ABSTRACT

Objectives: A significant indirect impact of COVID-19 has been the increasing elective waiting times observed in many countries. In England’s National Health Service, the waiting list has grown from 4.4 million in February 2020 to 5.7 million by August 2021. The objective of this study was to estimate the trajectory of future waiting list size and waiting times up to December 2025.

Methods: A scenario analysis was performed using computer simulation and publicly available data as of November 2021. Future demand assumed a phased return of various proportions (0%, 25%, 50%, and 75%) of the estimated 7.1 million referrals “missed” during the pandemic. Future capacity assumed 90%, 100%, and 110% of that provided in the 12 months immediately before the pandemic.

Results: As a worst-case scenario, the waiting list would reach 13.6 million (95% confidence interval 12.4-15.6 million) by Autumn 2022, if 75% of missed referrals returned and only 90% of prepandemic capacity could be achieved. The proportion of patients waiting under 18 weeks would reduce from 67.6% in August 2021 to 42.2% (37.4%-46.2%) with the number waiting over 52 weeks reaching 1.6 million (0.8-3.1 million) by Summer 2023. At this time, 29.0% (21.3%-36.8%) of patients would be leaving the waiting list before treatment. Waiting lists would remain pressured under even the most optimistic of scenarios considered, with 18-week performance struggling to maintain 60%.

Conclusions: This study reveals the long-term challenge for the National Health Service in recovering elective waiting lists and potential implications for patient outcomes and experience.

Keywords: computer simulation, COVID-19, elective backlog, elective performance, waiting lists, waiting times.

Introduction

The COVID-19 pandemic has exerted substantial pressure on healthcare services. Although the disease has mainly directly affected the acute emergency setting, the resulting reduction in hospital resources has constrained the capacity available for elective treatments. To manage the first wave of COVID-19 cases, many countries postponed all nonurgent procedures in public hospitals. Estimates suggest this contributed to more than 28 million operations being cancelled or postponed globally. This and subsequent pressures have led to the increase in waiting times that have been observed in many countries around the world.

Long waits are bad for both patients and healthcare providers. For patients, waiting time associates with worse health outcomes and can result in loss of independence, social isolation, and depression. Health inequalities, already increased during the pandemic, risk further widening with those able to pay, having the option to go private. For healthcare providers, additional and otherwise unnecessary costs may stem from the ongoing management of symptoms up until the point of definitive treatment. Costs may also be incurred for patients who, because of their waiting time, convert to the more disruptive and resource-intensive nonelective route—consider, for instance, the elevated fall risk of a delayed hip replacement or cataract patient. Long waits also affect the economy more widely, given the reduced contribution to the workforce made by those with impaired functional mobility and poorer health.

By Autumn 2021 (the time of writing), mass vaccination against COVID-19 has helped weaken the link between infection incidence and hospital bed occupancy. Although this has provided some stability for elective pathways, much uncertainty remains regarding the situation into 2022 and beyond. A substantial reduction in referrals has been observed in many countries during the pandemic: with societal restrictions relaxed (given the vaccine), how many of these will now return? And although elective capacity has mostly improved, will it be sufficient to effectively attack the backlog amassed in many healthcare systems? Especially if, while weakened, COVID-19 and associated infection control measures continue to restrict the hospital bed...
base beyond prepandemic levels [18–20] with waves of the disease leading to appreciable staff absence, such as that seen during the Omicron wave in late 2021. [21] There is also the matter of whether perennial recruitment issues in the health sector will limit the potential of additional funds pledged by many governments for accelerating elective recovery. [21]

These various factors, essentially relating to elective care demand and capacity, can be hypothetically flexed in generating scenario projections for future waiting list size and waiting times. Projections are important for a number of reasons. They help set realistic expectations for patients, enabling them to plan their lives better and make more informed decisions regarding alternative treatment options (such as conservative management or going private). They can be used by policy makers to more objectively assess the potential pros and cons of different resource allocation decisions, for example, the diversion of extra healthcare funds to elective care may be warranted if projections reveal worse than considered waits. In addition, they can be used to help estimate the additional capacity required to satisfy the aforementioned wait-driven healthcare demands, for example, the number of new community physiotherapists required to manage ongoing symptoms or the extra emergency care resources needed for those converting to the nonelective route.

