Editorial

Demand–capacity modelling and COVID-19 disease: identifying themes for future NHS planning

J. J. Pandit

Consultant, Nuffield Department of Anaesthetics, Oxford University Hospitals NHS Trust, Oxford, UK

Correspondence to: J. J. Pandit
Email: jaideep.pandit@sjc.ox.ac.uk
Accepted: 18 May 2020
Keywords: health services research; efficiency; modelling; COVID-19

This editorial shows how the principles of demand–capacity modelling lead to the conclusion that COVID-19 disease will require the UK NHS to run its services at a far higher working capacity (in terms of hospital, intensive care and social care beds, operating theatres and staff) than it has been used to. The analysis leads to key questions that should form the basis for rational planning in a post-COVID NHS.

Demand–capacity modelling in the NHS: a brief history

How much capacity does a hospital, or an individual surgical service need, to manage its demand? The correct answers are readily provided by several easily accessible tools designed to guide planning [1–3]. ‘Capacity’ is a combination of both physical (space, equipment, infrastructure, etc) and functional factors (staff numbers, and time for which they are contracted to work). ‘Demand’ is often measured in terms of the number of patients presenting for treatment, but more accurately is the time it takes to provide the care [3]. There is an important and consistent result of all calculations: in setting optimum capacity, variation in demand is more influential than average demand. Figure 1 shows, greatly simplified, that with a consistent arrival of 100 patients daily, each requiring a bed, then 100 beds are needed as capacity. Providing <100 results in a backlog or ‘waiting list’ for treatment. As demand becomes more variable, the required capacity increases, despite the average demand being the same. For example, if between 80 and 120 patients arrive daily, then the same 100 bed capacity is insufficient for half the time (and on the remaining days there is excess, wasted bed capacity). Clearly, a capacity of >100 beds is needed to capture most or all of the demand, albeit at the cost of wasted bed capacity some of the time. These simplified calculations yield the same result if the quantity ‘number of patients’ is converted to ‘time required for treatment’ and ‘number of beds’ converted to ‘staff hours contracted to provide care’ (some of the models use sophisticated ‘queuing theory’) [4].

Yet, the NHS has historically rejected these perfectly accurate calculations and chosen instead a different path: that of focusing on ‘flow’ of patients. In the very simplified example above: if 100 elective patients arrive per day, then the NHS will provide staffing for only 50 beds; arranging 50 arrivals in the morning and 50 in the afternoon, with a focus on more rapid turnover [5]. More detailed modelling that has been applied to real situations of course takes into account factors like variable arrivals, different durations of treatment or bed stay, and delays to discharges [1–5]. Figure 1 and the outline above offer only the simplest of illustrations. This cornerstone of NHS planning for decades is felt desirable due to the considerable cost savings in infrastructure and staff. The result is a decline in inpatient beds (see also Supporting Information, Figure S1), fall in staff numbers (as a proportion of work done; see also Supporting Information, Figure S2) and closure or merger of hospitals [6]. It is argued to have resulted in a leaner (i.e. lower capacity) and more ‘efficient’ (i.e. better utilised) system, with less waste. The philosophy is also seen in ‘just-in-time’ ordering for equipment and consumables [7].
Instead of stockpiling needles, sutures, clothing, etc ('just in case'), only the estimated bare minimum are held on site and regular ordering ensures adequate levels are present on any given day. Storage costs, space and waste (through time-expired products) are all reduced in a 'lean' way.

The 'patient flow' approach requires very careful systems planning and it is a matter of debate whether it has been successful [8, 9] (see also Supporting Information, Figures S3 and S4). For example, patients ultimately need to flow out of hospital and the reduction in capacity in social care has been identified as a potent rate-limiting step, leading in turn to 'bed blocking' or 'delayed transfers of care' [10] (see also Supporting Information, Figure S5).

