



THE DOCTOR *MIGHT* SEE YOU NOW

Healthcare rationing in the NHS before and after the Covid-19 pandemic

By Dr Robert Sutton

BRIEFING PAPER

EXECUTIVE SUMMARY

- The National Health Service Constitution promises universal access to care, free of charge, based solely on clinical need. This premise is extremely popular with the British public. No major British political party has indicated any intention of changing this promise.
- This popular goal is a fiction. It is impossible for any healthcare system to provide unlimited care. Rationing, of some description, is inevitable because of limited resources. The NHS hides the extent of rationing by using indirect means, often avoiding the associated difficult ethical questions.
- Some systems ration healthcare through prices. The NHS rations healthcare through opaque criteria such as waiting times, clinical criteria and administrative hurdles. The end result for patients is near identical to a system that rations through prices.
- Despite claims that treatments are made based on best-value, there are often arbitrary, political decisions made by politicians and the NHS regarding which diseases and treatments to prioritise. These decisions are made with limited disclosure of reasoning or public involvement. This rewards the loudest and best-connected lobbyists.
- The rationing of healthcare has grown steadily with an ageing population, increasing treatment costs and limited financial resources. The Covid-19 pandemic has magnified rationing pressures.
- Indirect rationing of healthcare resources by the NHS will likely continue to a greater degree after this pandemic, with cancelled and deferred treatments and diagnostic procedures necessitating more severe rationing. Waiting lists, already unwieldy, have grown dramatically. As of February 2020, there were 4.7 million people in England waiting on routine operations, a record number. It is unclear whether the NHS will be able to fulfil its constitutional obligations.
- The promises of increased spending made ahead of the 2019 UK general election will have only a marginal impact on patient access and outcomes. Healthcare has unlimited potential demand, meaning no matter how much money is spent there will always be rationing.
- There is evidence that the NHS is not using its resources efficiently. The UK has fewer practising physicians per 100,000 population than 26 of the 27 EU member states, despite above average spending levels. The UK has more medical graduates per capita than Germany, yet has fewer practicing physicians because of the NHS' difficulty in retaining staff.

- If the Government wants to ensure high quality healthcare outcomes for the United Kingdom they should:
 - (1) increase transparency by disclosing the extent and nature of rationing in the NHS, including the patients who are refused care;
 - (2) undertake structural reforms, learning from global best practice, with the aim of increasing competition, productivity and efficiency, thereby reducing the need for rationing;
 - (3) revisit the long-term plan for the NHS to ensure it is able to fulfil its obligations to the British public.

ABOUT THE AUTHOR

Robert Sutton is a junior doctor in Wales and recent graduate of the University of Oxford Medical School. He previously worked as a parliamentary staffer.

The National Health Service (NHS) offers healthcare to residents of the United Kingdom (UK). The breadth of services provided, and the equitable access to these services, has been heralded by British politicians across successive parliaments as a source of national pride since its formation in the aftermath of the Second World War.

The provision of free medical services, paid for through general taxation, was a key recommendation of the 1942 Beveridge Report, which proposed a programme of radical social reform offering the public a safety net “from the cradle to the grave.”¹ These proposals achieved enough popularity during the deprivation of wartime Britain that the post-war Labour government passed the 1946 National Health Service Act, implemented in 1948. The true level of popularity at the time has been questioned.² Nevertheless, the NHS continues to be judged to this day against the idealised post-war narrative.

There has been little political desire for any deviation from Beveridge’s original vision since the formation of the NHS. Today, the NHS is bound by a set of principles, laid out in the NHS Constitution, which oblige it to provide “a comprehensive service, available to all,” access based on “clinical need, not an individual’s ability to pay” and to aspire to “the highest standards of excellence and professionalism.”³

These ideals clash with the challenges of an ageing population, growing costs for new treatments and inevitable financial constraints. The sustainability of the current model and whether it is truly able to fulfil its constitutional obligations is therefore unclear.

A healthcare system which is “free to all at the point of care” must necessarily be limited in some way because of the substantial costs of providing the service. There are doctors, nurses and support staff that must be remunerated, buildings constructed and maintained, costly equipment and treatments made available. The NHS rations healthcare through various means in order to provide “free” services. There is a necessary trade-off between the breadth of services available and the accessibility and quality of those services if they are to remain “free”.

This paper examines the causes and consequences of the financial constraints the NHS operates within, how these constraints lead directly to *de facto* rationing in healthcare (which is never made explicit), and the implications for future service delivery. To provide an ethical justification for healthcare rationing within the NHS, it is important that its users are aware of the restrictions which it necessarily imposes. Political consensus is meaningless if these points are not broached.

1 Sir William Beveridge, “Social Insurance and Allied Services,” November 1942 (London, UK: HM Stationery Office)

2 Kristian Niemietz, “Diagnosis: Overrated – An analysis of the structural flaws in the NHS,” 4 December 2015, (London, UK: Institute of Economic Affairs), pp. 10-17

3 “The NHS Constitution for England,” last updated 1 January 2021, (London, UK: Department of Health & Social Care)

The NHS has a reputation as a “political football,” too important to be neglected yet too politically volatile to reform significantly.⁴ But ignoring important policy discussions does not make the problem go away. Instead, much as a patient may neglect having a check-up for fear of what their doctor might find, policymakers have avoided looking too closely at the NHS for fear of what their discoveries might suggest. The strain placed by the current Covid-19 pandemic has illustrated the fragility of the NHS when exposed to surges in demand, and the necessity of undertaking a closer inspection.

And as with a condition treated too late in its natural course, by neglecting serious discussions about the future of the NHS we risk delaying tackling the problem until it is too late to remedy. It is the hope that this paper will encourage sincere, if difficult, discussions about patient access to healthcare in the UK and what might be done to ensure the NHS can live up to its constitutional ideals.

THE ORIGINS OF HEALTHCARE RATIONING IN THE NHS

A society has limited resources which must be assigned to meet an infinite number of possible demands. The UK government directs some of those resources through taxation to fund the NHS, which is then free at the point of use. If a service has a broad scope, large user base and no cost for its users (as with the NHS) then there will be effectively unlimited demand.

There is no clear limit to how many services could fall under the umbrella of healthcare desirable to NHS users. An effectively infinite number of services could be introduced and further demand would still exist. This is particularly true given the pace with which new medical technologies and treatments are being introduced. Demand for healthcare services is effectively unlimited. The resources to meet this demand are not.

Decisions must ultimately be made for which services to commission. The basis for any such allocation will inevitably be, in some sense, arbitrary. The number of good causes is endless, yet a decision has to be made according to some criteria. The complexity of the NHS bureaucracy exacerbates the difficulty. Sweeping decisions to commission or deny certain services will inevitably lead to some form of injustice, intermittently wasteful or lacking.

It does not then follow that merely rerouting more money towards the NHS will allow us to better meet demand and reduce the need for rationing. There is likely to be some truth to this, at the margin, but how much requires further investigation. More money is not a necessary condition, and certainly not a sufficient one, to overcome the reality of rationing as a means of approaching unlimited demand with a limited budget. This is especially true within an inflexible and inefficient bureaucracy.

⁴ Abi Rimmer, Gareth Iacobucci, “NHS becomes political football as electioneering kicks off,” *BMJ* 2019;367:l6375

A healthcare system requires resources which cover an enormous range in terms of cost, liquidity and scarcity. More medicines, equipment, protective gear and medical professionals can be “bought” within domestic or global markets, and we can increase access by spending more and outbidding rival healthcare systems. Yet there is obviously an upper limit to this, as we compete for scarce resources using limited capital.

