

## Predictive and prescriptive analytics for multi-site modelling of frail and elderly patient services

Elizabeth Williams, Daniel Gartner & Paul Harper

To cite this article: Elizabeth Williams, Daniel Gartner & Paul Harper (21 Apr 2025): Predictive and prescriptive analytics for multi-site modelling of frail and elderly patient services, Journal of the Operational Research Society, DOI: [10.1080/01605682.2025.2487650](https://doi.org/10.1080/01605682.2025.2487650)

To link to this article: <https://doi.org/10.1080/01605682.2025.2487650>



© 2025 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group



[View supplementary material](#)



Published online: 21 Apr 2025.



[Submit your article to this journal](#)



Article views: 198



[View related articles](#)



[View Crossmark data](#)

# Predictive and prescriptive analytics for multi-site modelling of frail and elderly patient services

Elizabeth Williams , Daniel Gartner  and Paul Harper 

School of Mathematics, Cardiff University, Cardiff, United Kingdom

## ABSTRACT

Many economies are challenged by the effects of an ageing population, particularly in sectors where resource capacity planning is critical, such as healthcare. This research addresses the operational challenges of bed and staffing capacity planning in hospital wards by using predictive and prescriptive analytical methods, both individually and in tandem. We applied these methodologies to a study of 165,000 patients across a network of 11 hospitals in the UK. Predictive modelling, specifically Classification and Regression Trees, forecasts patient length of stay based on clinical and demographic data. On the prescriptive side, deterministic and two-stage stochastic optimisation models determine optimal bed and staff planning strategies to minimise costs. Linking the predictive models with the prescriptive optimisation models, generates demand forecasts that inform the optimisation process, providing accurate and practical solutions. The results demonstrate that this integrated approach captures real-world variations in patient LOS and offers a 7% cost saving compared to average-based planning. This approach helps healthcare managers make robust decisions by incorporating patient-specific characteristics, improving capacity allocation, and mitigating risks associated with demand variability. Consequently, this combined methodology can be broadly extended across various sectors facing similar challenges, showcasing the versatility and effectiveness of integrating predictive and prescriptive analytics.

## ARTICLE HISTORY

Received 6 August 2024  
Accepted 27 March 2025

## KEYWORDS

OR in health services;  
healthcare management;  
mathematical programming;  
stochastic modelling; data  
mining

## 1. Introduction

Ageing populations put pressure on both the economy and healthcare resources worldwide with medical care expenditures on the rise (United Nations, 2019). Ageing is one of the most common and well-known risk factors for most chronic diseases (Macnee et al., 2014). According to the United Nations (2019), an elderly person can be defined as 65 years old and over, whilst a frail person is classified as one who is at high risk of falling into dependency as a result of a negative event, such as an injury, fall or disability (Xue, 2011). Care for frail and elderly patients has more challenges and barriers in providing care due to a lack of resources or specialised models for care delivery (Heydari et al., 2019). Frail and elderly patients often suffer from multi-morbidity and can take longer to recover in a hospital with more staffing hours and resources required. These patients not only share resources with other patient types but are unique in terms of having their own service specialty within hospitals, known as care of the elderly (COTE). There is often difficulty in grouping these patients together for

length of stay (LOS) prediction, as there are many different factors which cause longer lengths of stay.

To address the challenges to healthcare services caused by ageing populations and the medical complexities of caring for frail and elderly patients, healthcare analytics has proven beneficial. While analytical techniques have been reported to yield up to 25% reductions in annual healthcare costs (Dash et al., 2019), our study focuses on a specific operational domain-hospital bed and staff capacity planning. Within this context, our integrated predictive and prescriptive models achieve a 7% cost reduction compared to traditional average-based approaches.