The changes forced by COVID-19 disease

The fundamental weakness of this 'flow' strategy has been laid bare by COVID-19 disease. In modelling terms, the impact of the acute phase pandemic has been to increase mean/total demand on services. When faced with the possibility (realised in all-too-horrible scenes in many countries like Italy) that thousands if not millions of patients would in a short space of time present to hospitals with a range of COVID-related symptoms – on top of the regular presentations for non-COVID treatment – the realisation rapidly dawned that no 'optimisation of patient flow' policy could possibly work. In graphical terms (Fig. 1), the peak demand would be off the scale (what the Prime Minister referred to as a 'sombrero' picture of demand). The result could have been complete collapse of the NHS with an uncontrollable rise in all deaths. At this point, the 'patient flow' focus was abandoned. There was no choice except to create actual physical and functional capacity [11, 12] (see also: https://www.kingsfund.org.uk/publications/critical-care-services-nhs). This was achieved in three main ways: a reduction/cessation of elective work (and even some emergency work, such as transplantation); physical infrastructure build (e.g. Nightingale hospitals); and staff recruitment (e.g. bringing back recently retired healthcare staff). Moreover, 'demand' was managed by asking people to stay at home (the lockdown policy 'flattening the sombrero'), and also by more explicit advice on selecting patients for escalations of treatment based on age and frailty criteria [13].

There are sporadic reports of some hospitals being temporarily overwhelmed, but this is to be expected as there is inevitably geographical–temporal variation in patient attendances (underlining the principle that variation is the all-important metric). The strategy of increasing real capacity has worked, reflected in acute bed occupancy being just 59% (compared with a norm of > 90% in many hospitals; see also Supporting Information, Figure S6) and critical care occupancy just 78% [14]. This is the very purpose of 'capacity'; that a proportion remains unused (Fig. 1).

How do we rationally set capacity?

Although the modelling can be quite mathematically detailed, the principles can be understood by reference to Fig. 1, and by introducing simple probability theory. Any capacity less than the peak of demand (for the variable demand conditions shown) will yield a certain probability of being able to meet the total demand. If, say, the mean demand is being measured as 100 h of service, and at its peak is 200 h, then it follows the setting that a capacity of exactly 100 h yields 50% probability of meeting demand; setting a capacity of 200 h offers 100% probability, and setting a capacity in between 100 and 200 h yields a probability between 50 and 100%, depending exactly where within this range it is set. Setting capacity at 100%
probability of demand capture is very expensive and may not be attainable. It has been demonstrated, in various ways, that setting the capacity to at least c. 80% of the maximum peak deviation of demand from the mean value (in this example at ~180 h) will have a corresponding c. 80% probability of capturing the demand overall, and offers a reasonable balance. The technical explanation lies in probability density functions of different data distributions being similar at this point (further explained in reference [3]). Giannakeas et al., used sophisticated modelling that included estimates of population-weighted and age-stratified probabilities of COVID-19 cases requiring acute care hospitalisation, coupled with estimated length of stay data; nevertheless, the estimated probability of capturing the demand for the set critical care capacity in Ontario, Canada, was 78.6% [12]. Using such approaches, planners can set the capacity at the 80%, 85% or whichever probability level they choose; the point is they know quantitatively where they stand.

What happens after COVID-19?
At some point the number of new COVID-19 cases will decline from 4000 a day to 400 a day, then 40 a day and so on. The dilemma is what to do with all this fresh, unused capacity?

One financially tempting option might be to go back to the ‘optimise patient flow’ strategy: close all the beds again and dismiss all the staff. This might temporarily help the national deficit, ballooned by Chancellor Sunak’s recent but necessary generosity, but COVID-19 will never go away. Even if new cases are down in the future, or so quickly rise back up to 4000 (as second or third waves), a single, overlooked COVID patient in a ward could rapidly spread the disease to other patients, and to staff, and lead us back to the very crisis we sought to avoid. The high NHS bed occupancies (see also Supporting Information, Figure S6) have been identified as a concern in infection spread [15]. Expressed another way: if a COVID-19 incidence of 40 per day justified national lockdown and economic meltdown ‘on the way up’ as a bad news story, why should the same incidence of 40 per day be viewed differently ‘on the way down’, as a good news story?