In 1950, the NHS budget as a proportion of GDP stood at 3.5%.⁵ After reaching a peak of 7.6% in 2019, it fell to around 7.2% in 2017. The reversal of this decline was a key Conservative Party pledge ahead of the 2019 UK general election.⁶ The decline has been cited by those arguing that “underfunding” is the root cause of inadequacies in the ability of the NHS to meet public demand.⁷ International comparisons are also frequently used, with the relatively higher public spending in Canada, Japan, Germany, France and the United States compared to the UK being heralded as a goal for future NHS budgeting.⁸

Yet in real terms, public health spending has grown by an average of 3.6% per year over the past 60 years.⁹ The UK also spends a higher percentage of our GDP on publicly funded healthcare than 28 of the 37 OECD nations, and above the average for the group (Figure 1).¹⁰ By this, the NHS is not markedly underfunded relative to other nations’ health services. Nor does it seem that a modest and relatively recent decline in relative spending can take the full blame for the NHS not being able to match public demand. Were the UK to inflate spending to match our international peers at the top of the scale, it would not eradicate the need for rationing because, as discussed, there would continue to be infinite potential demands on the system from the public.

5 John Appleby, “70 years of NHS spending,” 21 March 2018, Nuffield Trust

6 Jonathon Holmes, “What have the parties pledged on health and care?” 28 November 2019, The King’s Trust

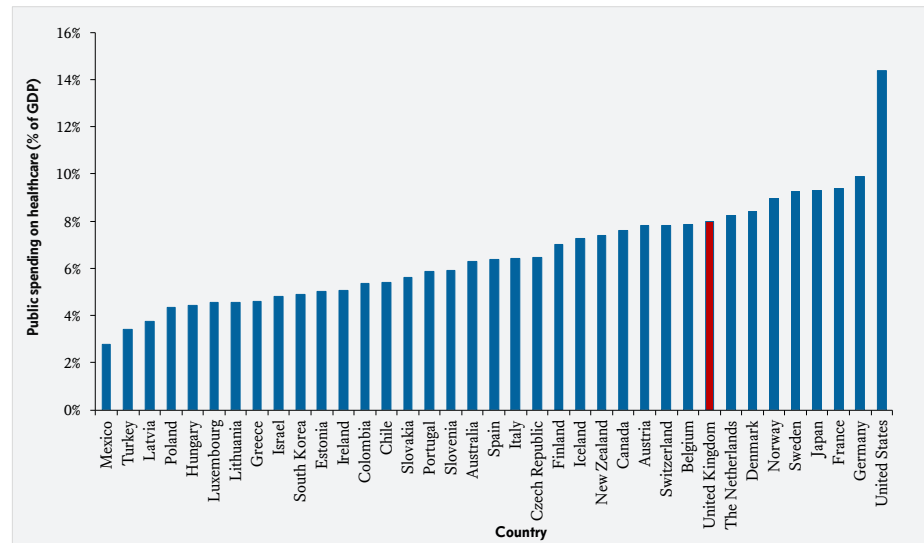
7 Polly Toynbee, “These brutal cuts to the NHS will haunt the Conservatives,” 25 October 2019, The Guardian

8 “How does UK healthcare spending compare with other countries?” 29 August 2019, Office for National Statistics

9 “Health spending as a share of GDP remains at lowest level in a decade,” 30 July 2019, The Health Foundation

10 “Health spending,” last accessed 17 January 2021, OECD

**FIGURE 1. PUBLIC SPENDING ON HEALTHCARE (% OF GDP)
ACROSS THE OECD IN 2019**



The UK spends above average, and has a higher level of spending than 28 of the 37 nations.

Furthermore, increasing the fraction of the public purse spent on health services requires either a declining proportion be spent in other important areas – social care, education, defence, foreign aid – or a tax hike, thereby reducing other economic activity and lowering wages. To justify either would require it be demonstrated that the principal cause of lacklustre care and rationing in the NHS is insufficient public spending, that the reduction in rationing achieved through increasing funding would justify withdrawing that money from other public services or the taxpayer and that such a pattern would be sustainable against a long-term trend of an ageing population, growing comorbid disease burden and rising treatment costs. In order to reduce rationing through increased public spending these conditions would have to be met.

Matching demand for healthcare services is also complicated by dramatic changes over much shorter timescales. This is demonstrated during yearly surges in demand during the colder months (so-called “winter crises”) and by the current Covid-19 pandemic.¹¹ Relatively illiquid resources are used to provide healthcare to a population, including medical professionals, treatments and facilities.

It is generally difficult, costly and time-consuming to obtain more of these. The lengthy process of training medical professionals, developing new treatments and building hospitals puts a limit on how quickly we can expand available resources in response to demand. This is apparent when a surge in demand outpaces our ability to grow a relatively fixed pool of resources. Higher demand necessitates stricter rationing.

¹¹ “NHS winter pressures,” The Health Foundation. A collection of posts from The Health Foundation on winter pressures in the NHS.

Staffing levels allow for straightforward international comparisons. Between 2009 and 2020, the number of doctors increased from 95,410 to 121,142 (27%), nurses from 278,470 to 302,471 (9%) and scientists, therapeutics and technical staff from 118,917 to 147,554 (24%).¹² In contrast, the number of permanent general practitioners (GPs) fell from 28,631 in 2015 to 26,435 in 2020, a fall of 7.7% and well short of the target of 5,000 new GPs.¹³ Mental health professionals declined between 2009 and 2017 from 99,457 to a low of 88,698 but have since increased to 98,649.¹⁴

These staffing levels are some of the lowest in the OECD.¹⁵ The UK has fewer practising physicians per 100,000 population than 26 of the 27 EU member states, despite above average spending levels. We also maintain chronic shortages of staff in certain specialties, in particular paediatrics, psychiatry and general practice, although this improved in 2020.¹⁶ The 2019 general election party manifestos displayed the growing sense of political urgency to remedy these shortages. One of the Conservative Party's manifesto priorities was to increase the number of nursing staff by 50,000 by 2024, and party leader Boris Johnson had stated that the NHS, rather than Brexit, was his true priority.¹⁷

Steps have been taken to increase the throughput of students for professionally-qualified healthcare positions. But increasing student numbers might not be sufficient to increase staffing levels. The UK already produces more medical graduates per capita than Germany, yet has 284.1 practising physicians per 100,000 population compared to Germany's 431.1.¹⁸ Retention of medical graduates through the full course of postgraduate medical training is worsening. The proportion progressing directly from foundation to specialty training has fallen from 83.1% in 2010 to 37.7% in 2018, with 14.8% no longer practising medicine.¹⁹ Reasons cited for leaving medicine include work-life balance, pay, inflexible schedule, fragmented teamwork, and lack of training opportunities.²⁰ This means substantial taxpayer resources are being spent training doctors who do not ultimately go on to practice medicine. Putting more graduates into an unappealing system does not necessarily translate into better outcomes.

