A predictive model for the post-pandemic delay in elective treatment

Romy Nehme *, Alena Puchkova, Ajith Parlikad

Institute for Manufacturing, University of Cambridge, 17 Charles Babbage Road, Cambridge, CB3 0FS, UK

ABSTRACT

The COVID-19 pandemic had a major impact on healthcare systems across the world. In the United Kingdom, one of the strategies used by hospitals to cope with the surge in patients infected with SARS-Cov-2 was to cancel a vast number of elective treatments planned and limit its resources for non-critical patients. This resulted in a 30% drop in the number of people joining the waiting list in 2020–2021 versus 2019–2020. Once the pandemic subsides and resources are freed for elective treatment, the expectation is that the patients failing to receive treatment throughout the pandemic would trigger a significant backlog on the waiting list post-pandemic with major repercussions to patient health and quality of life. As the nation emerges from the worst phase of the pandemic, hospitals are focusing on strategies to prioritise patients for elective treatments. A key challenge in this context is the ability to quantify the expected backlog and predict the delays experienced by patients as an outcome of the prioritisation policies. This study presents an approach based on discrete-event simulation to predict the elective waiting list backlog along with the delay in treatment based on a predetermined prioritisation policy. The model is demonstrated using data on the endoscopy waiting list at Cambridge University Hospitals. The model shows that 21% of the patients on the waiting list will experience a delay less than 18-weeks, the acceptable threshold set by the National Health Service (NHS). A longer-term scenario analysis based on the model reveals investment in NHS resources will have a significant positive outcome for addressing the waiting lists. The model presented in this paper has the potential to be an invaluable tool for post-pandemic planning for hospitals around the world that are facing a crisis of treatment backlog.

1. Introduction

The Coronavirus SARS-Cov-2 (COVID-19) pandemic resulted in an unprecedented global health, economic, and social crisis and has exposed the vulnerabilities of healthcare systems around the world due to the scarcity of hospital resources. Many healthcare systems, including the NHS in the UK, diverted elective treatment resources towards the frontline as a main mitigation strategy throughout the pandemic. Consequently, a significant number of elective patients accumulated on the waiting list. For instance, in February 2021, 4,698,348 waiting list patients were recorded across England, 35.5% of which incurred a delay exceeding 18-weeks [1], the acceptable threshold set by the NHS.

The implications of prolonged waiting time include increased mortality rates for patients, especially non-prioritised patients [2–4]. A recent study [5] examining historical data across England found that an average two month delay per cancer patient

(considering a three-month lockdown) would result in 2.25% additional lives lost with a backlog on referrals of 25%. Similar studies on chronic cardiac disease [6] indicated significantly higher mortality rates among patients pending structural procedures.

Given these implications, the main factors hindering elective treatment as COVID-19 stabilises are: shortages in staff due to sickness or enforced quarantine; inefficiencies in the supply-chain of surgical material (e.g. consumables, equipment); restricted availability of suitable operating theatres; rise in expenses for patients and insurance companies due to elaborate treatment protocols; and triaging non-emergency cases according to the risk–benefit ratio for the patient and community [7].

In fact, Brown et al. [8] highlight the limitations within elective treatment services, particularly endoscopy, well before the pandemic exposed the healthcare system's vulnerabilities. The authors cite 'issues with data availability, quality and use' as a recurring challenge. Particularly, the endoscopy datasets should be more comprehensive to present an indication of the reason the patient was referred for treatment. The authors also consider the process of accessing and processing these datasets complex due to inconsistent publishing schedules and formats across different datasets. Aside from how the data is presented and its timeliness, Greenhalgh et al. [9] and Guerriere [10] argue that a key

issue is the absence of sufficient skills to analyse the available performance data and translate it into operational improvements.

This study pursues the lack of data or visibility on the waiting list backlog to incite operational improvements across the recovery of elective waiting lists post-pandemic. Particularly, tracking the backlog would inform policy planning (e.g. challenge thresholds for acceptable waiting time), short- and long-term investments in expanding hospital resource capacities, short-term hospital resource allocation, as well as offer estimates on the expected delay in treatment for waiting list patients. To address these issues, a predictive model for the delay in post-pandemic elective treatment is introduced based on a pre-determined scheduling policy. The proposed model-based on discrete-event simulation — estimates the short- and long-term trajectory of the elective waiting list to gauge the time period and resources required for recovery. The model is applied to the endoscopy department of Cambridge University Hospitals (CUH) as a case study to demonstrate its practical capability.

The main research aims of the paper include: (i) presenting a flexible granular simulation model that can be adapted to different waiting list use cases and optimisation objectives; (ii) offering short- and long-term delay time and waiting list statistics that can be filtered down to different patient characteristics (e.g. priority class, elective procedure); (iii) informing waiting list policy planning and resource allocation through highlighting pain points (e.g. types of patients waiting longer, short resources); and (iv) gauging the benefit of post-pandemic treatment capacity investments limiting prolonged patient delays.

This paper is structured as follows: Section 2 presents a review of the literature, Section 3 establishes the predictive model, and sections 4 and 5 detail the undertaken case study along with the findings. Sections 6 and 7 finally list the limitations of the study and conclude potential future prospects.

2. Literature review

In the literature, various studies have addressed patient waiting list problems and optimisation of resources in hospitals. Daldoul et al. [11] propose a stochastic mixed integer programming model to optimise resources such as physicians and nurses in order to minimise patient waiting time in the emergency department. The impact of transferring non-urgent patients as a result of emergency department overcrowding is examined by Nezamoddini et al. [12]. The transfer minimises patient waiting times while maintaining the same capacity of resources. The model suggested in Oh et al. [13] was used to investigate problem areas concerning process flow, resource allocation, and operational policies in order to optimise an emergency department's throughput. Bhattacharjee and Ray [14] propose a model of a hospital appointment system that recognises different patient classes for a radiology department's scanning machine. Multiple appointment system policies are evaluated with patient waiting time and resource utilisation being the monitored criteria for success. Both models [13,14] are implemented using discrete-event simulation. Mahmoudzadeh et al. [15] target rising waiting times for healthcare services by investigating patient scheduling policies using a robust optimisation approach.

Data analytics have been widely explored in healthcare to fulfil a multitude of patient care targets. Among these targets, waiting time serves as a common measure for a healthcare system's performance. In fact, most NHS trusts have continuously exceeded their maximum waiting targets [16]. As a result, waiting time predictions have been well-explored within pre-pandemic literature.

2.1. Pre-pandemic predictive models

Several authors adopted a machine learning methodology to predict patient waiting time among other KPIs. Kaul et al. [17] conduct a comparative study among different prediction-based algorithms (CARE, COHESY, and HARM) highlighting their key features and limitations. While this study [17] explores different machine learning algorithms, Curtis et al. [18] propose a universal predictive model for patient waiting time by assessing nine machine learning algorithms according to a set of evaluation criteria (e.g. root mean square error). The research objectives comprise encouraging superior staff responsiveness and patient satisfaction. Sun et al. [19] employ Quantile Regression to predict waiting time in real-time given a certain set of patient characteristics (e.g. patient acuity) to improve patient satisfaction and quality of care. The resulting forecasts incurred a prediction error of only 9 to 16 min. Joseph et al. [20] and Goncalves et al. [21] both adopt the Random Forest Regression methodology to predict waiting time and pinpoint the variables with the highest predictive power.

While these studies have generated reasonably sound results (e.g. accuracy of 50.09% [21]), other authors have approached the problem using more holistic methodologies to mitigate the limitations of 'black box' machine learning techniques. In fact, Liu et al. [22] isolate the root cause of macro-level features (e.g. high waiting time and length of stay) using agent-based simulation modelling to improve resource allocation and strategic planning. Babashov et al. [23] adopt a discrete-event simulation model highlighting system bottlenecks and quantifying the impact of resource levels to reduce patient waiting time. Chong et al. [24] model patient flow using a system dynamics model to examine the trade-off between waiting time, occupancy, and safety outcomes.

