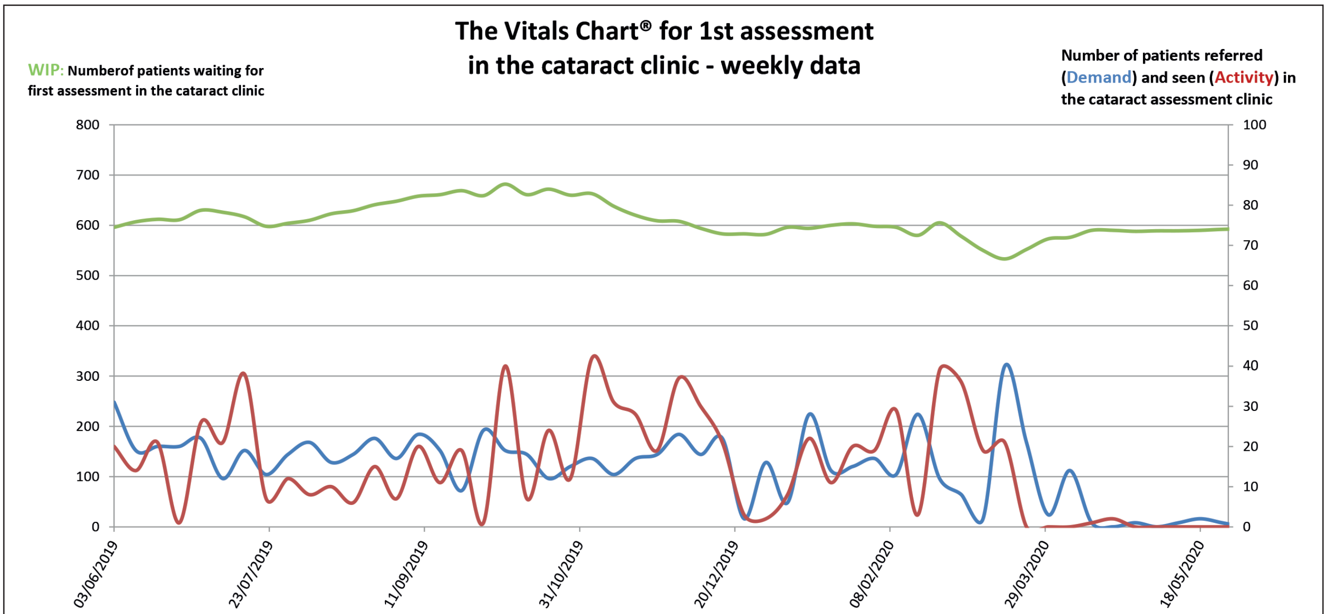
If this team want to achieve an average lead-time of six weeks (42 days) from referral to being seen in the clinic, they will need to have a WIP of 15 patients / week x 6 weeks = 90 patients. That means draining out 600 minus 90 patients = 510 patients. If they want to deal with this back-log over the next year, then they will need to assess 510/52 ~ 10 more patients / week in addition to the 'normal' average demand of 15 patients / week that can be expected once the upstream services resume.

The team can also begin to test other scenarios, e.g. 'What would be the impact on the cataract theatre downstream? What additional cataract OP activity could they perform without overwhelming the theatre resources?'

This means creating and updating the verified Vitals Chart® for every resource in the system [5].

2. Modelling the impact of variation in an unstable system

Little's Law is very useful but it is based on averages and doesn't take into account the impact of variation in demand on the required capacity to achieve a stable system. This is because it is not possible to save unused resource-time from a period when the demand was less than average for one when the demand is greater than average [9,10]. As a consequence, we need a little more resource-time capacity, on average, to achieve the required activity and make the