HOSPITAL BEDS

The number of hospital beds in England has more than halved over the past 30 years, from 299,000 in 1987/88 to 141,000 in 2019/20.²¹ This currently equates to

¹² "NHS staffing tracker: Hospital services," 30 June 2020, Nuffield Trust

¹³ "NHS staffing tracker: General practice," 30 June 2020, Nuffield Trust

¹⁴ "NHS staffing tracker: Mental health and learning disability," 30 June 2020, Nuffield Trust

¹⁵ "Healthcare personnel statistics – physicians," August 2020, Eurostat

¹⁶ "Specialty recruitment: round 1 – acceptance and fill rate," 14 July 2020, Health Education England

¹⁷ "Election results 2019: Boris Johnson's victory speech in full," 13 December 2019, BBC News

¹⁸ "Health at a Glance 2019," 7 November 2019, OECD iLibrary

¹⁹ Tom Moberly, "More doctors are taking a break from training after foundation programme," *BMJ* 2019;364:l842

²⁰ Hannah Wilson, Arabella Simpkin, "Why are so many doctors quitting the NHS?" 6 February 2020, *The BMJ Opinion*

²¹ Leo Ewbank, James Thompson, Helen McKenna, Siva Anandaciva, "NHS hospital bed numbers: past,

2.5 beds per 1,000 citizens. The same number is 2.8 for the USA, 3 for Spain, 3.2 for Italy, 6 for France, 7.4 for Austria and 8 for Germany. Pressure on bed space also appears to have risen, with rising occupancy rates and declining lengths of stay. From 2011 to immediately preceding the pandemic, occupancy rates increased from 86.6% to 89.4%.²² The average length of stay has been squeezed from 5.3 days to 4.5 days.²³

This is not necessarily an indicator of declining access to healthcare. Hospitals must aim for a “Goldilocks” level of bed occupancy; not too high that they cannot accept new patients and not too low that bed space is being wasted. To determine whether these trends reflect growing demand and resulting rationing, or merely more efficient use of bed space, more information is needed.

Readmission data is one useful source. Between 2013/14 and 2018/19, the emergency readmissions rate (the proportion of patients being admitted who had been discharged in the previous 30 days) grew steadily from 12.5% to 14.3%.²⁴ Readmissions is an indicator that an increasing proportion of patients are being discharged with unmet needs. This is more consistent with an NHS in which pressure on bed space necessitates rationing of healthcare for inpatients.

THE CHALLENGES OF COMMISSIONING: AGEING CITIZENS, COSTLY TREATMENTS

THE COMMISSIONING PROCESS

How does the NHS decide which health services to commission? There is an effectively unlimited number of good causes to choose. Funding must be allocated to staffing, infrastructure, equipment, technologies and medicines, and within each such category further decisions must be made on how best to spend that allocation.

In the case of new medicines, the path taken from research and development to eventually being approved for use on the NHS is expensive and time-consuming.²⁵ Following initial approval by the Medicines and Healthcare products Regulatory Agency (MHRA), the National Institute for Health and Care Excellence (NICE) assesses whether the treatment represents good value for money and makes the decision whether to recommend it to the NHS. An incremental cost-effectiveness ratio (ICER) is calculated to determine the cost difference for the increased effectiveness compared to the previous treatment standard.

A low ICER is favourable, representing a low incremental cost for the gain in effectiveness. When calculating “effectiveness,” NICE relies on the quality-adjusted

present, future,” 26 March 2020, The King’s Fund

22 “Hospital bed occupancy,” last updated 15 October 2020, Nuffield Trust

23 Leo Ewbank, James Thompson, Helen McKenna, Siva Anandaciva, “NHS hospital bed numbers: past, present, future,” 26 March 2020, The King’s Fund

24 “Emergency readmissions,” last updated 17 September 2020, Nuffield Trust

25 Ben Collins, “Access to new medicines in the English NHS,” 28 October 2020, The King’s Fund

life year (QALY), a controversial means of evaluating years of life gained and the “quality” of those years in terms of symptom burden and general health.²⁶ An ICER of below £20,000 per QALY is considered good value for money and will generally be recommended to the NHS; £20,000 to £30,000 will require a more careful evaluation of the strength of evidence; and above £30,000 will generally be rejected. These figures are somewhat arbitrary, but have been reached over time as a balance between financial restraints and healthcare demands. There is direct price-based rationing occurring at this stage of the commissioning process with minimal patient involvement.

EXCEPTIONS TO THE RULE

NICE’s ICER criteria provide an intuitive utilitarian justification for commissioning of new treatments. It also allows for previously approved treatments which have been reevaluated and deemed to be poor value to be removed and placed on the NHS “blacklist.”²⁷ Yet there have been many exceptions which have bypassed NICE’s cost-effectiveness criteria, typically involving emotive conditions which exert political pressure on NHS leadership.

The recent approval of the cystic fibrosis treatment lumacaftor/ivacaftor (trade name Orkambi) is one high-profile example. Despite an initial ICER of almost £220,000, persistent campaigning and public pressure contributed to the government yielding and renegotiating for its approval.²⁸ The Cancer Drugs Fund is another example, providing funding for expensive new cancer treatments which were previously rejected by NICE, and QALY price limits have increased for end-of-life treatments.²⁹ This includes so-called “orphan” drugs for which no alternative is available.³⁰ There are numerous such examples of treatments which would normally be rejected by NICE being approved for use in the NHS.

WHAT NICE GIVES, THE NHS TAKES AWAY

The NHS has a statutory obligation to implement NICE recommendations within 3 months.³¹ Yet this has not prevented the NHS previously pushing back, through indirect means, and thereby restricting access. The growing number of exceptions to NICE’s criteria is a further challenge to keeping NHS spending within its budget. The high cost of sofosbuvir, a treatment for hepatitis C, at £80,000 per QALY, led the NHS to apply quotas and clinical criteria to restrict access and

²⁶ Oya Asim, Stavros Petrou, “Valuing a QALY: Review of current controversies,” *Expert Review of Pharmacoeconomics & Outcomes Research* 2005;5:667-9

²⁷ “NHS England Board Paper: Items which should not be routinely prescribed in primary care: findings of consultation and next steps – for decision,” 30 November 2017, NHS England

²⁸ “Appraisal consultation document: Lumacaftor-ivacaftor for treating cystic fibrosis homozygous for the F508del mutation,” March 2016, National Institute for Health and Care Excellence; “NHS England concludes wide-ranging deal for cystic fibrosis drugs,” 25 October 2019, NHS England

²⁹ “Cancer Drugs Fund,” accessed 17 January 2021, NHS England; Kalipso Chalkidou “Evidence and values: paying for end-of-life drugs in the British NHS,” *Health Economics, Policy and Law* 2012;7:393-409

³⁰ Tom Powell, Michael O’Donnell, “NICE appraisals of rare diseases,” 12 March 2019, House of Commons Library

³¹ “Guide to the processes of technology appraisal,” last updated 30 May 2018, National Institute for Health and Care Excellence. See section 1.5.

control spending.³² Even when NICE opens the door to patient access, the reality of NHS budgetary constraints demands rationing.

In a complex world of changing resources and demands, effective allocation to perfectly match the two is extraordinarily difficult if not practically impossible. Resources vary over time: staff enter and leave the workforce; new treatments are introduced and older ones are proven ineffective; new hospitals are built and old ones closed down. Demand is similarly dynamic: there is a yearly cycle of high winter demand and unpredictable events such as the Covid-19 pandemic require drastic action to avoid services being overwhelmed. Planners need to find ways to allocate resources to meet citizens' expectations.

This requires rationing. Despite the NHS Constitution stating that medical treatment should be available depending solely on clinical need, not ability to pay, the reality of an inflexible bureaucracy operating within financial constraints has necessitated healthcare rationing.

MECHANISMS OF HEALTHCARE RATIONING IN THE NHS

Rationing is a necessary process. It can be exacerbated dramatically when demand is volatile and resources relatively fixed. Access to services can be severely restricted in such circumstances.