The studies listed above offer significant insight on healthcare predictive models. However, when it comes to the application pursued in this paper, a strong historical basis – an imperative and universal factor for the accuracy of these models – for pandemic patient delays sufficient for isolating patterns is lacking. The rest of this section explores approaches from existing literature on overcoming this shortcoming and adopted predictive models for unprecedented events.

2.2. Pandemic predictive models

Throughout the pandemic, predictive analytics have been utilised to forecast infection rates, admissions rates, and waiting list size among a multitude of KPIs.

When it comes to predicting waiting list size and the corresponding costs, authors have relied on aggregated hospital data to draw out conclusions. In fact, based on NHS data, Macdonald et al. [25] estimate the building backlog of elective procedures to accentuate the current state of waiting lists and propose solutions. The authors highlight the diminishing compliance with pre-pandemic waiting time standards (87% in 2019 vs. 83.2% in 2020). They anticipate 400,000 cases to be missed every month. Sud et al. [5] also examine the impact of different scenarios on the waiting list backlog for cancer referrals based on historical age- and cancer stage-stratified data. For every prospective level of backlog (ranging from 25% to 75%), the number of lives potentially lost is quantified. The implications of capacity investments are similarly measured in terms of the number of lives potentially saved. Fowler et al. [26] model the expected number of surgeries performed, the resources required, and the cost of delayed surgery. Garcia-Rojo et al. [27] conduct an observational descriptive study to assess the impact of the pandemic on urology surgical waiting lists in high-volume hospitals. The

resulting mean waiting time observed substantially exceeded that of 2019 as well as acceptable delay thresholds in some cases. However, given the descriptive methodology design adopted in these studies, the analysis lacked a long-term outlook on patient waiting time.

Other authors approached the problem using more analytical methodologies. Oussedik et al. [28] model the orthopaedic pathway as a dynamic system. The research objective comprises estimating the potential number of patients on elective waiting lists and proposing recovery strategies. The model suggests that while financially burdensome, expanding service capacity by 30% substantially reduces the time required to restore pre-COVID waiting list levels. Joshi et al. [29] employ machine learning (based on an optimisation algorithm) fed by internal hospital data to create an interactive predictive analytics tool. The tool offers real-time estimations on the expected backlog clearance time, overtime required, and potential costs associated with backlog reduction, assuming an optimal assignment of resources. While both studies focus on resource availability and allocation, the model proposed in this paper offers estimations on delay time which would highlight the groups of patients requiring resource investments.

Wood [30] also approach predictions using a more systematic methodology, discrete-time simulation. The research objective is to quantify the implications of the pandemic on the backlog and estimate the service capacity required to restore pre-pandemic waiting list levels. The monitored outputs concern the waiting list size and the proportion of waiting patients in compliance with the 18-week standard. Given that the study was conducted back in March 2020, a series of simplifications were considered for the distribution of patient arrivals and priority classes. Ho et al. [31] serves as another study that directly pursues pandemic waiting lists. The model anticipates the cumulative deficit between the actual and expected procedures based on historical data, a method similarly considered in this paper. The authors postulate that even with mitigation measures, it could take longer than a year to eliminate the implications of the pandemic. Given that the model relies on national aggregated patient data, high-level conclusions regarding the overall trajectory of the waiting list are presented. Thus, the limitations restricting these studies [30,31] lies within the limited granularity in monitoring waiting time, irrespective of patient characteristics.

To summarise the above, the main limitations of both pre- and post-pandemic predictive models are the limited granularity in monitoring waiting time, irrespective of patient types and characteristics, and the lack of a long-term outlook on the trajectory of the waiting list post-pandemic.

Many papers [13,32], and [23] highlighted the value of simulation modelling and, in particular, discrete-event simulation as a highly informative tool for healthcare planning to minimise waiting lists and maintain acceptable waiting time thresholds, as well as to investigate the impact of what-if scenarios on various KPIs. The authors in Salmon et al. [33] provide a literature review of applications of simulation modelling to emergency departments. The rise in recent publications indicate the potential of simulation in reaching a clinical audience.

The model introduced in this paper addresses the limitations identified above using simulation modelling given its model-centricity as opposed to data-centric machine learning algorithms. That is, rare disruptions or soft factors (e.g. waiting list patient mortality) with little data basis to be picked up by pattern recognition can also be incorporated for accuracy. The model employs predictive analytics supported by waiting list policy planning to offer a short- and long-term outlook on the waiting list backlog, which directly corresponds to the research aims targeted by this study.

3. Model

The proposed model adopts discrete-event simulation to represent the waiting list system. This methodology is considered given its granular representation of the system, breaking it down to patient-by-patient events marking referral to treatment, entering the waiting list, receiving treatment, and leaving the system. This granularity also encourages the inclusion of different patient characteristics that define each event (e.g. a patient requiring a colonoscopy would spend 2.5 h receiving treatment versus 3 h in the case of gastroscopy patients). The patient characteristics would then characterise the predictions of delays in treatment generated by the model. The objectives of the model are detailed in Section 3.1 followed by a description of the model logic in Section 3.2. A comprehensive overview of the case study is then presented including the data used, the experimentation performed to set up the simulation, and the final implementation of the model (Section 4). This structure follows the 'Strengthening the Reporting of Empirical Simulation Studies' (STRESS) guidelines for discrete-event simulation as presented in [34].

3.1. Objectives

3.1.1. Model purpose

The objectives of this model are threefold: (i) estimating the average delay time for each patient type in the short-term, (ii) estimating waiting list levels in the long-term for a given treatment capacity and a set of resources, and (iii) studying the impact of inflated treatment capacities.

3.1.2. Model outputs

In fact, the model offers a wide range of outputs collected while the model is running. The following data was particularly collected to gain a short- and long-term outlook on the expected delay time for elective patients:

1. Short-term Performance and Delay: Proportion of patients waiting less than 18-weeks and average delay in treatment
2. Short-term Patient-specific Delay: For every patient type, confidence intervals for the average delay in treatment
3. Long-term Waiting List Trajectory: Waiting list levels for a pre-determined timeframe

3.1.3. Experimentation aims

The sensitivity of these outputs in regards to treatment capacity is also studied using the model. Different levels of increased treatment capacity are considered with the objective of determining the most optimal level to mitigate the implications of the pandemic. These scenarios are further explained in Section 5.4.

3.2. Logic

3.2.1. Base model overview and logic

To represent the waiting list system and break it down to different events dictating a patient's journey, the model consists of two separate yet interdependent flows: patient and signal flow. The patient flow represents patients joining the waiting list with a set of characteristics/attributes and undergoing treatment. This flow allows the model to keep track of granular patient-by-patient time accounts.

The signal flow serves as a series of periodic triggers for waiting list updates (e.g. updating the priority number of waiting patients every five minutes) and patient releases to treatment when resources are available. These entities support the collection of waiting list KPIs.