Referring back to Fig. 1, it is essential to understand the post-COVID world in terms of demand–capacity modelling. For years or decades to come, COVID’s key impact will be to increase the overall variability of hospital attendances and demand for NHS services (regardless of whether there are actually second or third ‘waves’ of pandemic). COVID-19 will do this even if its incidence becomes low and it does not markedly increase the mean demand; even at low incidence it will affect variability of demand. The mathematical proof of this lies in an equation called the ‘logistic map’ (Appendix 1), which is widely used to model population growth [16,17], synonymous with disease spread, for viral populations [https://arxiv.org/ftp/arxiv/papers/2003/2003.05681.pdf]. Although the initial growth/spread of virus is initially exponential, the logistic map shows that later, population size/infection rates become cyclical. This variability is greatly affected by very small changes in the underlying growth rate (infectivity, or r or R0 value) that do not themselves change the mean, and in turn this is reflected in demand on services overall (Fig. 2). An accompanying article explains in more detail the terms in the logistic map equation [17].

As we have seen, the only rational way to manage this is a sustained increase in real capacity (Fig. 1). Quantitatively, the amount of capacity we will need can be estimated quite accurately using the models and calculations referred to in the first paragraph. So, we already have the tools to plan ahead [1–5]. This will entail a fundamental shift away from the ‘patient flow’ focus. Of course, optimising flow is no bad thing, and good lessons will have been learned to influence efficient practice, but this should no longer be viewed as a substitute for real capacity, in the way it has been in the past.

Key questions to underpin future planning
If rational demand–capacity matching is to be at the core of future NHS thinking, then specific questions relating to setting capacity and managing demand can be refined:

1 How do we set capacity? At all levels – hospital level, individual operating theatres, outpatients, etc – capacity planning should be undertaken using recognised models that will capture most of the demand most of the time (see section above). A corollary is that at other times, this capacity will be unused, so the NHS will have to learn to live with empty beds and underfilled operating lists. Bed occupancy rates > 90% (see also Supporting Information, Figure S6) must become a thing of the past if COVID-19 is to be contained at all.

Operating theatre capacity will have to increase for at least two reasons. One is that the additional time now taken to don and doff personal protective equipment and allow for air changes etc., all greatly add to the mean and variance of individual operating times [3]. This does not require any change to models used to schedule lists – all it does is increase the values for individual operations inputted into these models. A second is the creation of COVID vs. non-COVID care pathways (see below).
Departmental-level capacity will have to increase by ensuring adequate staff numbers. It will be impossible to deliver a safe service in the post-COVID world with anaesthesia staffing shortfalls of 10–30% as is currently commonplace [3].

2 How do we fund the increased capacity? It is well accepted that NHS funding has for years been inadequate for the rise in demand, especially in the realm of social care. Moreover, irrational funding models create perverse incentives [18]. One example relevant to anaesthesia is that current tariffs do not reward efficient planning of operating lists, but rather, encourage cherry-picking of some inherently profitable operations over others [19]. Another example is the existing taxation and pension arrangements, that make it equally profitable for individuals to sit at home as to engage in NHS work [20]. It is difficult to see how these funding models are compatible with any COVID-19 management strategy.

3 How do we manage COVID in patient pathways? The routine and regular testing of all patients and staff is common sense as part of clinical and risk management (see the accompanying article for its role in the infectivity growth rate R0 value [17]). But it also creates issues for capacity setting. Identifying COVID-19–positive patients will logically require them to be directed down dedicated ‘COVID’ vs. ‘non-COVID’ care pathways (or even, directed to ‘COVID’ vs. ‘non-COVID’ hospitals). This is made explicit in the specialty’s own template for re-activation of elective services (https://icmanaesthesiacovid-19.org/clinical-guidance). This duplication of resources clearly means an increase in total capacity, with the COVID pathway capacity being relatively underutilised. Moreover, a proportion of staff will turn out to be (asymptomatic) COVID positive and this in turn will reduce capacity.

4 How do we manage demand? Hitherto, demand management in the NHS has involved setting clinical care pathways and thresholds for hospital referral [21], and perhaps covertly some rationing of services [22]. COVID-19 has overtly caused the NHS to limit access to services, so the opportunity is there for a bolder and more transparent debate about a sharper definition of the services that the NHS should provide in the longer term. This might indeed result in some treatments no longer be offered [23]. This would mitigate some of the increased costs of providing the increased NHS capacity required. A corollary is the need to create alternative models of funding or care for those services no longer to be provided by the NHS. These will be political/societal decisions as to which services are provided or not, but anaesthetists are in a key position to provide insights to inform those judgements.