RATIONING IS A REALITY OF THE NHS

The NHS Constitution states that service must be available to all, depending solely on clinical need. The public expects the NHS to meet their treatment requirements. Politicians are aware of the considerable political damage which would be caused by explicit denial of service. Yet the perfect matching of resources and demand is incompatible with the reality of the UK's current market for healthcare services.

The result of this political denialism is that rationing occurs through opaque, indirect measures. NICE does this on price grounds long before treatment choices are made available to patients and clinicians. Other measures, while sidestepping the unconstitutionality and political embarrassment of an outright denial, nonetheless serve to restrict access in a manner which, from a consequentialist perspective, has the same endpoint of restricting access to healthcare.

Consider an example of a 74-year-old lady with osteoarthritis of the knee. The pain and loss of functionality make it extremely difficult for her to perform many of her activities of daily living: getting dressed, washing, preparing meals and shopping all become extraordinarily difficult. Her GP refers her to an orthopaedics consultant but also informs her that the current waiting time is several months, and that even if she is deemed eligible, the waiting list for knee replacement surgery is just

³² Piotr Ozieranski, Lawrence King, "Wales approves new hepatitis C drug while England deliberates," *BMJ* 2016;355:i6499

as long. She is informed that there is no way to get this operation done quicker. She dies shortly after her 75th birthday after acquiring pneumonia in the community, without having had her knee replacement operation.

In this scenario, there is no explicit denial of service. The patient *believes* she will have the treatment she expects. But the endpoint is the same as if her GP had simply refused to refer her for further evaluation and treatment: she did not receive the treatment which she required. This may seem like bad luck, but it is one of a range of mechanisms employed systematically throughout the NHS to ration healthcare. In this case, treatment is delayed for an older and sicker patient who has a higher chance of dying while awaiting treatment. Indirect discrimination occurs against such patients.

RATIONING BY ANY OTHER NAME

How else is healthcare rationed in the NHS, and is rationing truly a necessity? Could we not simply find ways to cut unnecessary costs and improve productivity, thereby avoiding the need for healthcare rationing? Theoretically, yes. But there will be a natural limit to this, and the solution to how one can restructure incentives to absorb increased demand within a publicly funded organisation such as the NHS has evaded successive governments.

It is the path of least resistance for service providers to implement methods which indirectly ration healthcare rather than striving for reform. Increasing productivity requires structural and cultural changes within an organisation, which are difficult to achieve, particularly if employees may not feel a personal stake in the organisation's success.

In rationing healthcare, the quality of service can be lowered, or the number of patients using it reduced. Lowering service quality might include smaller teams to perform procedures, reduced clinical follow up or requirements that patients pay for over-the-counter medications and supplies (a rare example of politically acceptable price-based rationing applied directly to service users).³³ These methods are all used to some degree for different services, and they each “lower” quality to variable degrees. Although some could arguably be implemented without lowering standards, there will be a natural limit to how far we can whittle down a service without eventually impacting patient care.

Restricting patient access can be done in a number of ways. As discussed above, lengthy queues can restrict access in a manner which differentially targets older and sicker patients. One can seek to deter patients from help-seeking behaviours (by making treatment inconvenient or otherwise unappealing), reschedule and thereby delay treatments (with the same effect as queueing), apply administrative burdens for clinicians (such as extraordinary funding requests) and implement clinical criteria to narrow eligibility.³⁴

³³ “Guidance on conditions for which over the counter items should not routinely be prescribed in primary care,” last accessed 17 January 2021, NHS England

³⁴ Gareth Iacobucci, “Exceptional requests for care surge as rationing deepens,” *BMJ* 2017;358:j3188

In the case of the Covid-19 pandemic, the opportunities to ration patient access have expanded with restrictions on face-to-face consultations. Denying patients in-person consultations and implementing lengthy telephone waiting times to book appointments will result in some patients turning away with their needs unmet.³⁵ Requiring patients to have pre-booked appointments to attend emergency departments could have a similar effect.³⁶ Each hurdle serves to filter out more potential patients.

At a national commissioning level, NICE uses quality-adjusted life years, a measure which has long been recognised as discriminating against treatments for older and sicker patients, as a way of deciding which treatment options to fund.³⁷ Other mechanisms, more subtle and difficult to quantify, can ration healthcare without this explicit refusal of service.

QUANTIFYING HEALTHCARE RATIONING

If rationing must occur, how might it be quantified? As a consequence of limited resources, excess demand, inadequate planning and political hesitancy, we might expect it to be concealed to a certain degree to limit political fallout. There is the possibility that attempts to “game the system” by service providers result in published statistics providing an inadequate representation of the reality of patient access. This was the case when ambulances were reported delaying admitting patients into emergency departments in order to better meet the 4-hour wait target.³⁸

The NHS Targets are performance indicators specified in the Constitution which outline a standard the public can expect.³⁹ Whether targets are effective or result in improving outcomes is controversial.⁴⁰ NHS Digital publishes data regularly on a range of indicators, and health policy institutions such as the Nuffield Trust, the Health Foundation and the King’s Trust provide user-friendly interfaces to track these.⁴¹ They allow us to indirectly observe the mechanisms and effects of rationing.

Queueing, or waiting times, is one of the more straightforward mechanisms to quantify. The NHS publishes figures for different clinical waiting times. The key outcome measured for timeliness of acute treatment is the 4-hour wait for emer-

35 Nick Bostock, “Millions of patients ‘avoiding calls to GP’ during COVID-19 pandemic,” 25 April 2020, GPonline

36 Owain Clarke, “Coronavirus: People told to phone ahead before going to A&E,” 13 July 2020, BBC News

37 David Schwappach, “Resource allocation, social values and the QALY: a review of the debate and empirical evidence,” *Health Expectations* 2002;5:210-22

38 Robert Watts, Laura Donnelly, “Don’t leave patients in ambulances to hit A&E targets, hospitals told,” 27 October 2012, *The Telegraph*

39 Elizabeth Parkin, “NHS maximum waiting time standards,” 27 March 2020, House of Commons Library

40 Ian Blunt, “Fact or Fiction? Targets improve quality in the NHS,” 13 February 2015, Nuffield Trust

41 “Data and information,” NHS Digital

gency hospital treatment.⁴² In the subacute category, cancer referrals and treatment are subject to NHS timeliness targets. And in the elective category, various procedures such as hip- and knee-replacement surgery have nationally collected and published waiting times. A consistent trend across diverse statistical measures could strengthen our confidence in any interpretation.