Fig. 1 summarises the overall logic of the model. The sensitivity analysis in Section 5.4 adopts the same methodology with the only variable being treatment capacity or the number of patients that can be treated in a given time period.
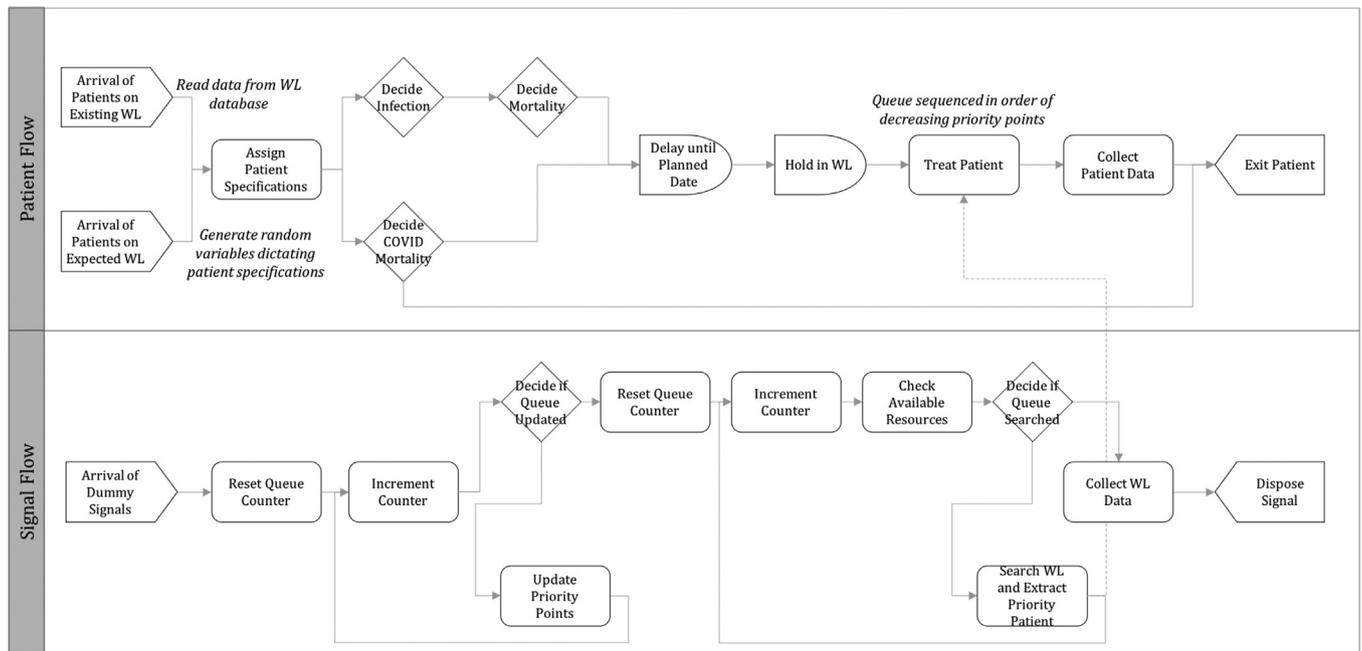
Fig. 1. Discrete-event simulation model blueprint.

3.2.2. Components

1. **Entities:** Patient entities represent actual patients joining the waiting list and undergoing treatment. They are characterised by a set of attributes such as priority, waiting list lane, procedure required, and time of arrival. The objective of this flow is to collect the patient-by-patient data, most importantly the time spent in the waiting list. Signal entities represent dummy time-based signals with the objective of periodically collecting waiting list KPIs, most importantly the number of patients waiting.

2. **Entity Activities:**
Patient Flow: When it comes to patient entities, two types of patient arrivals are considered: current waiting list and expected waiting list arrivals. Once current waiting list patients enter the system, they are assigned a set of characteristics or attributes (e.g. priority class, type of treatment, elective planned date, and time of arrival) indicated by the actual waiting list.

Once expected waiting list patients arrive, a priority class and type of elective treatment is assigned by generating a random variable dictated by a predetermined distribution. The time of arrival is also recorded while an elective planned data is allocated based on the elective treatment sought out by the patient. Depending upon the application, a set of additional characteristics relevant to the treatment time or type of resources required can also be assigned.

After the waiting list arrivals along with their specifications have been established, a series of decisions dictating the patient's potential COVID-19 infection and/or mortality are undertaken. When it comes to mortality, patients leave the system directly. In the case of infection, patients' mortality is similarly reassessed with a higher mortality rate. Otherwise, the patients are considered sufficiently healthy to endure the delay time within the waiting list. Healthy patients are directly transferred to a hold (representing the waiting list) until their elective planned date approaches. Once they are due to receive treatment, patients are transferred to a queue. The patients are released for treatment in accordance with the signal system detailed in the next section. Upon treatment, the patient seizes a series of resources (e.g. clinicians, theatres) based upon the assigned type of elective treatment. Thereafter, the patient leaves the system.

Signal Flow: Dummy signals arrive every interval (e.g. every five minutes) to trigger priority updates and patient treatment signals. Within every iteration, the priority number of each patient within the waiting list is updated based on the prioritisation classification policy. For the CUH case study, the calculation of priority corresponding to their pre-determined classification system is detailed in Section 4.3.1.

Thereafter, each signal verifies whether enough resources are idle to treat patients. If this condition is satisfied, the patient seeking the available resources with the highest priority is released towards treatment. The signal similarly enters a loop to ensure that all patients in the queue requiring currently idle resources are directed towards treatment. Otherwise, the signal is directly disposed of.

3. **Resources:** Treatment resources include the clinicians and theatres required for different types of elective procedures.

4. **Queues:** The waiting list queue holds all patients ready for elective treatment waiting for available resources. These patients are sequenced in order of decreasing priority number.

4. Case study overview

The model introduced in this study serves a multitude of practical applications in mitigating waiting list backlog implications and planning for recovery. The specific case of endoscopy has been considered in this study given its substantial waiting lists relative to other types of elective treatment. In fact, during the first six weeks of the pandemic, Longcroft-Wheaton et al. [35] reported an 88% reduction in treatment capacity within endoscopy in the UK. The pandemic is assumed to end on July 19th, 2021, when all restrictions have been abolished in the UK.

CUH waiting lists serve as the only data source to populate the proposed model. As it stands now, within CUH's endoscopy department, a sample size of 10,194 patients has been recorded

as incomplete pathways, 45% of which incurred a delay exceeding 18-weeks so far. Referring to Section 1, the aforementioned figures validate the comparability of the selected application to the overall state of waiting lists in England.

4.1. Model inputs

The model inputs include patient-specific and hospital-specific datasets. The former define the mix and volume of patients arriving to the waiting list system including:

1. Current Waiting Lists: Current backlog of patients on the waiting list with a set planned date for treatment. The patient characteristics defining this set of patients include the following:

   Waiting list lane. The current waiting list provided by CUH's endoscopy department is characterised by two lanes modelled in this study: *elective planning* (patients that have already been diagnosed and are looking to follow-up on their diagnosis around a certain elective planned date), *elective diagnostic* (patients looking to get diagnosed as soon as possible, no elective planned date).

   Pathway. The pathway defines the type of elective treatment the patient is seeking. The considered pathways include routine, two-week wait, urgent, and cancer pathways.

   Endoscopy procedure category. Endoscopy procedures fall under the following three main categories: gastroscopy, colonoscopy, and flexi sigmoidoscopy. These categories exclude specialities such as hepatology and bronchoscopy which have minimal impact on the flow of patients as they require speciality resources (e.g. clinicians) not shared with other endoscopy procedures. Theatre time for these cases is also blocked out separately and are not included in this model's treatment schedules.

   Priority class. The defined priority classes are listed as follows in order of decreasing priority: P1 — Urgent, P2 — High, P3 — Moderate, P4 — Low, P5, and P6. P5 and P6 are characterised as patient-initiated delays whose degree of urgency is periodically reviewed.

2. Expected Waiting Lists: Expected patients arriving to the waiting list system in the future for urgent or non-urgent treatment. The characteristics defining this category of patients are similar to that of the current waiting list except that they are entirely based on estimations.

3. Mortality and Infection Rates: The probability of a patient's early departure from the waiting list due to COVID-19 infection or mortality.

Hospital-specific inputs gauge the availability and distribution of resources across different specialisations including:

1. Treatment Capacity: For every type of treatment, the duration of time required to treat a single patient.

2. Treatment Resources: For every type of treatment, the type of resources required and the availability of these resources across the week.

3. Prioritisation Policy: The policy based on which patients on the waiting list are prioritised and scheduled for treatment.

4.2. Pre-processing

4.2.1. Current waiting lists

The current waiting list dataset consists of a list of waiting patients seeking endoscopy treatment, each characterised by a set of patient specifications. The considered specifications include time of arrival, elective planned date, pathway, endoscopy treatment, and priority class. Serving as the actual current waiting list, this data is directly fed into the model.

This dataset is also used to draw out conclusions on the distribution of patient characteristics across the waiting list lanes considered. These distributions are used to extrapolate any discrepancies in the data as well as assign patient characteristics to the expected arrivals introduced in the next section.