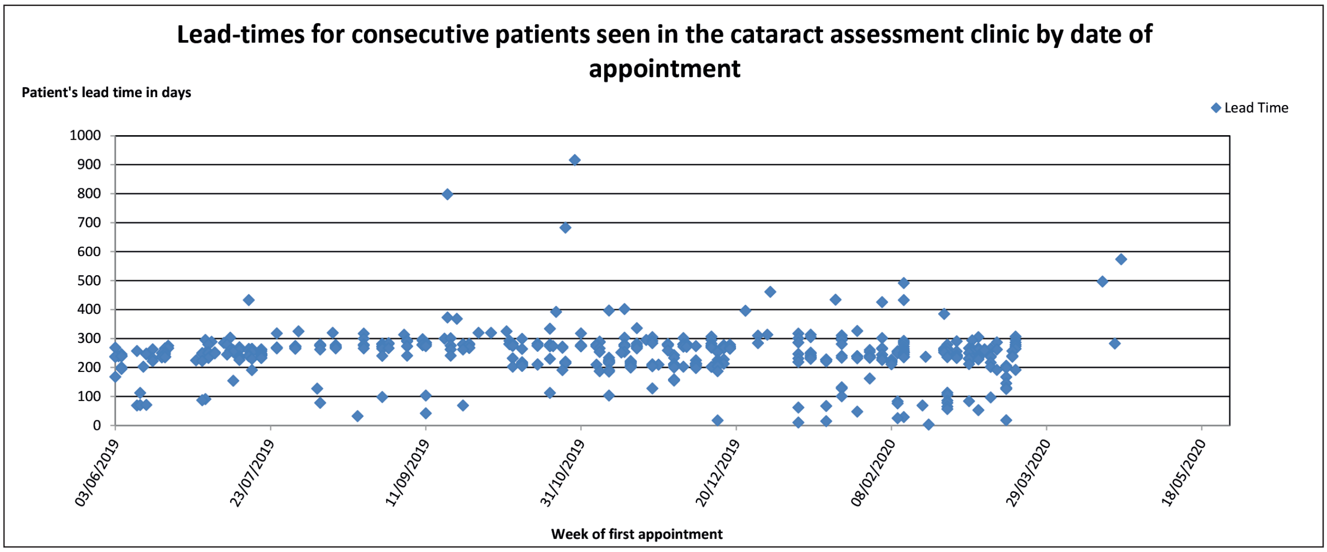


Figure 4: The weekly Vitals Chart® for first assessment in the cataract clinic (courtesy Northern Devon Healthcare NHS Trust).

system resilient to variation. The Vitals Chart® data can then drive more sophisticated 'stock-and-flow' models to predict the required resource-time capacity to meet the lead-times predictably [11].

Clinical teams can then measure the cycle-times for each task involved in the process and calculate the weekly workload at each resource (required weekly activity x cycle time) and the manhours required / week [8]. Some cycle-times may increase post-COVID due to the increased infection control measures and the team can compare the required with the available manhours / week.

Gantt charts are one way of testing scenarios involving average cycle-times [8,12] and discrete event simulations (DES) allow the variation in multiple interacting factors to be taken into account [13].

Opportunities for improving flow within our current resources

Once we understand the required resource time to meet the expected demand and drain out any WIP, then there is an opportunity to free-up resource time by reducing the 'carve-outs' in the pre-COVID schedule and resources [6,14].

For example, the glaucoma service described previously [14] had four separate clinics based on the urgency and source of referral. This means that the inherent variations in the demand for each sub-clinic results in some clinic slots being left empty and other clinics overbooked [15]. However, since clinical process and equipment required for all glaucoma patients is the same, could these clinic resources be pooled?

Conversely there is a real danger of carving-out more resource, e.g. for 'COVID positive patients' and this will compromise the service still further.

Summary

In the COVID-19 pandemic, we have learned just how quickly we can redesign services to meet changing demand and cycle-times. We also learned that predictive modelling is constrained by the initial 'estimates' of demand [16]. We need a model that is updated in real-time by a feedback loop with the measures of the actual flow (demand, activity, WIP and lead-times) [17]. This means correcting the underlying data query to capture all the patients in the system. Teams that have Vitals Charts® updated in real-time now have the ability to use their implicit understanding of their clinical processes to keep the patients in their care safe from the harm of delay [5,18]. They then have to opportunity to use the same data in more sophisticated models to predict what is likely to happen in response to changing circumstances.