Against these possible benefits, there has been relatively little interest from the academic community in projecting waiting lists, especially during COVID-19 and at the level needed to inform regional or national policy. Before COVID-19, much of the effort involved bespoke projects at an individual hospital or specialty level through use of operational research techniques such as queuing theory and computer simulation. [22–25] In very few models is the effect of long waits on patient health and behavior captured, which is a particularly important aspect to appreciate in the COVID-19 era.

Of the COVID-19–related studies, early investigators have projected the size of the orthopedic [26] and all-specialty surgery [27] waiting lists in England, albeit without the kind of mechanistic modeling required to estimate waiting times. Although not in the peer-reviewed literature, others have presented conceptual dynamical modeling frameworks for COVID-19 application; however, a lack of representation of individual patient movements prohibits the meaningful consideration of waiting time. [28] Accounting for this within an alternative and more granular dynamical modeling technique, waiting times and list size have been projected at the outset of the pandemic, under various scenarios relating to possible levels of demand and capacity. [29] Nevertheless, this, as with the other mentioned efforts, does not capture the effect of the long waits that have since emerged as the pandemic has evolved on patient behavior.

The objective of this study is to estimate the trajectory of future waiting list size and waiting times in England’s National Health Service (NHS) up to December 2025, under various scenarios considered plausible at the time of writing. The remainder of the article is structured as follows. The “Waiting List Model” section details development and specification of a computer simulation to model the problem. The “Study Setting and Scenarios” section outlines the setting of the study and the range of scenarios considered. In the “Model Calibration” section, the calibration of the model is covered. Projections are presented in the “Results” section. Finally, a discussion on the strengths, limitations, and practical implications is provided in the “Discussion” section.

Methods

Waiting List Model

Although the particular organization of elective waiting lists will differ based on country, specialty, and clinic, [30,31] there are some general principles that can be used to guide model development. First, patients in need of treatment join the waiting list, typically following a general practitioner (family doctor) referral. They are then treated according to their priority (based on clinical need), waiting time (elapsed time since referral), and complexity (in terms of healthcare activity required—some will require much more than others). Aggregate treatment rates are limited by available resources, for example, diagnostic capacity, the number of consultants and surgeons, and operating theater slots. Finally, some patients will leave the waiting list before treatment as waits increase.

Computer simulation is used to model this dynamical behavior. The 2 key model inputs are the future number of referrals and capacity for treatment, both of which can be varied within a scenario analysis. The Poisson distribution is used to capture variability that would reasonably be expected in the referral rate, with the distribution mean set equal to the given daily number of referrals according to the scenario in question. This essentially assumes that one referral is independent of another and is a common choice in modeling variability in healthcare arrivals. [32] To capture patient complexity, referrals are partitioned on arrival to one of a number of classes, each of which awards a different number of points to the referral. The patients selected for treatment on each simulated day (bounded above by capacity for that day) are determined probabilistically with chances of selection guided by the number of points (e.g., if there are 2 waiting patients—of classes awarding 10 and 20 points, respectively—and there is capacity to treat only one, then the first is selected with a 1-in-3 probability and the second with a 2-in-3 probability). This ensures that referrals of the highest-point classes, representing the least complex cases, have the greatest chance of being selected for treatment and thus have the least waiting time. It also ensures that treatment order reflects waiting time because it is more difficult for waiting patients to escape selection for treatment with each passing day.