4.2.2. Expected waiting lists

The expected waiting list arrivals are derived from CUH's historical endoscopy data. Assuming no further COVID-19 waves decapitating the capacity for elective treatment, waiting list levels should account for the natural annual growth of the waiting list as well as the significant disruption in arrivals amid the pandemic. That is, the expected number of arrivals post-pandemic (after COVID-19 restrictions subside on July 19th) are considered a combination of the following components [31]:

(a) Expected number of arrivals in the current and subsequent years assuming the pandemic has not occurred

(b) Missed arrivals during the pandemic or patients that would have sought out treatment had there been no pandemic

Pandemic-free Arrivals. To model (a), the upward trend characterising the number of waiting list arrivals between 2015 and 2019 (shown on the left in Fig. 2) is examined by means of regression modelling. As a result, the annual number of arrivals is extended to cover years 2020 to 2026.

To transform the projected annual arrivals to estimated daily arrivals, the distribution of arrivals within each previous year is studied using a piece-wise regression approach. The resulting projected daily arrivals are thereby transformed into proportions of the annual number of arrivals to model the distribution within the year rather than the actual magnitude of arrivals.

Exploiting the overall annual trend of arrivals as well as their proportional distribution within a specific year, the expected number of patients joining the waiting list for 2021–2026 was extrapolated. These values disregard the implications of COVID-19 to model the pandemic-free arrivals. The resulting patient arrivals are shown on the right in Fig. 2.

Missed Arrivals. Within the years 2020 and 2021, there has been a notable deterioration in the number of arrivals. In fact, the reduction in arrivals is attributed to the fact that patients have learned to 'live with their condition' to avoid hospital-onset infection until normality has been restored. This deterioration is extrapolated up until July 19th using a similar regression approach to mark the end of the COVID-19 pandemic. In fact, the overall upward trend in the restoration of 'normal' patient arrivals is quantified using Summer 2020 data (when restrictions were lifted).

Thereafter, the difference between the pandemic-free expected arrivals (a) and the actual arrivals collected from CUH for the duration of the pandemic serves as the missed arrivals (b). These arrivals are shifted towards the period following July 19th under a higher priority given the time waited so far.

Summing the two components, the resulting arrivals are shown in Fig. 3. The analysis is conducted across the highlighted post-pandemic surge in patient arrivals (July 2021–Jan. 2023, 562 days). As for the patient attributes characterising each arrival, the distributions derived from the current waiting list are used to randomly assign specifications to each patient.

4.2.3. Mortality and infection rates

There are a series of risks and implications resulting from prolonged waiting time, most importantly the increased potential of COVID-19 infection and mortality. The latter comprises infection-instigated and waiting time-instigated mortality. Given the absence of data on how these risks were amplified by the pandemic within CUH, the rates provided by Moreno et al. [6] were considered.
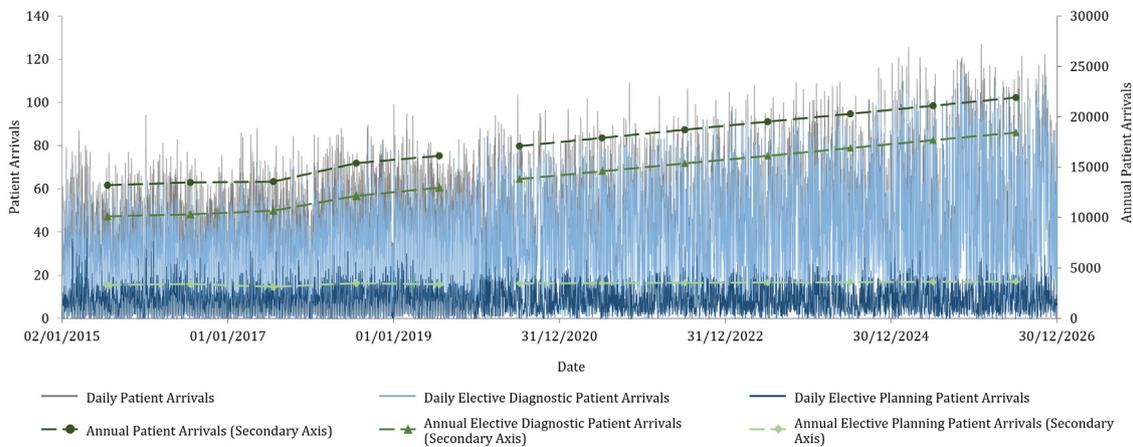
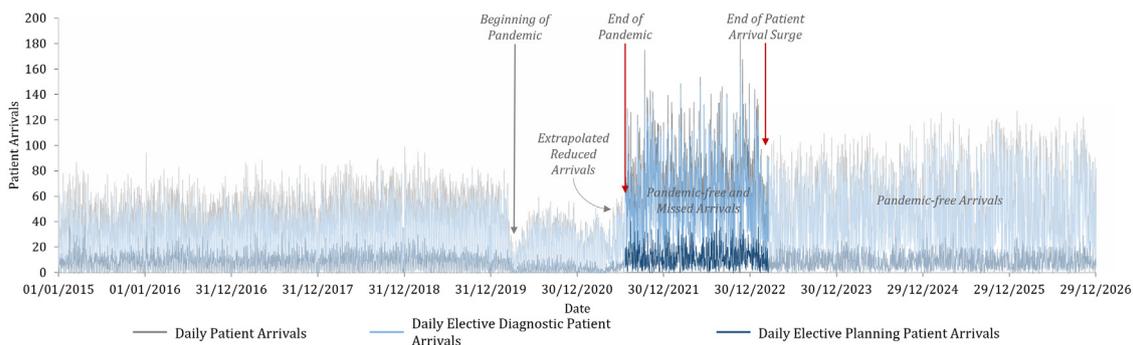**Fig. 2.** Expected patient arrivals disregarding pandemic implications.

Fig. 3. Patient arrivals combining upward trend and pandemic implications.

4.2.4. Treatment capacity

To model the treatment capacity of the endoscopy department, historical data on the number of procedures performed per week is collected. Thereafter, a distribution is fitted for each procedure category using @Risk 8.0. The software relies on bias-corrected Maximum Likelihood Estimation (MLE) for fitting. The fit of the distributions is tested via Akaike (AIC), Bayesian (BIC), and Average Log-Likelihood.

Apart from the micro-behaviour of capacity, the endoscopy treatment capacity is assumed to grow at a fixed annual rate. In fact, the number of gastroscopies, colonoscopies, and flexible sigmoidoscopies performed in the United Kingdom (UK) is characterised by a compounded annual growth rate (CAGR) of 1.20%, 5.03%, and 5.05% respectively considering 2016 to 2019 [1]. Brown et al. [8] attributed this annual growth to population size, population age profile, cancer incidence, new cancer referral guidelines, and the introduction of new treatment technologies among other yearly variations. For the purpose of this study, CUH is assumed to have followed a similar growth pattern from 2021 to 2026. The CAGRs are directly fed into the model as an annual inflation rate for treatment capacity.

4.2.5. Treatment resources

Resource availability has been a prominent and universal constraint for the re-introduction of elective treatment throughout the pandemic. While treatment capacity serves as the maximum number of patients that can be treatment, the available treatment resources determine which type of patients are treated and when. In fact, there are certain specialised resources who can only cater to a set of procedures and patients. Thus, the type of clinicians and theatres required for every endoscopy procedure are collected along with their corresponding schedule across a typical week.

To select a resource for a particular patient, the model investigates the availability of clinicians and theatres able to accommodate the procedure in-question. In case more than one complying clinician and/or theatre are available, the model assigns the more specialised resource (ability to perform fewer procedures). This allows the system to support a wider variety of incoming patients with a more comprehensive set of resources.

4.2.6. Prioritisation policy

The prioritisation policy implemented at CUH comprises scheduling patients in decreasing order of waited time within each priority class. To model such a system, the following equation is used to quantify the priority number of each patient i:

$$PriorityPoints_i = PriorityWeight_{PriorityClass_i} \\ + CurrentTime - PlannedDate_i \tag{1}$$

The priority weight serves as a predetermined constant allowing the transformation of the priority class into a time-equivalent value to be added to the delay in treatment as it stands now. The assignment of this value is further explored in Section 4.2.7. In practical terms, this constant indicates the number of additional days a patient would wait before they are deemed eligible for the next priority class. For instance, a patient belonging to P3 whose delay exceeds the gap between the priority weights of P3 and P2 would be scheduled before a newly arriving P2 patient.