The devastation COVID-19 is bringing to care homes, and its predilection for the elderly is an unfortunate form of forced demand management, outside the control of any system or individual. It brings into focus issues around the expectations of longevity in care of the elderly (see also Supporting Information, Figure S7), the role of advanced directives and, even, the role of assisted dying [24]. The last has been explicitly recognised as a potentially ethically justifiable...
standpoint [25], with the main arguments being that permitting assisted dying enables consenting patients to avoid negative quality-adjusted life years (QALYs), while enabling the resources consumed by these patients to provide additional QALYs for others [26]. A pandemic brings these issues into sharp focus. The Supreme Court of Quebec extended the provisions of Canadian law, enabling those with deaths that were ‘reasonably foreseeable’ to opt for assisted dying [27], specifically to avoid contracting COVID-19 and dying a death they had not chosen [28]. Any review in the UK would need to accommodate the need to provide appropriate care for all, prioritising care to maximise outcomes, and also allowing people who do not wish to have inappropriate escalations of care the right to exercise those choices.

Conclusion

Whether we like it or not, COVID-19 has forced us to increase real capacity in the NHS and necessitated difficult decisions of demand management; things we have done poorly in the past. Since COVID-19 will never go away, these issues will remain. The only choice we now have is whether to let COVID dictate a series of emergency plans as it waxes and wanes, or whether we proactively develop a strategy that offers more stable service delivery. The four questions above are not the only relevant ones. One can imagine numerous very specific issues, but I believe almost all questions pertaining to future NHS planning can be mapped onto at least one of these four. Framing the core issues in terms of demand–capacity presents an opportunity for anaesthetists, through their professional societies, to help influence the national agenda – and will help us learn to live with COVID-19 disease.

Acknowledgements

JP served for 3 years as National Clinical Associate to the NHS New Care Models Program, and registered with Oxford University Consulting, works with Deloitte plc as an Associate to advise NHS Trusts on theatre efficiency. He is elected member of Council, Royal College of Anaesthetists and Chair of the Safe Anaesthesia Liaison Group. The views expressed are personal (although consistent with information at https://icmanaesthesiacovid-19.org/clinical-guidance). No other competing interests declared.

References

1. Gallivan S, Utley M, Treasure T, Valencia O. Booked inpatient admissions and hospital capacity: mathematical modelling study. British Medical Journal 2002; 324: 280–2.

2. Cayiri T, Veral E. Outpatient scheduling in health care: a review of literature. Production and Operations Management 2003; 12: 519–49.

3. Pandit JJ. Practical Operating Theatre Management. Cambridge: Cambridge University Press, 2019.

4. Bailey NTJ. Queueing for medical care. Applied Statistics 1954; 3: 137–45.

5. Tlapa D, Zepeida-Lugo CA, Tortorella GL, et al. Effects of lean healthcare on patient flow: a systematic review. Value in Health 2020; 23: 260–73.

6. The King’s Fund. The NHS in a Nutshell. 2020. https://www.kingsfund.org.uk/projects/nhs-in-a-nutshell (accessed 17/05/2020).

7. Snowdon AW, Wright A. Case study: supply chain transformation in the UK National Health Service. Healthcare Quality 2018; 21: 24–7.

8. Polito T, Watson K. Just-in-time under fire: the five major constraints upon JIT practices. Journal of the American Academy of Business 2006; 9: 8–13.

9. Martin RM, Sterne JA, Gunnell D, Davey Smith G, Frankel S. NHS waiting lists and evidence of national or local failure: analysis of health service data. British Medical Journal 2003; 326: 188.

10. Simpson C, Marnoch G. Bed-blocking in the National Health Service in Scotland: a study of bed-blocking in Scottish National Health Service trusts; its nature and extent. Health Bulletin 1999; 57: 99–107.

11. Shoukat A, Wells CR, Langley JM, Singer BH, Galvani AP. Projecting demand for critical care beds during COVID-19 outbreaks in Canada. Canadian Medical Association Journal 2020; 192: E489–E496.