INTERNATIONAL COMPARISONS OF QUEUEING

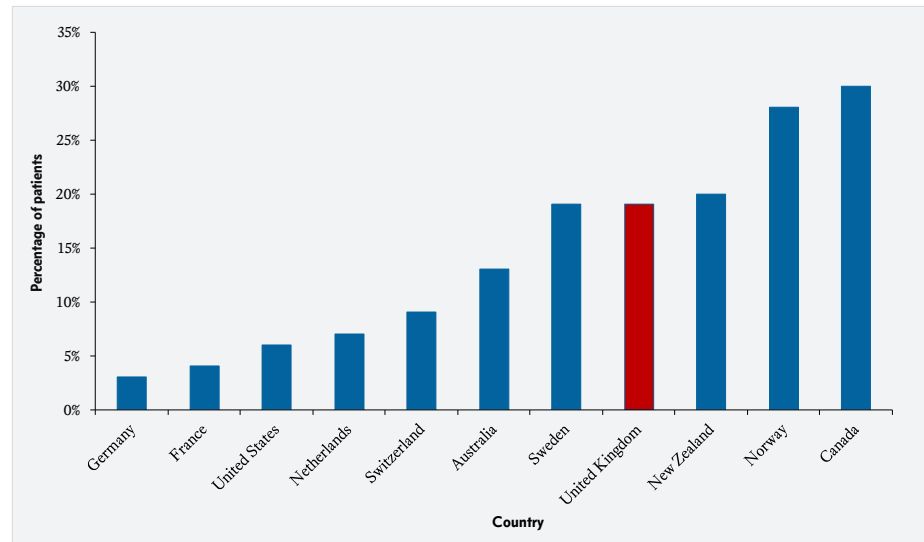
These metrics also allow us, to a limited extent, to make international comparisons of the length of waiting times as a proxy for the severity of rationing. Data is limited here, but the UK seems to perform around average across a range of measures. The 2016 Commonwealth Fund International Health Policy Survey of Adults and the OECD iLibrary compile internationally comparable statistics on access and waiting times.⁴³

The UK performance on ability to get same- or next-day appointments, waiting times for specialist appointments (See Figure 2) and waiting times for elective surgery (See Figure 3) are distinctly average. Many nations which seem to consistently outperform us – the Netherlands, Australia, Denmark, Finland, New Zealand – are also among those using a smaller proportion of their GDP on public healthcare spending. More public spending correlates weakly with these proxy measures for rationing.

42 The 4-hour wait is the most widely discussed and politically contentious of the NHS Targets. For a discussion, see: “What’s going on with A&E waiting times?” last updated 25 March 2020, The King’s Fund

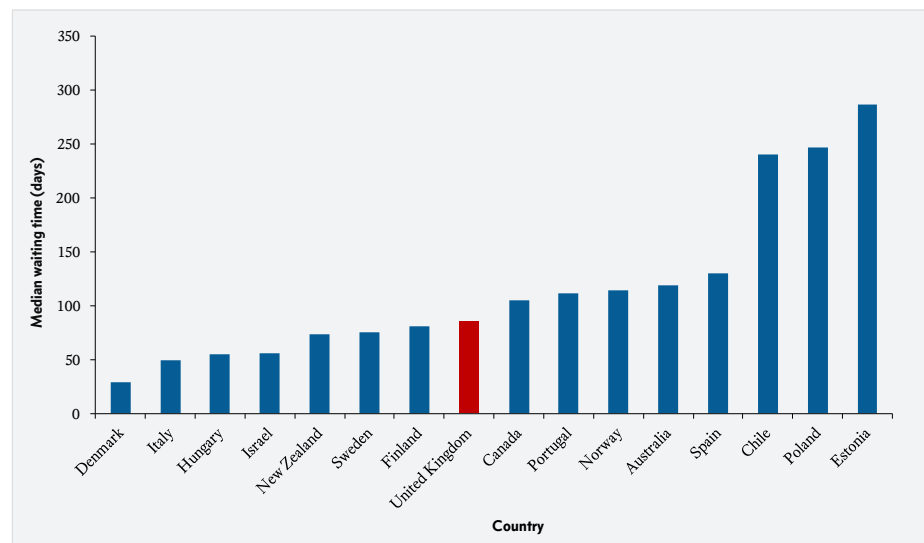
43 “2016 Commonwealth Fund International Health Policy Survey of Adults,” 16 November 2016, The Commonwealth Fund; “Health at a Glance 2019: Waiting times for elective surgery,” 7 November 2019, OECD iLibrary

FIGURE 2. PERCENTAGE OF PATIENTS REQUIRING A SPECIALIST APPOINTMENT WHO WAITED TWO MONTHS OR LONGER (2016)



Percentage of patients who saw or needed to see a specialist in the past 2 years and were required to wait two months or longer for the appointment. Based on data from the 2016 Commonwealth Fund International Health Policy Survey.

FIGURE 3. HIP REPLACEMENT MEDIAN WAITING TIMES (2017)



Based on data from "Health at a Glance 2019," 7 November 2019, OECD iLibrary.

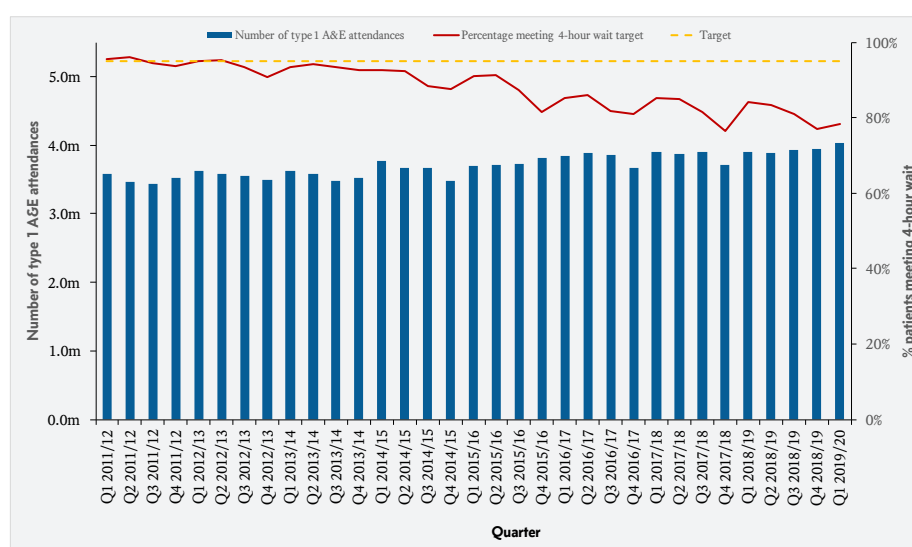
ACCIDENT & EMERGENCY

Domestically, the Nuffield Trust and the Health Foundation maintain a wide range of indicators of healthcare service and access in the UK through their Quality-Watch programme.⁴⁴ To quantify rationing within the UK over time, we consider the trajectory over previous years and in recent months during the Covid-19 pandemic. Together, these paint a worrying picture of a steady decline in healthcare access followed by an acute decompensation in service provision as a result of the pandemic.

⁴⁴ "QualityWatch," Nuffield Trust

The 4-hour wait target is much publicised as a proxy for acute care provision.⁴⁵ The target for the proportion of patients being seen within 4 hours was lowered from 98% to 95% of patients in 2010, an early signal of the difficulty in achieving it and the political concern about failing to do so. In spite of this lowering of the target, it has not been met since July 2015. By the third quarter 2019/20 it had fallen to an all-time low of 71%. Since then it has improved sharply to 90%, but this coincided with a marked decrease in the number of patients presenting to A&E (Figure 4). The number who presented in the first quarter of 2020/21 was down 44% on one year previously. It is unclear whether this reduced demand will adversely impact patient outcomes, or whether many of those previously presenting to A&E were inappropriately attending.

FIGURE 4. A&E ATTENDANCES AND PERCENTAGE MEETING 4-HOUR WAIT TARGET OVER TIME



Plot of total number of type 1 A&E attendances Q1 2011/12 to Q1 2020/21 and the percentage of type 1 attendances being seen within 4 hours. Dashed line represents 95% target. There has been a steady increase in the number of attendances over the last decade and a corresponding fall in the percentage being seen within 4 hours. The recent uptick in percentage being seen within 4 hours is difficult to interpret given the dramatic fall in A&E attendances as a result of the Covid-19 pandemic.