4.2.7. Assumptions

While the actual endoscopy waiting list consists of multiple specialists, only two lanes were modelled given the severity of their backlog. Although the treatment resources directed to the remaining disregarded lanes were excluded, clinicians do hold some overlapping capabilities across multiple lanes. Thus, the
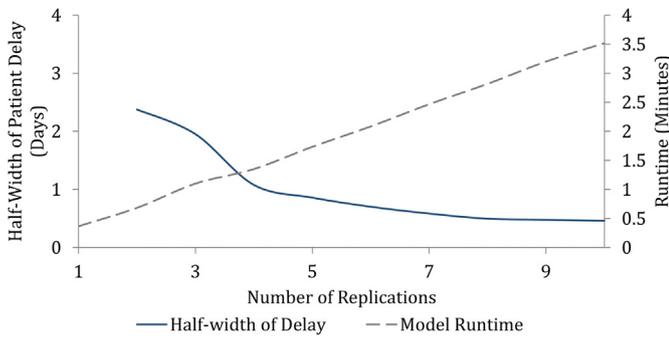
Fig. 4. Model runtime versus half-width of delay time for a different number of replications (sample of two months).

Fig. 5. Distribution of priority classes across current waiting list (sample size: 10,194 patients).

treatment capacity dedicated to elective planning and diagnostic patients was not completely isolated.

4.3. Experimentation

4.3.1. Model initialisation

To first ensure the model setup reflects the actual system, a warm-up period is considered. The warm-up period accounts for the time required from the date the simulation starts (May 12th) to arrive at the post-pandemic stage (July 19th) after which there is a surge in waiting list levels. This phase is excluded from the analysis, allowing sufficient time for the waiting list to attain its steady state — in this case, a continuous upward trend. As a result, only the post-pandemic delay is reported.

4.3.2. Model run length and estimation approach

The model is run for five years until the end of year 2026 to capture both short- and long-term waiting list outlooks. To cover the variation in the stochastic patient characteristic assignments and treatment capacity, replications are considered imperative. For the months of June and July 2021, up to 10 replications were considered. To gauge the minimum number of replications required to capture this variation, the half-width of the delay in treatment (95% confidence interval) and model runtime were monitored . After experimenting with both variables for a sample of June and July 2021 data, four replications were deemed sufficient to substantially narrow down this confidence interval to one day while curbing computation time as indicated in Fig. 4.

4.4. Implementation

The model was implemented on Arena Simulation Software 16.0 (professional license) on a Windows 64-bit operating system. Random sampling was conducted for a set of patients joining the waiting list within a time period of one week to test the validity of the model assignments.

4.4.1. Model execution and calibration

To calibrate the model in accordance with the selected case study, a parametric approach is derived from Wood [30]. That is, referring to Eq. (1), the priority weight corresponding to each priority class is deemed a variable that is iteratively altered until the model mimics the real system with respect to a set of KPIs that best represent the system: average delay in treatment, minimum delay in treatment, maximum delay in treatment, performance or the proportion of patients waiting longer than 18-weeks.

Calibration is conducted using CUH's 2019 waiting list given the availability of patient arrivals as well as the corresponding delay in treatment. The year 2019 was chosen as a reference given
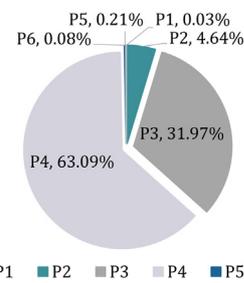
the relative stability in treatment capacity, similarly to the years 2021 to 2026.

To perform this calibration, Arena Process Analyser Software is employed given its ability to run multiple scenarios simultaneously. After considering over 70 different sets of priority weights, by trial and error, the set of priority weights rendering the model a true reflection of reality is deemed {154, 123, 101, 32, 15, 0} for P1 to P6 respectively. CUH's prioritisation policy seems to be heavily biased towards the first three priority classes. In fact, to achieve 2019's actual KPIs, P1 to P3 patients are scheduled early on followed by a significant gap in P4 to P6 treatment. Referring to Fig. 5, P3 and P4 clearly serve as the most common priority classes when it comes to endoscopy treatment. Thus, their priority weights hold significant consequences on the output. In practical terms, this gap indicates that a patient belonging to priority class P4 would be able to wait around 70 days until their condition is reclassified as P3. As for the remaining priority classes, the lack of sufficient data deems them irrelevant for this application and obstructs any conclusions.

While these selected priority weights minimise the gap separating the actual from the model-derived 2019 KPIs, a substantial error rate of 0.4% for the average delay and a 3% unit difference in performance was yielded. Possible sources of error include changing patient demographics, the turnover of clinicians with different capacities, and the introduction of a more systematic prioritisation policy within the model with P1 to P6 priority classes.

5. Discussion

Once the validity of the model as a true reflection of the actual system has been established, results are generated for a time horizon equating to the length of the post-pandemic surge in patient arrivals.

5.1. Short-term performance and delay

Serving as a primary KPI for a healthcare system's overall success, waiting time is monitored by assessing performance along with the overall delay. Currently, among the patients still waiting for treatment, 55% have been waiting for less than 18-weeks. After simulating the progress of the waiting list, given the prolonged surge in patient arrivals post-pandemic, performance fell to 21.3%, a long way from NHS's 92% standard. In just under 600 days, 21,494 patients were serviced with an average delay of 6 months (versus 3 months in 2019). In fact, the distribution of delays is indicated in Fig. 4. While most patients waited 4 to 7 months, a maximum delay of over 11 months was incurred.
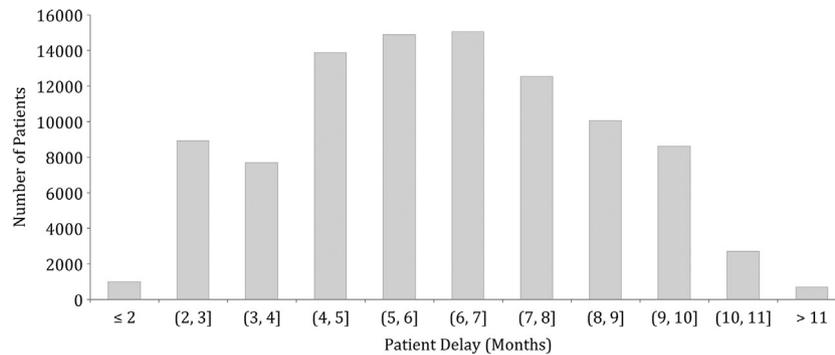
Fig. 6. Distribution of the delay in treatment across post-pandemic surge in patient arrivals.

5.2. Short-term patient-specific delay

After considering the macro performance of the system, the delay in treatment is studied at a micro-level considering the different patient characteristics. In fact, the average delay and corresponding 95% confidence intervals are indicated in Fig. 6 with respect to the waiting list lanes, pathways, priority classes, and endoscopy procedures respectively.

Waiting List Lanes. Fig. 7(a) indicates a significant 14.13% jump in the average delay of elective diagnostic patients with respect to elective planning. This gap can be attributed to the fact that the bulk of diagnostic patients are classified at a P4 priority while planned patients are relatively homogeneously distributed between P3 and P4 priority classes.

Pathways. Considering Fig. 7(b), the wide confidence interval surrounding the two-week wait and cancer pathways suggests the absence of sufficient patients to infer any robust conclusions. When it comes to routine and urgent patients, urgent cases seem to incur a substantially higher delay in treatment. While this may seem unconventional, most urgent cases fall under the elective diagnostic waiting list lane which is mainly characterised by the P4 priority class. Routine cases are heavily concentrated within the elective planning waiting list and hence, are fairly distributed across P3 and P4 priority classes.

Priority Classes. Fig. 7(c) validates the prioritisation policy applied for patient assignments. The relative jump in the delay between priority classes is dictated by the set priority weight. This is supported by the sharp shift between P3 and P4, mirroring the similar shift in priority weights discussed in Section 4.2.7.