12. Giannakeas V, Bhatia D, Warkentin MT, Bogoch II, Stall NM. Estimating the maximum capacity of COVID-19 cases manageable per day given a health care system’s constrained resources. Annals of Internal Medicine 2020; 173: 407–10.

13. National Institute of Health and Care Excellence. COVID-19 rapid guideline: critical care in adults, NG159. 2020. https://www.nice.org.uk/guidance/ng159 (accessed 12/05/2020).

14. West D. NHS hospitals have four times more empty beds than normal. Health Service Journal 2020. https://www.hsj.co.uk/acute-care/nhs-hospitals-have-four-times-more-empty-beds-than-normal/7027392.article (accessed 05/05/2020).

15. Goldstein ND, Ingraham BC, Eppes SC, Drees M, Paul DA. Assessing occupancy and its relation to healthcare-associated infections. Infection Control and Hospital Epidemiology 2017; 38: 112–4.

16. Eilersen A, Jensen MH, Sneppen K. Chaos in disease outbreaks among prey. Scientific Reports 2020; 10: 3907.

17. Pandit JJ. Managing the R0 of Covid-19: mathematics fights back. Anaesthesia 2020. https://doi.org/10.1111/anae.15151.

18. Mears A. Gaming and targets in the English NHS. Universal Journal of Management 2014; 2: 293–301.

19. Abbott T, White SM, Pandit JJ. Factors affecting the profitability of surgical procedures under ‘Payment by Results’. Anaesthesia 2011; 66: 283–92.

20. Pandit JJ. Pensions, tax and the anaesthetist: significant implications for workforce planning. Anaesthesia 2016; 71: 883–91.

21. NHS England. Demand management good practice guide. 2016. https://www.england.nhs.uk/publication/demand-management-good-practice-guide/ (accessed 05/05/2020).

22. The King’s Fund. Rationing under the radar. 2016. https://www.kingsfund.org.uk/blog/2016/08/nhs-rationing-under-radar (accessed 05/05/2020).

23. New B, Klein R. The rationing debate: defining a package of healthcare services the NHS is responsible for. British Medical Journal 1997; 314: 503–9.
24. Sinmyee S, Pandit VJ, Pascual JM, et al. Legal and ethical implications of defining an optimum means of achieving unconsciousness in assisted dying. *Anaesthesia* 2019; 74: 630–7.

25. Berger JT. Pandemic preparedness planning: will provisions for involuntary termination of life support invite active euthanasia? *Journal of Clinical Ethics* 2010; 21: 308–11.

26. Shaw D, Morton A. Counting the cost of denying assisted dying. *Clinical Ethics* 2020; 15: 65–70.

27. Truchon v. Attorney General of Canada, Superior Court of Quebec, No. 500–17-099119-177, 9/11/19.

28. Banerjee S. Jean Truchon, who fought to expand medically assisted death, receives procedure. National Post (Canada), 2020. https://nationalpost.com/pmn/news-pmn/canada-news-pmn/jean-truchon-who-fought-to-expand-medically-assisted-death-receives-procedure (accessed 19/05/2020).

Appendix: Logistic map equation

\[ x_{n+1} = rx_n(1 - x_n) \]

where \( x_n \) is a number between zero and one that represents the ratio of the existing population (of virus, or people infected) to a theoretical maximum possible population. The parameter \( r \) is the driving parameter that determines the growth rate. The equation is recursive: the result for \( x_{n+1} \) is fed back into the term \( x_n \) for the next iteration. Each iteration is then a ‘generation’ of population growth. The relationship of the term \( r \) with infectivity and viral growth is discussed in an accompanying article [17].

Supporting Information

Additional supporting information may be found online via the journal website.

**Figure S1.** Annual number of inpatient beds, NHS England.

**Figure S2.** Shortfalls in NHS staffing vs. demand.

**Figure S3.** Data on percentage of patients seen within the 4 h target in Accident & Emergency, England, for three successive years.

**Figure S4.** Percentage of patients meeting the 18-week referral to treatment target.

**Figure S5.** Delays in transfer of care from hospital, expressed as total delayed bed days/total occupied bed days.

**Figure S6.** Bed occupancy (%) over successive years (by quarter).

**Figure S7.** Total NHS spending by age (average per person per year).