As previously mentioned, patients may also leave the waiting list before treatment. This is known in queuing theory parlance as reneging. To address this, it is assumed that each waiting patient has a waiting time–dependent probability of reneging, which is sampled on each simulated day. The probability is calculated using a sigmoidal function, whose value increases with waiting time. Specifically, the cumulative Weibull distribution function is used, with shape and scale parameter given by $k$ and $l$, respectively. If the patient is determined to have reneged, they are removed from the waiting list. Within each simulated day, reneging occurs before treatment selection with new referrals being simulated last. This ensures that a patient cannot renege or be selected for treatment on the same day as their referral.

A single run of the simulation from the first to the last day in the considered projection period accounts for just one way in which events could pan out. To capture realistic variability and uncertainty, the random number seed used in the simulation loop is changed and another run is performed. This may mean a different number of arrivals on each day, a different selection of patients, and different patients reneging at different times. After a sufficient number of runs are performed, model outputs are combined to calculate the various projections of waiting list size.
waiting times, and the number of patients reneging for the scenario in question. Appendix 1 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2022.06.016 summarizes the simulation approach through pseudo code. Appendix 2 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2022.06.016 contains the completed research checklist appropriate to this type of study (the Strengthening The Reporting of Empirical Simulation Studies guidelines).33

Study Setting and Scenarios

The above model is applied to the elective waiting list of England’s NHS as of August 2021 (the latest month of available data at the time of the study). At such time, elective waits had shown some signs of improvement although the backlog of patients awaiting treatment was still rising (Fig. 1). Specifically, the 18-week referral-to-treatment (RTT) statistic, measuring the percentage of patients waiting under 18 weeks, had improved from its 46.8% low in July 2020 and stabilized at approximately 68%. The RTT metric is the principal barometer of elective performance in the NHS and is assessed against a 92% constitutional target.34 It is increasingly more instructive to consider the numbers waiting over 52 weeks, which increased 270-fold from February 2020 (before the pandemic) to its peak in March 2021. All data were obtained from a publicly available central NHS source.35

The scenarios examined through the model (Table 1) are defined over the period from September 2021 to December 2025 and relate to the 2 previously mentioned “unknowns” at the time of the study: the proportion of missed referrals returning and future capacity for treatment (“Introduction” section). Although monthly referrals had returned to prepandemic levels by August 2021, an estimated 7.1 million had been “missed” since March 2020.36 In our analysis, we consider that 0%, 25%, 50%, and 75% of these will return within the 12 months from September 2021 (normally distributed, μ = 182 days, σ² = 55, and truncated between 0 and 365 days) on top of a “baseline” referral rate equivalent to the mean from March 2019 to February 2020 (ie, the last full 12 months before the pandemic). An increasing referral growth rate based on the rate observed in the 5 years prepandemic is also considered, which is linear (P < .001) and increases by approximately 50,000 each year (this represents 2.8% annual growth in the latest prepandemic year). Relative to total elective treatments from March 2019 to February 2020, we consider future capacity at 90%, 100%, and 110% for the period until December 2025 (phased in over the first 6 months). The lower limit reflects constrained elective capacity assuming sustained COVID-19 demand, whereas the upper limit accounts for success in

Table 1. Scenarios examined within this study.

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Proportion of referrals “missed” during the pandemic to return in 12 months from September 2021, %</th>
<th>Elective capacity relative to the number of treatments performed in 12 months to February 2020, %</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>110</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>100</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>90</td>
</tr>
<tr>
<td>4</td>
<td>25</td>
<td>110</td>
</tr>
<tr>
<td>5</td>
<td>25</td>
<td>100</td>
</tr>
<tr>
<td>6</td>
<td>25</td>
<td>90</td>
</tr>
<tr>
<td>7</td>
<td>50</td>
<td>110</td>
</tr>
<tr>
<td>8</td>
<td>50</td>
<td>100</td>
</tr>
<tr>
<td>9</td>
<td>50</td>
<td>90</td>
</tr>
<tr>
<td>10</td>
<td>75</td>
<td>110</td>
</tr>
<tr>
<td>11</td>
<td>75</td>
<td>100</td>
</tr>
<tr>
<td>12</td>
<td>75</td>
<td>90</td>
</tr>
</tbody>
</table>