The recent COVID-19 pandemic has demonstrated the crucial need for analytics to be used and continually developed within the healthcare setting to provide highly beneficial results (Charles et al., 2024). The main challenges faced by healthcare managers are the hundreds of beds and staff to manage, which creates an excessive number of options for decisions (Best et al., 2015; Proudlove et al., 2007). Demand and capacity are further complicated because no two patients are exactly alike. Because bed capacity planning typically relies on averages, it fails to take into consideration the stochastic

**CONTACT** Elizabeth Williams  [Williamsem20@cardiff.ac.uk](mailto:Williamsem20@cardiff.ac.uk)  School of Mathematics, Cardiff University, Cardiff, United Kingdom

 Supplemental data for this article can be accessed online at <https://doi.org/10.1080/01605682.2025.2487650>.

© 2025 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

nature of the healthcare industry (Abdelaziz & Masmoudi, 2012; Harper & Shahani, 2002). This research contributes to the literature by incorporating natural variability through the use of classification and regression trees (CART).

This research was motivated by collaboration with a large healthcare network that manages frail and elderly patients across multiple hospitals. Healthcare administrators highlighted significant challenges in bed and staffing capacity planning, where reliance on average-based predictions led to inefficiencies, resource shortages and increased costs. The aim of this study was to develop integrated predictive and prescriptive analytics models that address these operational issues by incorporating patient-specific variability, thereby enhancing planning accuracy and reducing costs.

To the best of our knowledge, this is the first study to examine the value of linking predictive and prescriptive analytics for the capacity planning of frail and elderly healthcare resources. We approach the problem as follows:

1. We begin by utilising CART methodologies with a large dataset consisting of over 165,000 patient records. This analysis focuses on identifying groupings of patients aged 65 and over who share similar clinical and demographical attributes that affect the LOS in hospital. CART was chosen for this analysis because of its ability to handle large datasets effectively and produce easily interpretable visual representations of decision rules. These groupings help us understand the factors that influence LOS and form the basis for our predictive modelling.
2. Next, we develop both deterministic and stochastic mixed-integer programming models to plan bed and staffing resources. The deterministic model assumes that future demand is known and fixed, while the stochastic model accounts for uncertainty in patient admissions and LOS by incorporating multiple scenarios into the planning process. These models aim to optimise resource allocation to minimise costs while meeting patient care requirements.
3. We then integrate the CART-based groupings into our optimisation models. By using the demand forecasts generated from the predictive models, we inform the deterministic and two-stage stochastic optimisation models. This step ensures that the variations and uncertainties captured by the predictive analysis are directly utilised in the resource planning process, leading to more accurate and practical solutions.
4. Finally, we analyse the value of the stochastic solution (VSS), to assess the benefits of implementing the stochastic model over the traditional deterministic

model. VSS measures the improvement in decision-making that results from considering uncertainty in the optimisation process. By comparing the outcomes of both models, we demonstrate the advantages of incorporating stochastic elements into capacity planning, such as better handling of demand variability and improved resource allocation.

The goal of this research is to develop and evaluate decision models that enhance capacity planning for frail and elderly patients in hospital settings. The CART models aim to forecast patient LOS based on clinical and demographic data, facilitating more accurate demand prediction and grouping patients by shared characteristics that influence LOS. These predictions serve as inputs to the optimisation models, which determine optimal bed and staffing levels under both deterministic and stochastic conditions. The integration of these predictive and prescriptive models enables healthcare managers to make informed decisions that minimise costs, accommodate demand variability and ensure the delivery of high-quality patient care. By understanding the value of predictive and prescriptive analytics, healthcare managers can avoid traps by planning resources based on averages. This, in turn, can lead to better health outcomes and more efficient use of resources.

The remainder of the paper is structured as follows. [Section 2](#) provides an overview of current literature in terms of CART analysis with a particular focus on frail and elderly patients. Moreover, deterministic and stochastic optimisation models for hospital bed planning are compared and contrasted with our work. In [Section 3](#), we first introduce CART, followed by the introduction of our deterministic and two-stage stochastic models. Then, we introduce an illustrative example to demonstrate the models' functionality. In [Section 4](#), we apply our models to real-world data from a network of hospitals, introducing variation into the model by utilising CART predictions for demands. Then, we provide comparisons between the deterministic and two-stage stochastic models by calculating the VSS. We offer a detailed discussion in [Section 5](#), providing managerial insights, discussing the generalisability of our results and outlining directions for future research. Our paper concludes with [Section 6](#).