Endoscopy Procedures. Fig. 7(d) demonstrates a homogeneous distribution of the delay across different endoscopy procedures. While Enteroscopy, Dilatation, PEG, TBBX, and Pouchoscopy correspond to a limited sample size, the remaining procedures seem to incur a delay of around 181 days. This consistency can be attributed to CUH's assignment of specialised treatment resources according to the relative number of patients requiring this speciality.

5.3. Long-term waiting list trajectory

The long-term waiting list trajectory is studied for the next five years, until the end of 2026. Referring to Fig. 8, the average number of patients waiting is indicated per day along with a shaded margin of error. The waiting list seems to maintain growth across the studied timeframe up to a peak of 37,400 patients. The initial sharp growth pertains to the post-pandemic arrival surge while the steady growth thereafter is attributed to the natural long-term inflation in patient arrivals. Meanwhile, the treatment capacity is tailored to the current reduced demand

as indicated by the slight decrease in waiting list levels at the beginning of the simulation. While the treatment capacity is inflated annually as per Section 4.2.4, this growth was insufficient to completely stabilise the waiting list in the long term. This can be attributed to the pandemic acting as a massive disruption and obstructing any growth in treatment capacity in 2020 and 2021 while patient arrivals continued to grow. The impact of this disruption is indicated in Fig. 8. As a result, capacity is lagging demand. In fact, this has been the case well before the pandemic. According to a report published by the Health Foundation [36], meeting the 92% performance standard pre-pandemic would have required the treatment of 500,000 additional patients per year until 2024. Thus, without a substantial investment in capacity, it was highly unlikely to attain the 18-week standard by 2024 with the existing infrastructure and staffing levels. Considering the pandemic has only exacerbated this vulnerability, it now seems completely unrealistic that the waiting list levels will recover by 2026 as seen in Fig. 8.

5.4. Sensitivity analysis

This section closely considers investing in expanding treatment capacities. Reiterating the model's significance as introduced in Section 1, the tool gauges the payback of investing in treatment resources by quantifying the resulting reduction in patient delays and performance. Given the homogeneous distribution of the delay across the different endoscopy procedures as indicated in Section 5.2, an overall inflation rate for the capacity is considered for all endoscopy treatments. This assessment was performed for capacity inflation rates of 20%, 40% and 60%. To that end, the implications of the resource investments are shown in Fig. 9.

Comparing inflation rates, the 40% increase in capacity seems to offer greater improvements in delay KPIs. This improvement exhibits the exponential behaviour characterising waiting list levels versus capacity. That is, as more patients miss treatment early on, the more likely incoming patients will have their treatment delayed. Certainly, there would be a limit to the growth in benefits after which the delay KPIs would plateau, as demonstrated by the 160% capacity. In fact, to attain the NHS's 92% standard for performance, approximately, a 55% treatment capacity investment is required across the post-pandemic surge in patient arrivals.

Fig. 10 presents the long-term waiting list trajectory for the said capacity levels along with their respective shaded margins of error. While a 60% capacity investment would reverse any pandemic implications on the waiting list, a 20% investment would stabilise and level up waiting list levels by the first quarter of 2023.
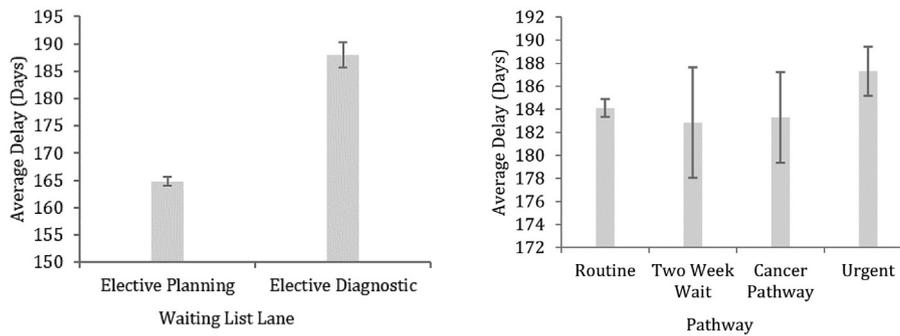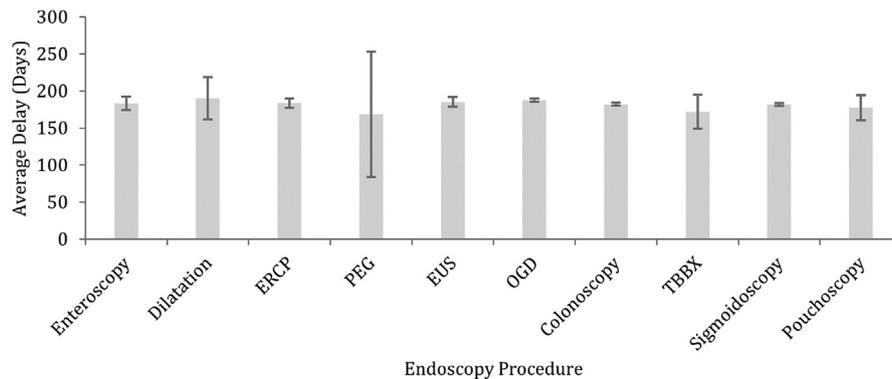
(a) Average delay across waiting list lanes

(b) Average delay across pathways
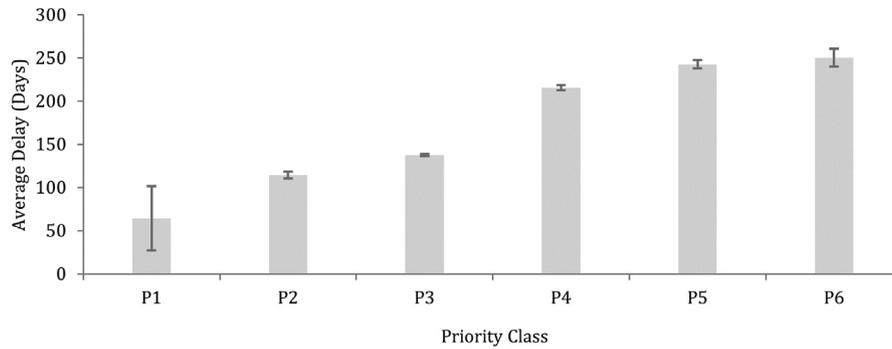
(c) Average delay across priority classes

(d) Average delay across endoscopy procedures

Fig. 7. Short-term patient-specific delay.

6. Constraints and limitations

The model's lack of operational practicality serves as the primary drawback of adopting a simulation methodology. In fact, the model is ill-suited for daily use by the hospital. The tool should be maintained regularly to ensure that the model is populated with updated data reflecting the system. Hence, the model setup as well as its maintenance is deemed highly expensive.

The generalisability of the findings can also be considered limiting. While the prioritisation policy implemented is universal across the UK healthcare system, other aspects of the model, particularly around treatment capacity, are specific to the application.

7. Conclusion

The undertaken research presents a clear indication of the severity of the imminent waiting list backlog as a rising threat to the well-being of patients across all medical fields. Not only is the current backlog presenting as a major burden based upon current treatment capacities, but the upcoming expected surge in patient arrivals serves as an exceptional circumstance. The proposed model presents a pandemic recovery tool that offers visibility on the short- and long-term trajectory of the backlog.