TAKE HOME MESSAGE

- Service redesign is an iterative process that requires teams to:
  - visualise the flow through their system and to recognise that: 'A system is only as good as its feedback loop' Gregory Bateson 1904-1980 [1]
  - feedback loops are updated weekly time series charts that show the demand (flow in), activity (flow out), the WIP (work-in progress, queue, waiting list) and patients lead-times at each resource in their system,
  - quickly understand and correct the ubiquitous error in the data queries which currently underestimate their waiting lists.
- Only once the flow data are correct, can mental arithmetic (e.g. Little's Law) or more sophisticated 'stock and flow' and discrete event simulations be used to predict what will happen when changes are made upstream or downstream.
- Then staff can use their clinical knowledge and initiative to redesign their services and provide the resource-time capacities required to keep their system resilient and meet their patients' lead-times safely.

References

1. Wikipedia: Gregory Bateson: https://en.wikipedia.org/wiki/Gregory_Bateson
2. World Economic Forum. This interactive map shows US electricity usage in real time: www.weforum.org/agenda/2016/09/united-states-electricity-usage-map-real-time
3. AA Milne: *Winnie-the-Pooh* (first published by Methuen and Co Ltd. 1928).
4. Silvester K. Diagnosing the Flow Constraint in an Endoscopy Service. Part 1: Recognising and Avoiding the Data Query Trap. *Journal of Improvement Science* 2016;**38**.
5. Talbot L. Visualizing the Referral to Treatment Time (RTT) Waiting List for Ophthalmology Patients. *Journal of Improvement Science* 2020;**66**.
6. Wright L. Diagnosing Long Waits in Radiology. *Journal of Improvement Science* 2017;**50**.
7. Little JDC: A Proof for the Queuing Formula: L = λW. *Operations Research* 1961;**9(3)**:383-7. Little's law: https://en.wikipedia.org/wiki/Little%27s_law
8. Silvester K. Designing Ophthalmology Services Part 1: How do we address the queues in a clinic? *Eye News* Online Exclusive: www.eyenews.uk.com/features/ophthalmology/post/designing-ophthalmology-services-part-1-how-do-we-address-the-queues-in-a-clinic
9. Wikipedia: Agner Krarup Erlang: https://en.wikipedia.org/wiki/Agner_Krarup_Erlang
10. Silvester K, Lendon R, Bevan H, et al. Reducing waiting times in the NHS: Is lack of capacity the problem? *Clinician in Management* 2004;**12(3)**:105-11.
11. Silvester K. Implementing Endoscopy Pre-Assessment: Part 2. *Journal of Improvement Science* 2016;**36**.
12. Pirie E. Diagnosing and Treating a 'Sick' Complex Multidisciplinary Clinic. *Journal of Improvement Science* 2019;**63**.
13. Dodds S. A Case Study of a Successful One-Stop Clinic Schedule Design using Formal Methods. *Journal of Improvement Science* 2012;**6**.
14. Silvester K. Designing ophthalmology services Part 2: How do we address the queues for a clinic? *Eye News* 2020;**26(6)**:32-7.
15. Adams A. Diagnosing Flow Constraints in a Fast Track Glaucoma Clinic. *Journal of Improvement Science* 2019;**61**.
16. "A terrible toll". *The Economist* 23 May 2020:81.
17. Dodds S, et al. An Acute Hospital Demand Surge Planning Model for the COVID-19 Epidemic using Stock-and-Flow Simulation in Excel: Part 1. *Journal of Improvement Science* 2020;**68**.
18. Jones C, et al. Improving the Delivery of Chemotherapy: Part 7. Implementing the Feedback Loops to Complete the 6M Design® Cycle. *Journal of Improvement Science* 2020;**64**.

(All links last accessed June 2020)

Website: Links to the *Journal of Improvement Science* and the Health Care Systems Engineering programme can be found at www.improvementscience.uk

Visit our website or scan QR codes for the previous articles in this series.

Scan for Part 1

Scan for Part 2

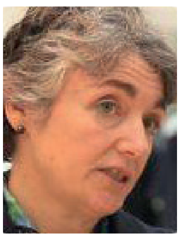
AUTHORS

Martin Dowdall MSc MMACP Grad Dip Phys,

Senior Manager, Service Transformation Team, Northern Devon Healthcare Trust.

Caspar Gibbon BM FRCOphth,

Consultant Ophthalmologist, Northern Devon Healthcare NHS Trust.



Kate Silvester BSc MBA FRCOphth,

Coach, Healthcare Systems Engineering Programme; co-founder, Journal of Improvement Science (JOIS); Honorary Associate Professor, University of Warwick.

Declaration of competing interests: None declared.