Note. Approximately 7.1 million elective referrals were “missed” during the pandemic from March 2020 to August 2021 and that 16.6 million treatments were performed in the 12 months to February 2020 (the last full year before COVID-19 was declared a pandemic).
government plans to increase elective capacity. No clear cyclical behavior could be detected in monthly elective treatments, and therefore, no seasonality is assumed for capacity.

**Model Calibration**

For parsimony, 2 classes are assumed for representing patient complexity, each of which must be assigned a probability of selection and a number of points (“Waiting List Model” section). Given that assignment to class A and B are complementary, it is necessary to define only one probability parameter (e.g., if class A probability is 0.8, then class B probability is 0.2). Given also that it is the relative difference in points that informs selection of patients for treatment, it is only necessary to define the number of points awarded to 1 class (here, the other is set equal to one).

These 2 model parameters relating to the complexity classes, alongside the 2 relating to the reneging function (k and l), were estimated through fitting to the latest 6 months’ data for waiting list size, 18-week performance, and numbers waiting over 52 weeks (Fig. 1). Accuracy was determined through mean absolute percentage error (MAPE), with larger weights applied to more recent data (according to an exponential decay function with a 1-month half-life). Modeling upon the actual referrals and capacity over the 6-month period, the accuracy of 1 million distinct parameter sets was computed. For each of the 1 million simulations, MAPE was calculated from the mean of 100 performed runs.

Figure 2. Model projections (mean and 95% confidence intervals), accounting for the pre-pandemic referral rate (March 2019-February 2020) with no trend growth and various proportions of the estimated 7.1 m referrals “missed” during the pandemic returning in the 12 months from September 2021.

m indicates million; RTT, referral-to-treatment.
All computational processes used R version 4.1.1 on a Microsoft Windows machine (i9-9980XE CPU, 64GB RAM).

Of the 1 million parameter sets, the top 1000 (by MAPE) were selected for use in obtaining the scenario projections (“Results” section). This limited over-reliance on one or a small number of parameterizations, which could otherwise promote undue confidence in model outputs. To reasonably address such uncertainty, projections included the mean of the 1000 trajectories and 95% confidence bands based on the 2.5% and 97.5% quantiles. Appendix 3 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2022.06.016 contains a comparison of model projections and actual values over the 6-month calibration period. This supports goodness of fit for waiting list size and 18-week performance, with all actual values contained within the modeled confidence bands. This also holds for the most recent 4 months of the number waiting over 52 weeks, although there is some larger discrepancy for the first 2 months. Parameter estimates, calculated from the top 1000 selected parameter sets, give a 0.90 mean (95% confidence interval 0.86-0.92) for the probability of a referral being assigned to complexity class A with a 110 mean (10-200) points awarded. One interpretation of this larger and less complex class is a lack of need for significant surgical procedure, which is supported by a similarly large proportion of elective treatments being carried out without hospital admission.35 The reneging parameter estimates give a 2.8 mean (2.4-3.0) for k and a 2654 mean (2025-3000) for l. Parameter estimates, including parameter correspondence, is further detailed in Appendix 4 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2022.06.016. Appendix 5 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2022.06.016 diagrammatically summarizes the overall modeling approach taken in this study.