After estimating the expected patient arrivals and treatment capacity for CUH's endoscopy department, the resulting delay was approximated at 6 months, with 78% of patients' delay exceeding the 18-week NHS standard. In the long-term, the waiting list is expected to maintain its upward trend in the next five years unless resource investments are initiated. Testing the potential expansion in resources yielded substantial improvements. In fact, a 55% capacity investment over the post-pandemic surge in waiting list levels was deemed necessary to maintain the 92% NHS standard for performance, validating the importance of considering recovery investments.
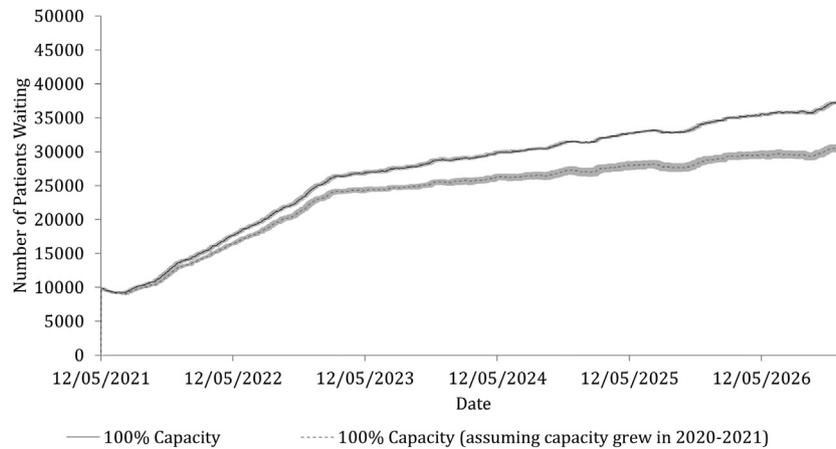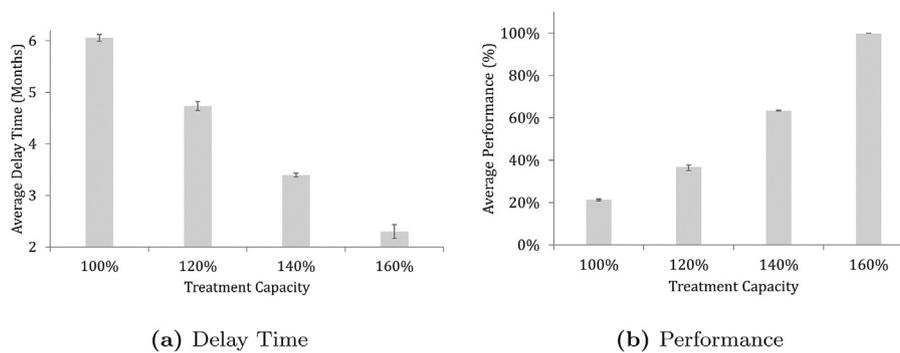
Fig. 8. Long-term waiting list trajectory (2021–2026).

(a) Delay Time **(b)** Performance

Fig. 9. Average delay and performance considering resource investments of 20%, 40%, and 60%.
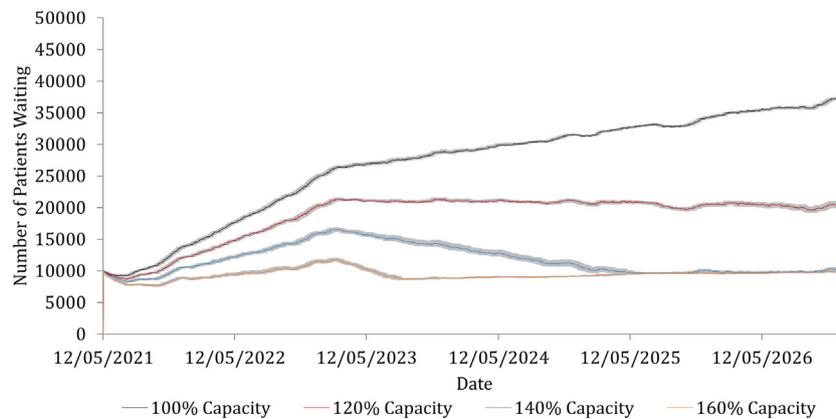

Fig. 10. Waiting list long-term trajectory considering resource investments of 20%, 40%, and 60%.

While the conclusions disclosed in this study should serve as guidance for pandemic recovery planning rather than certain predictions, the modelling process is generalisable beyond endoscopy. Similar to other elective fields, resources such as theatres and clinicians are shared between different patient types across a wide variety of procedures. Other factors not included in this model that may differ across other elective fields include multiple-stage treatment procedures and emergency cases.

Potential future prospects for this study primarily include: expanding the scope to include all endoscopy waiting list lanes and specialities to accurately gauge the available capacity for treatment; quantify the costs associated with expanding treatment capacity; testing different prioritisation policies from the literature to validate/refute the optimality of the current categorisation system in recovering waiting list levels as efficiently as possible.

CRediT authorship contribution statement

Romy Nehme: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing – original draft, Visualization, Project administration. **Alena Puchkova:** Supervision, Writing – review & editing. **Ajith Parlikad:** Conceptualization, Writing – review & editing.

Declaration of competing interest

References

[1] NHS England, NHS Improvement, Monthly diagnostic waiting times and activity, 2021.

[2] D.A. Martinez, H. Zhang, M. Bastias, F. Feijoo, J. Hinson, R. Martinez, J. Dunstan, S. Levin, D. Prieto, Prolonged wait time is associated with increased mortality for Chilean waiting list patients with non-prioritized conditions, BMC Public Health 19 (1) (2019) 1–11, http://dx.doi.org/10.1186/s12889-019-6526-6.

[3] K. Matsuo, H. Novatt, S. Matsuzaki, M.S. Hom, A.V. Castaneda, E. Licon, D.J. Nusbaum, L.D. Roman, Wait-time for hysterectomy and survival of women with early-stage cervical cancer: A clinical implication during the coronavirus pandemic, Gynecol. Oncol. 158 (1) (2020) 37–43, http://dx.doi.org/10.1016/j.ygyno.2020.05.019.

[4] C. Maringe, J. Spicer, M. Morris, A. Purushotham, E. Nolte, R. Sullivan, B. Rachet, A. Aggarwal, The impact of the COVID-19 pandemic on cancer deaths due to delays in diagnosis in England, UK: A national, population-based, modelling study, Lancet Oncol. 21 (8) (2020) 1023–1034, http://dx.doi.org/10.1016/S1470-2045(20)30388-0.

[5] A. Sud, B. Torr, M.E. Jones, J. Broggio, S. Scott, C. Loveday, A. Garrett, F. Gronthoud, D.L. Nicol, S. Jhanji, et al., Effect of delays in the 2-week-wait cancer referral pathway during the COVID-19 pandemic on cancer survival in the UK: A modelling study, Lancet Oncol. 21 (8) (2020) 1035–1044, http://dx.doi.org/10.1016/S1470-2045(20)30392-2.

[6] R. Moreno, J.-L. Díez, J.-A. Diarte, F. Macaya, J.-M. de la Torrre Hernández, O. Rodríguez-Leor, R. Trillo, J. Alonso-Briales, I. Amat-Santos, R. Romaguera, et al., Consequences of canceling elective invasive cardiac procedures during Covid-19 outbreak, Catheter. Cardiovasc. Interv. 97 (5) (2020) 927–937, http://dx.doi.org/10.1002/ccd.29433.

[7] K.P. Iyengar, V.K. Jain, A. Vaish, R. Vaishya, L. Maini, H. Lal, Post COVID-19: Planning strategies to resume orthopaedic surgery — Challenges and considerations, J. Clinic. Orthopaed. Trauma 11 (2020) 291–295, http://dx.doi.org/10.1016/j.jcot.2020.04.028.

[8] H. Brown, S. Wyatt, N. Gale, S. Croft, A. Turner, A. Mulla, Scoping the future: An evaluation of endoscopy capacity across the NHS in England, Cancer Research UK (September) (2015).

[9] J. Greenhalgh, S. Dalkin, K. Gooding, E. Gibbons, J. Wright, D. Meads, N. Black, J.M. Valderas, R. Pawson, Functionality and feedback: A realist synthesis of the collation, interpretation and utilisation of patient-reported outcome measures data to improve patient care, Health Serv. Deliv. Res. 5 (2) (2017) 1–280, http://dx.doi.org/10.3310/hsdr05020.

[10] M. Guerriere, Determining the utility of public reporting — Too early to judge, HealthcarePapers 6 (2) (2005) 62–67, http://dx.doi.org/10.12927/hcpap..17764.