CANCER REFERRALS, INVESTIGATION AND TREATMENT

Cancer waiting times support this picture of an acute decompensation in services for non-Covid-19 healthcare problems.⁴⁶ Delays in cancer treatment are an emotive issue and have been reported as a consequence of the pandemic.⁴⁷ Different parts of the referral pathway are subject to different waiting times. Following an urgent GP referral for suspected cancer, at least 93% of patients should be seen by a specialist within two-weeks. Following a decision to treat a new primary cancer, at least 96% of patients should start treatment within 31 days (although this num-

⁴⁵ "A&E waiting times," last updated 22 July 2020, Nuffield Trust

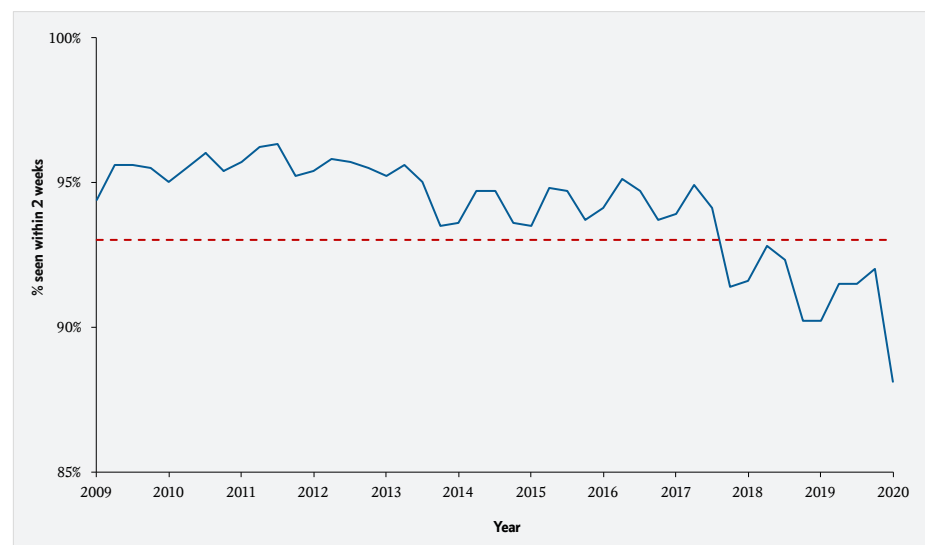
⁴⁶ "Cancer waiting times," 14 December 2020, Nuffield Trust

⁴⁷ Sharon Brennan, "Cancer treatment delayed as patient priority lists drawn up," 23 March 2020, HSJ

ber varies for different types of treatment). 85% of patients should have progressed from an initial GP referral to a first treatment within 62 days.

The two-week referral standard has been missed for over 2 years and was at an all-time low of 88.1% in the second quarter of 2020/21 (See Figure 5). The target for GP referral to first treatment within 62 days has not been met since 2013, and currently stands at 76.9%. The other main pathway besides GP referral is referral via screening, and the impact of the Covid-19 pandemic on this pathway is even more worrying.

FIGURE 5. PERCENTAGE OF PATIENTS REFERRED BY GP FOR SUSPECTED CANCER SEEN BY A SPECIALIST WITHIN 2 WEEKS



Percentage of patients referred by their GP on the suspected cancer pathway and seen by a specialist within the target time of 2 weeks. The target of 93% of patients to be seen within 2 weeks has not been met for the last 10 quarters, and has fallen to an all time low of 88.1% as of the second quarter 2020/21.

The target time from an initial screening service to a first treatment for cancer is the same as for the GP pathway at 62 days. Ninety percent of patients should meet this target. In the second quarter of 2019/20, 86.9% of patients were achieving this: below the target, but only slightly. By the second quarter of 2020/21, the number of patients meeting it had fallen to 64%.

Individual aspects of cancer treatment, such as chemotherapy, radiotherapy and surgery, are subject to different targets. The target of 94% for surgery has been missed for the last 8 quarters and has now fallen to its lowest level at 87.5%. As a result of the pandemic, timely interventions for cancer treatment have been limited, and modified Cancer Waiting Times guidance was issued stating that prioritisation should be implemented with downgrading or avoidance of referrals where capacity is limited.⁴⁸ This is worrying, and suggests that we will almost certainly miss many early cancer diagnoses, with associated worsening treatment outcomes.

Cancer is not the only diagnostic pathway which has suffered. Diagnostic testing across multiple modalities and specialities has been delayed.⁴⁹ There is a target time from specialty consultation to diagnostic test of 6 weeks, as part of the overarching 18-week referral-to-treatment (RTT) target. The 6-week target is measured as a composite of 15 key imaging, physiological and endoscopic procedures.

The current threshold of 99% has rarely been met since its introduction in the NHS Operating Framework 2012/13, but the proportion of patients not meeting this target has managed to stay below 5% since it was introduced. This proportion increased dramatically as elective diagnostic procedures were cancelled to build capacity during the early weeks of the pandemic. At its worst, 58.5% of patients failed to meet the 6-week target, and this number now stands at around 38%. In early 2020, the yearly test rate was around 2 million, but now stands at around 1.5 million.

This would be less concerning if the diagnostic pathway had sufficient capacity to handle the inevitable backlog of tests which we are presently building up. The waiting list for a diagnostic test more than doubled from a low of 411,569 in 2008 to a pre-pandemic high of 1,081,921, indicating that the difficulty in keeping up with demand is a long-standing issue. It has since jumped to 1,241,095, with a corresponding increase in the median wait time from 2.5 weeks to 4.1 weeks.

This demonstrates the impact of Covid-19 on non-Covid diagnostic testing. Resulting delays in diagnosis may lead to more advanced disease at diagnosis and worse prognosis. If non-Covid diagnostic testing capacity remains at present levels, and delayed access to care continues as a result of the pandemic, rationing will necessarily be implemented through a waiting list which has little hope of reaching pre-pandemic levels, much less of being cleared.

PRIMARY CARE SERVICES

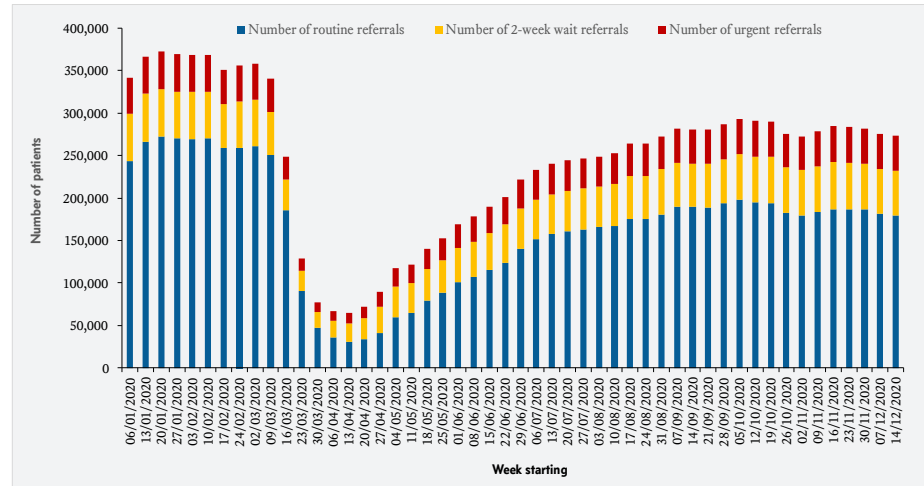
GP referrals fell dramatically during the lockdown across all referral types, not just those for suspected cancer.⁵⁰ Published data is broken up into routine, urgent, and suspected cancer (2-week wait) referrals. The typical number per week at the start of 2020 stood at 270,000 for routine, 43,000 for urgent and 55,000 for suspected cancer. These numbers fell to lows of 25,567, 8,950, and 15,959 during the pandemic, respective declines of 91%, 79%, and 71%.