**Results**

Assuming no long-term trend growth in referrals, the waiting list would peak at 13.6 million (95% confidence interval 12.4-15.6 million) in Autumn 2022 (Figs. 2 and 3) if 75% of referrals missed during the pandemic returned in the 12 months from September 2021 and only 90% of prepandemic elective capacity could be provided (scenario 12; Table 1). If 110% capacity can be achieved, then peak waiting list size would be reduced by approximately 2 million (scenario 10). If none of the missed referrals returned then, under the most optimistic scenario of those considered, the waiting list would reduce over the longer term, reaching 5.2 million (3.0-6.7 million) by December 2025 (scenario 1). In general, results show that regardless of the number of missed referrals that return, the waiting list will over time stabilize to levels determined by capacity (Figs. 2 and 4). This is owing to higher rates of reneging from the waiting list in the shorter term when the waiting list—and moreover waiting times—are larger. This is most prominent for scenario 12, where the proportion of patients reneging increases threefold to peak at 29.0% (21.3%-36.8%) in Spring 2023. Note that the estimated reneging function, inferred...
through model calibration ("Model Calibration" section), can be found in Appendix 6 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2022.06.016.

If 75% of missed referrals returned, then waiting time performance could fall <50% when measured by the 18-week RTT metric (scenarios 10–12). This compares with the 46.8% low recorded in July 2020 (Fig. 1). Over the longer term, above 60% performance can be achieved provided 110% capacity can be sustained. Indeed, above 60% performance could be maintained throughout the period should no missed referrals return. Note that the brief spikes in 18-week performance in Spring 2022—most noticeable for scenarios involving a greater return of missed referrals—are an artifact of the 18-week RTT equation, given the flood of new referrals which temporarily increase the numerator (number of patients waiting 18 weeks). Although the 18-week metric represents the constitutional target, it is increasingly more instructive to look at numbers waiting over 52 weeks. As a worst-case scenario, this would peak at 1.6 million (0.8–3.1 million) in Summer 2023 (scenario 12). If 110% capacity is provided, numbers can be limited to within 0.5 million over the longer term (c.f. 436 127 recorded in February 2020).

Accounting for referral growth at the rate observed in the 5 years prepandemic, results indicate a much greater challenge in restoring elective performance (Fig. 5), even under the most optimistic of scenarios considered, modeling suggests that capacity would be insufficient to prevent a sustained increase in waiting list size and deterioration in patient waiting times over the longer term.

Discussion

Despite the significant indirect impact of COVID-19 on elective performance, and the attendant media interest and public concern, there has been a deficit of applied modeling to project the future state of waiting lists and what this could mean for patients and healthcare services.

The results of this study reveal the long-term challenge for the NHS in recovering elective waiting lists following the early impacts of COVID-19. Even without accounting for a long-term trend growth in referrals, these projections raise numerous concerns for healthcare systems (Fig. 2). First, the additional resource requirements for managing ongoing symptoms: recommendations for long-waiting joint replacement patients include extra physiotherapy, dietetics, orthotics, psychology, and pharmaceuticals—all of which come at a cost. Second, the increased likelihood of converting to the costlier nonelective route: for instance, waiting over 20 weeks for planned cholecystectomy associates with a threefold increase in emergency treatment need when compared with lesser waits. Third, the exasperation of health inequalities: post–COVID-19 research conducted in England has found those living in more deprived areas are 1.8 times more likely to experience a 1-year wait than those from less deprived areas, with 47% of people surveyed unable to go private. Fourth, the negative economic impacts: while difficult to quantify, studies have recognized the wider implications of cataract treatment delays on employment, ability to care for others and the costs of increased motor vehicle crashes. The need to address these challenges will be greater should referrals grow according to the longer-term trend evident before the pandemic (Fig. 5).