[11] D. Daldoul, I. Nouaouri, H. Bouchriha, H. Allaoui, A stochastic model to minimize patient waiting time in an emergency department, Oper. Res. Health Care 18 (2018) 16–25, http://dx.doi.org/10.1016/j.orhc.2018.01.008.

[12] N. Nezamoddini, M.T. Khasawneh, Modeling and optimization of resources in multi-emergency department settings with patient transfer, Oper. Res. Health Care 10 (2016) 23–34, http://dx.doi.org/10.1016/j.orhc.2016.06.001.

[13] C. Oh, A.M. Novotny, P.L. Carter, R.K. Ready, D.D. Campbell, M.C. Leckie, Use of a simulation-based decision support tool to improve emergency department throughput, Oper. Res. Health Care 9 (2016) 29–39, http://dx.doi.org/10.1016/j.orhc.2016.03.002.

[14] P. Bhattacharjee, P.K. Ray, Simulation modelling and analysis of appointment system performance for multiple classes of patients in a hospital: A case study, Oper. Res. Health Care 8 (2016) 71–84, http://dx.doi.org/10.1016/j.orhc.2015.07.005.

[15] H. Mahmoudzadeh, A.M. Shalamzari, H. Abouee-Mehrizi, Robust multi-class multi-period patient scheduling with wait time targets, Oper. Res. Health Care 25 (2020) 100254, http://dx.doi.org/10.1016/j.orhc.2020.100254.

[16] Department of Health, Great Britain, The NHS Improvement Plan, 2004.

[17] C. Kaul, A. Kaul, S. Verma, Comparitive study on healthcare prediction systems using big data, in: 2015 International Conference on Innovations in Information, Embedded and Communication Systems, ICIIECS, IEEE, 2015, pp. 1–7, http://dx.doi.org/10.1109/ICIIECS.2015.7193095.

[18] C. Curtis, C. Liu, T.J. Bollerman, O.S. Pianykh, Machine learning for predicting patient wait times and appointment delays, J. Amer. College Radiol. 15 (9) (2018) 1310–1316, http://dx.doi.org/10.1016/j.jacr.2017.08.021.

[19] Y. Sun, K.L. Teow, B.H. Heng, C.K. Ooi, S.Y. Tay, Real-time prediction of waiting time in the emergency department, using quantile regression, Annal. Emerg. Med. 60 (3) (2012) 299–308, http://dx.doi.org/10.1016/j.annemergmed.2012.03.011.

[20] A. Joseph, T. Hijal, J. Kildea, L. Hendren, D. Herrera, Predicting waiting times in radiation oncology using machine learning, in: 2017 16th IEEE International Conference on Machine Learning and Applications, ICMLA, IEEE, 2017, pp. 1024–1029, http://dx.doi.org/10.1109/ICMLA.2017.00-16.

[21] F. Gonçalves, R. Pereira, J. Ferreira, J.B. Vasconcelos, F. Melo, I. Velez, Predictive analysis in healthcare: Emergency wait time prediction, in: 9th International Symposium on Ambient Intelligence, Springer, 2018, pp. 138–145, http://dx.doi.org/10.1007/978-3-030-01746-0_16.

[22] Z. Liu, D. Rexachs, E. Luque, F. Epelde, E. Cabrera, Simulating the micro-level behavior of emergency department for macro-level features prediction, in: 2015 Winter Simulation Conference, WSC, IEEE, 2015, pp. 171–182, http://dx.doi.org/10.1109/WSC.2015.7408162.

[23] V. Babashov, I. Aivas, M. Begen, J. Cao, G. Rodrigues, D. D'Souza, M. Lock, G. Zaric, Reducing patient waiting times for radiation therapy and improving the treatment planning process: A discrete-event simulation model (Radiation Treatment Planning), Clin. Oncol. 29 (6) (2017) 385–391, http://dx.doi.org/10.1016/j.clon.2017.01.039.

[24] M. Chong, M. Wang, X. Lai, B. Zee, F. Hong, E. Yeoh, E. Wong, C. Yam, P. Chau, K. Tsoi, et al., Patient flow evaluation with system dynamic model in an emergency department: Data analytics on daily hospital records, in: 2015 IEEE International Congress on Big Data, IEEE, 2015, pp. 320–323, http://dx.doi.org/10.1109/BigDataCongress.2015.54.

[25] N. Macdonald, C. Clements, A. Sobti, D. Rossiter, A. Unnithan, N. Bosanquet, Tackling the elective case backlog generated by Covid-19: The scale of the problem and solutions, J. Public Health 42 (4) (2020) 712–716, http://dx.doi.org/10.1093/pubmed/fdaa155.

[26] A. Fowler, T. Dobbs, Y. Wan, R. Laloo, S. Hui, D. Nepogodiev, A. Bhangu, I. Whitaker, R. Pearse, T. Abbott, Resource requirements for reintroducing elective surgery during the COVID-19 pandemic: Modelling study, Br. J. Surg. 108 (1) (2021) 97–103, http://dx.doi.org/10.1093/bjs/znaa012.

[27] E. García-Rojo, C. Manfredi, R. Santos-Pérez-de-la Blanca, Á. Tejido-Sánchez, B. García-Gómez, M. Aliaga-Benítez, J. Romero-Otero, A. Rodriguez-Antolín, Impact of COVID-19 outbreak on urology surgical waiting lists and waiting lists prioritization strategies in the post-COVID-19 era, Actas. Urol. Esp. 45 (3) (2021) 207–214 (english edition), http://dx.doi.org/10.1016/j.acuroe.2021.02.005.

[28] S. Oussedik, S. MacIntyre, J. Gray, P. McMeekin, N.D. Clement, D.J. Deehan, Elective orthopaedic cancellations due to the COVID-19 pandemic: Where are we now, and where are we heading? Bone Joint Open 2 (2) (2021) 103–110, http://dx.doi.org/10.1302/2633-1462.22.bjo-2020-0161.r1.

[29] D. Joshi, A. Jalali, T. Whipple, M. Rehman, L.M. Ahumada, "P3": An adaptive modeling tool for post-COVID-19 restart of surgical services, JAMIA Open 4 (2) (2021) ooab016, http://dx.doi.org/10.1093/jamiaopen/ooab016.

[30] R.M. Wood, Modelling the impact of COVID-19 on elective waiting times, J. Simul. (2020) 1–9, http://dx.doi.org/10.1080/17477778.2020.1764876.

[31] K.M.A. Ho, A. Banerjee, M. Lawler, M.D. Rutter, L.B. Lovat, Predicting endoscopic activity recovery in England after COVID-19: A national analysis, Lancet Gastroenterol. Hepatol. 6 (5) (2021) 381–390, http://dx.doi.org/10.1016/S2468-1253(21)00058-3.

[32] T. Comans, A. Chang, L. Standfield, D. Knowles, S. O'Leary, M. Raymer, The development and practical application of a simulation model to inform musculoskeletal service delivery in an Australian public health service, Oper. Res. Health Care 15 (2017) 13–18, http://dx.doi.org/10.1016/j.orhc.2017.07.002.

[33] A. Salmon, S. Rachuba, S. Briscoe, M. Pitt, A structured literature review of simulation modelling applied to emergency departments: Current patterns and emerging trends, Oper. Res. Health Care 19 (2018) 1–13, http://dx.doi.org/10.1016/j.orhc.2018.01.001.

[34] T. Monks, C.S.M. Currie, B.S. Onggo, S. Robinson, M. Kunc, S.J.E. Taylor, Strengthening the reporting of empirical simulation studies: Introducing the STRESS guidelines, J. Simul. 13 (1) (2019) 55–67, http://dx.doi.org/10.1080/17477778.2018.1442155.

[35] G. Longcroft-Wheaton, N. Tolfree, A. Gangi, R. Beable, P. Bhandari, Data from a large western centre exploring the impact of COVID-19 pandemic on endoscopy services and cancer diagnosis, Frontline Gastroenterol. 12 (3) (2021) 193–199, http://dx.doi.org/10.1136/flgastro-2020-101543.

[36] T. Gardner, C. Fraser, S. Peytrignet, Elective care in England: Assessing the impact of COVID-19 and where next, Health Found. (2020).