These numbers have not yet returned to their pre-pandemic baseline. Assuming typical weekly referral numbers given above, based on referral data for 2020 (Figure 6), approximately 435,000 urgent referrals, 424,000 suspected cancer referrals and 5,326,000 routine referrals which would normally have been made did not happen. There has been a dramatic restriction of primary care services, likely due to both avoidance of healthcare services by patients and reluctance of GPs to make referrals.

⁴⁹ "Diagnostic test waiting times," last updated 15 October 2020, Nuffield Trust

⁵⁰ "NHS e-Referral Service (e-RS) open data dashboard," last accessed 17 January 2021, NHS Digital

FIGURE 6. NUMBER OF WEEKLY ROUTINE, URGENT AND 2-WEEK WAIT GP REFERRALS (2020)



Weekly GP referral numbers during 2020 for routine, urgent and 2-week wait referrals. Starting in March, the number for each referral type fell precipitously, and have yet to return to pre-pandemic levels.

This is against a backdrop of a steady decline in patient access to GP services for at least the past decade, which has set the stage for a worsening RTT time.⁵¹ Patients self-report difficulty getting through to their practice over the phone, getting a timely appointment, and being able to see their preferred GP. The decline in the number of full-time GPs is an exacerbating factor, and this in turn contributes to pressure on A&E departments, which some patients resort to attending after struggling to access primary care services.

The number of hospital beds available for A&E departments to admit patients has fallen over the past decade, from 144,455 to 128,935.⁵² This pressure on bed space has manifested with some patients being prematurely discharged or not otherwise having their needs met, with an increasing readmission rate as a result.⁵³

With both primary and secondary care capacity stretched well before the peak of the first wave of the pandemic, and anticipating a surge in Covid-19 patients, instructions were issued to NHS England on 17 March 2020 to urgently discharge patients who were considered medically fit into the community, in order to free up bed space.⁵⁴ Those discharged into care homes contributed to long-standing pressures.⁵⁵ This has been cited as a key contributor to the rapid early spread of Covid-19 in care facilities.⁵⁶

⁵¹ "Access to GP services," last updated 14 August 2020, Nuffield Trust

⁵² "Hospital bed occupancy," last updated 15 October 2020, Nuffield Trust

⁵³ "Emergency readmissions," last updated 17 September 2020, Nuffield Trust

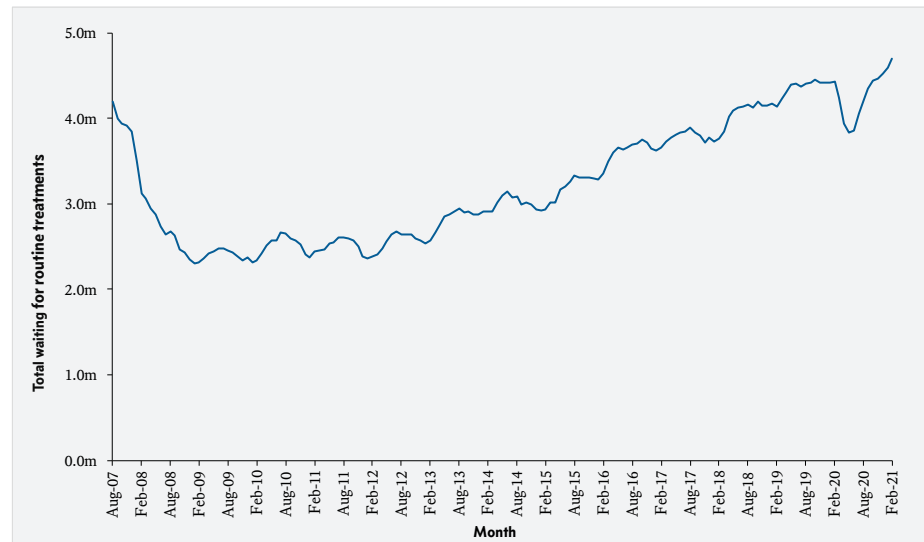
⁵⁴ Sir Simon Stevens, Amanda Pritchard, "Important and urgent – next steps on NHS response to COVID-19," 17 March 2020, NHS England

⁵⁵ "Care home bed availability," last updated 22 July 2020, Nuffield Trust

⁵⁶ Bryan Christie, "Covid-19: 338 patients with the virus were discharged from Scottish hospitals to care homes," BMJ 2020;371:m4225

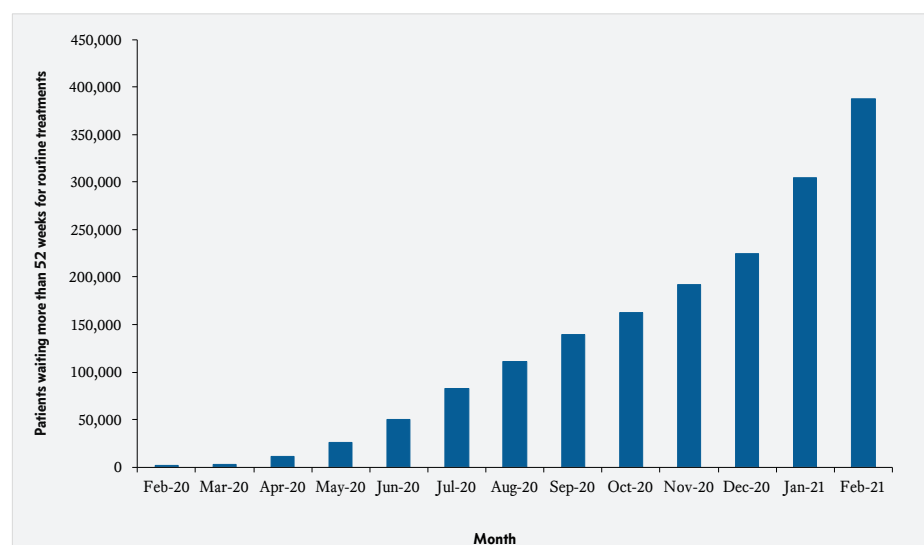
The latest data paints a particularly worrying picture of the impact of the Covid-19 pandemic on waiting for routine treatments. There has been a substantial increase in waiting times under the Referral to Treatment (RTT) pathways. As of February 2021, there were a record 4.7 million people waiting for non-urgent operations and procedures in England. There are now over 388,000 people who have waited more than a year for routine treatments, compared to 1,600 people before the pandemic. This backlog will add substantial additional pressure over the coming years.

FIGURE 7. PATIENTS WAITING FOR ROUTINE TREATMENTS AND PROCEDURES, ENGLAND, AUG 2007 TO FEB 2021



Incomplete Referral to Treatment (RTT) pathways, NHS England

FIGURE 8. PATIENTS WAITING MORE THAN 52 WEEKS FOR ROUTINE TREATMENTS AND PROCEDURES, ENGLAND, FEB 2020 TO FEB 2021



Incomplete Referral to Treatment (RTT) pathways, NHS England

The NHS has encountered many controversies regarding end of life issues. These controversies have been raised again as a result of the unique pressures of the Covid-19 pandemic. The use (and alleged misuse) of end of life treatment pathways and Do Not Attempt Cardiopulmonary Resuscitation (DNACPR) orders are emotive issues which can cause conflict between clinical staff and patients and their relatives.