Increasing capacity is a possible solution to address waiting list growth and improve waiting times. In October 2021, the UK Treasury pledged a further £6 billion to address the elective backlog in England, on top of £8 billion already earmarked over 3 years. We calculate, using an assumed £2933 average elective pathway cost, that this £14 billion would buy extra capacity to treat a total 4.8 million patients, equivalent to 1.6 million annually. This amount is approximately equivalent to the additional

Figure 4. Long-term convergence of mean waiting list projections across the considered capacities, accounting for the prepandemic referral rate (March 2019-February 2020) with no trend growth and various proportions of the estimated 7.1 m referrals “missed” during the pandemic returning in the 12 months from September 2021. m indicates million.
requirement for scenarios involving 110% of prepandemic capacity, given the 16.6 million treatments performed in the 12 months before the pandemic. Nevertheless, our modeling suggests that more investment would be required to improve long-term 18-week performance much beyond the 60% level. Along with financial budget, attracting and retaining a suitable workforce is another challenge that must be met to increase capacity. Consideration should also be given to the shorter-term effect of further COVID-19 waves on staff absence, given significant sickness among NHS staff because of the Omicron variant in late 2021. Other options include improving productivity and, of course, reducing demand, potentially through greater promotion of conservative management approaches.

Turning to limitations, although a satisfactory model fit to historical data has been possible for waiting list size and 18-week performance, there is a less convincing approximation to numbers waiting over 52 weeks (“Model Calibration” section). Further work may be used to explore whether greater accuracy could be achieved with the use of additional complexity classes, beyond the 2 considered here. It should also be acknowledged that although patient complexity is captured, the current model does not appreciate referral priority and its influence on treatment order. Future investigators may wish to explore the extent to which incorporation of this may improve model accuracy. Nevertheless, given the various ways that different specialties and clinics prioritize treatment for waiting patients, it is likely to be difficult
to capture priority within a national, specialty-wide application of the model.

Alternatively, modeling may be conducted at an individual specialty level, which could allow for incorporation of specialty-specific referral priorities and moreover could provide greater accuracy in model calibration (given the ability to obtain better model fits to data generated by more homogeneous dynamical processes). Results could thereafter be aggregated for a specialty-wide view of the format presented in this study. Modeling at this granularity would also allow for consideration of specialty-specific demand and capacity factors, that is, particular known or expected resource constraints or referral trajectories. If projected waiting times were penalized in a specialty-specific manner, then service level capacity allocations could be optimized. In addition to requiring some form of wait-dependent outcome quantification (eg, to what measurable extent could a 52-week cardiology wait be “worse” than a 52-week dermatology wait), future investigators should also be cognizant of the additional data requirements and complexity of such an approach.

Supplemental Materials

Supplementary data associated with this article can be found in the online version at https://doi.org/10.1016/j.jval.2022.06.016.

Article and Author Information

Accepted For Publication: June 14, 2022
Published Online: xxxx
doi: https://doi.org/10.1016/j.jval.2022.06.016

Author Affiliations: Modelling and Analytics, Bristol, North Somerset and South Gloucestershire, Clinical Commissioning Group, National Health Service, Bristol, England, UK (Howlett, Wood); Centre for Healthcare Innovation and Improvement, School of Management, University of Bath, Bath, England, UK (Wood).

Correspondence: Richard M. Wood, PhD, Modelling and Analytics, Bristol, North Somerset and South Gloucestershire, Clinical Commissioning Group, National Health Service, South Plaza, Marlborough Street, Bristol, England BS1 3NX, United Kingdom. Email: richard.wood16@nhs.net

Author Contributions: Concept and design: Howlett, Wood
Acquisition of data: Howlett, Wood
Analysis and interpretation of data: Howlett, Wood
Drafting of the manuscript: Wood
Statistical analysis: Howlett
Supervision: Wood

Conflict of Interest Disclosures: The authors reported no conflicts of interest.

Funding/Support: The authors received no financial support for this research.

Acknowledgment: The authors are grateful to the anonymous reviewers for their helpful comments and suggestions that have improved the quality and legibility of this article.

REFERENCES

1. Iacobucci G. COVID-19: all non-urgent elective surgery is suspended for at least three months in England. BMJ. 2020;368:m1106.


41. Wood RM. Unravelling the dynamics of referral-to-treatment in the NHS. Health Syst. 2021;10(2):131–137.

42. Oliver D. Do the NHS workforce plans really add up? BMJ. 2020;370:m3203.