One of the best known controversies surrounded the Liverpool Care Pathway (LCP).⁵⁷ This care pathway was ultimately withdrawn in 2013 due to public outrage over its application.⁵⁸ Yet the ethical challenges over how to manage those patients unlikely to recover from their current illness, or who are unlikely to benefit from cardiopulmonary resuscitation (CPR) in the event of their deterioration has persisted.

A recent report from the Care Quality Commission (CQC) highlighted a large number of relatives who felt they had not been appropriately consulted when DNACPR decisions were being put in place and the use of inappropriate “blanket” DNACPR decisions early in the pandemic.⁵⁹ Public misconceptions regarding the likelihood of success of CPR attempts certainly contributes to disagreements over DNACPR decisions, but it remains imperative that if they are to be justified ethically, the use of DNACPR must be due to transparent criteria rather than pressure on services.

AMBULANCE SERVICES

Patients self-presenting to A&E are not the only ones whose access to treatment has been restricted. Ambulance time response data⁶⁰ and internal guidance on patient triage⁶¹ during the pandemic paints a picture of a system which was already subject to considerable pressure and decompensated rapidly during the pandemic.

Ambulance calls are triaged into Category 1, Category 2, and Category 3 depending on urgency, with Category 1 calls the most urgent. Category 1 response times have remained relatively stable during the pandemic. But this disguises ways in which access to ambulance services has been rationed.

Guidance was issued during the pandemic implementing a new raised threshold for driving a patient to hospital by ambulance. Thus, some patients who would have previously been admitted based on their illness were denied access. Response times for Category 2 and Category 3 calls increased between January and March 2020, from 21 to 32 minutes and 57 to 90 minutes respectively. Maintaining Category 1 response times has come at the cost of drawing resources away from other

⁵⁷ “Liverpool Care Pathway: ‘They told my family I was dying,’” 15 August 2013, BBC News

⁵⁸ “Overhaul of End of Life Care system,” 15 July 2013, Department of Health and Social Care

⁵⁹ “Protect, respect, connect – decisions about living and dying well during COVID-19,” 18 March 2021, Care Quality Commission

⁶⁰ “Ambulance response times,” 1 May 2020, Nuffield Trust

⁶¹ George Arbuthnott, et al. “Revealed: how elderly paid price of protecting NHS from Covid-19,” 25 October 2020, The Sunday Times

cases. Some GPs were even pressured to deter patients from seeking hospital care in the event of deteriorating clinical condition.

Many other areas of NHS service provision will be affected by the current pandemic and the policy responses which have been implemented. The size of the impact for many of these, such as cancellation of operations, is not yet clear, as the collection and publication of certain official statistics has been halted.⁶² It may take months or years before the full scale of healthcare rationing, due partly to the pandemic and partly to background trends of demand steadily outpacing resources, becomes apparent.

CONCLUSION AND RECOMMENDATIONS

Rationing of treatments and diagnostic procedures, which has been in place for many years through indirect measures, has grown considerably in recent months as the government focuses on building capacity for Covid-19 patients at the expense of other health services.

Once the pandemic is over, even if funding increases considerably, rationing will continue. This will be either direct, by cancelling operations and denying patients services which they otherwise would have accessed prior to the pandemic, or indirectly, by ostensibly offering these services but using mechanisms such as queueing and administrative hurdles to achieve the same endpoint of preventing a patient from receiving care.

Yet the pandemic also gives an opportunity for the UK government to begin to make inroads into much needed reform of the NHS. By having its hand forced by Covid-19, there is a chance to implement changes, based on international best practice, in order to get the logjam of non-Covid cases moving again. Three key steps should be taken to achieve this: the case must be made to the public for the necessity of structural reform post-Covid; competition must be increased and procurement processes streamlined within the UK healthcare system; and the long-term plan for the NHS must be revised to ensure that the vision for its future is feasible and sustainable – at present it is neither.

MAKING THE CASE FOR NECESSARY REFORM

There is an ethical imperative to acknowledge the existence and prevalence of healthcare rationing within the NHS, especially as its encroachment upon emergency department care, cancer treatment times, access to diagnostic testing and primary care worsens. Those treatments, tests and referrals which were deferred due to the pandemic must be identified. Waiting times must be calculated from this data and communicated to the public in a transparent manner.

⁶² "COVID-19 and the production of statistics," last accessed 17 January 2021, NHS England. "Due to the coronavirus illness (COVID-19) and the need to release capacity across the NHS to support the response, we paused the collection and publication of some of our official statistics."

These times will undoubtedly fall short of the constitutional obligations of the NHS Targets and cause some political embarrassment. But keeping patients waiting indefinitely with ever-changing waiting times is ethically unjustifiable and means patients are not able to make informed decisions about their care. By presenting the reality of post-pandemic waiting times, however unpalatable they might be, individual patients will be able to make their own decisions about how to seek health services to best meet their needs.

INCREASING COMPETITION AND STREAMLINING PROCUREMENT

In order to get the backlog of cases moving, it is clear that business as usual for the NHS will not suffice. Structural reforms will be necessary to increase competitiveness and to ensure that patients are able to efficiently progress through their care pathways. International case studies are useful here. Those nations which outperform us in patient outcomes and waiting times with similar per-capita health spending – Australia, Germany, Finland, Denmark, New Zealand, Israel – manage to achieve universal or near-universal coverage alongside a high degree of patient choice from private complementary or supplementary insurance.

The private health sector in the UK is underutilised relative to its international comparators. Only around 10% of patients have private coverage compared to typical figures of 30–80% for the above nations.⁶³ With the overwhelming majority of UK citizens relying on a single provider, if waiting times are breached due to limited capacity, there is little recourse for the patient. Denmark, for instance, gives citizens the right to access public health services within fixed waiting times, and where those targets are missed, the citizen may be referred to a private provider to ensure timeliness of treatment. The UK has no such remedy, and patient care suffers as a result.

All patients whose treatment on the NHS was deferred during the pandemic should be provided with new, binding waiting times. Where the target is not met, the option for referral to private service providers should be made and funded publicly, with procurement methods such as reverse auctions used to ensure good value for money for the taxpayer. Simultaneously, for those citizens who can afford to use private health service providers, and thereby take pressure off the NHS, incentives such as tax rebates should be in place to encourage them to do so. The goal should be to increase patient throughput while maintaining service quality and encouraging those who can afford to shop outside the NHS for elective procedures to do so.

REVISITING THE LONG-TERM PLAN FOR THE NHS

It is clear that the role of the NHS in the lives of British citizens will need to be revised in the aftermath of the pandemic. Failures of service provision have been made clear, and the myth that the NHS does not enforce healthcare rationing has been exposed. This will be exacerbated over the next few years if the UK government insists on sustaining this myth. Instead, the long-term plan for the NHS and its role within the UK healthcare system should be rearticulated. Once the neces-

⁶³ “International health care system profiles,” last accessed 25 January 2021, The Commonwealth Fund

sity of this revision is acknowledged, steps should be taken to reform the NHS and allow it to live up to its promise of access based solely upon clinical need. At present, it is manifestly failing to do so, and underperforming on the international stage.

Political reluctance to tackle a subject as sensitive as the NHS has long been the most significant barrier to meaningful reform. Yet healthcare rationing will increasingly affect citizens' lives as the NHS struggles to cope with the backlog of cases which this pandemic has produced. Access to healthcare will be denied as a result.

It must be a political priority to discuss what steps need to be taken by the NHS to ensure it fulfils its constitutional obligations and to challenge the popular misconception that it does not ration healthcare for its users. Meaningful improvements will be near-impossible until the true scale of rationing is more widely recognised.