## 2. Related work

This section aims to provide insight into existing operational research (OR) literature specifically focusing on hierarchical CART methodologies along with deterministic and stochastic modelling techniques. The first subsection will focus on CART analysis applied to frail and elderly patients and the second subsection will focus on deterministic and

stochastic modelling techniques applied to healthcare settings.

Ensuring healthcare facilities possess adequate resources like staff, equipment, and beds to meet patient demand efficiently is essential for effective healthcare operations management (Ardakani et al., 2023; Costa et al., 2003; Harper, 2002). Within operational research, capacity planning employs predictive modelling and optimisation techniques to anticipate patient flow, regulate hospital bed occupancy, and allocate resources effectively. Previous literature reviews have delved into these methodologies, focusing on their application in optimising healthcare resource utilisation (Chanchaichujit et al., 2019; Humphreys et al., 2022). Various mathematical models, including queueing theory and simulation, are utilised to analyse patient arrival patterns, LOS and resource utilisation, enabling healthcare providers to make informed decisions regarding staffing levels, facility expansions, and resource allocation. Capacity planning plays a crucial role in optimising healthcare delivery, minimising wait times, and enhancing patient satisfaction while ensuring cost-effectiveness and resource utilisation (Bhattacharjee & Ray, 2014; Erhard et al., 2018; Grange et al., 2024; Liping et al., 2018; Nasrabadi et al., 2020).

A wide variety of literature reviews have been published within the healthcare OR field. Typically, literature reviews are more focused on specific methods (Fildes et al., 2008), specific problems (Marynissen & Demeulemeester, 2019), or specific patient types (Williams et al., 2021). Fildes et al. (2008) explored different forecasting techniques in the OR domain whilst, Marynissen and Demeulemeester (2019) focused on different methods but applied to the context of appointment scheduling within hospitals. Williams et al. (2021) presented a literature review highlighting the use of Operational Research and Management Science (OR/MS) methodologies in addressing the challenges associated with the care of frail and elderly patients. These reviews demonstrate a wide variety of research taking place within the healthcare domain. Our research aims to overlap between these reviews, by applying OR methodologies to elderly and frail patient care.

### **2.1. CART analysis applied to frail and elderly**

CART is a machine learning method for constructing prediction models from data. The algorithm constructs a decision tree which is structured hierarchically. The decision tree asks a series of questions that decide groups into which the data is sorted, in the form of binary recursive partitioning, where each node is split into two groups. CART analysis has often been applied to healthcare for

predictive analytics, covering a wide range of medical settings. Within frail and elderly care, the literature can be grouped into two subgroups. The first grouping focuses on illnesses typically suffered by these patients, with the second grouping having a more generic medical setting however using elderly and frail as the patient groups of interest.

#### **2.1.1. Use of CART within hospitals**

The works of Byeon (2015), Ius et al. (2020) and Watanabe et al. (2018) demonstrated the use of CART models within hospitals, focusing on age as a predictive factor or limiting the analysis to older populations. Byeon (2015) developed a CART-based tool to predict endocrine disorders in the elderly, identifying higher prevalence categories for depression and obesity. Ius et al. (2020) applied random forests to analyse treatment outcomes, using age as a continuous variable to show it influenced survival, though not as the strongest factor. Watanabe et al. (2018) applied CART to determine different risk factors of rotator cuff tears between elderly and young patients, find age to be the most significant predictor. The next category of splitting is then different depending on the age group of the patient. These studies highlight CART's effectiveness across diverse age-related health conditions.

#### **2.1.2. Use of CART across hospitals and community care**

Expanding to hospital and community-wide care, Kuo et al. (2019), Lam et al. (2019) and Passmore et al. (2011) demonstrate the broader application of CART models. Kuo et al. (2019) used CART models to develop a system for identifying social frailty in the elderly, incorporating 15 variables (e.g., age, BMI, income). Random forests and C5.0 classification models were found to have the highest prediction accuracy of 0.97. Lam et al. (2019) used CART to identify whether frailty would be an indicator of recurrent fallers over the age of 65 in the community. The results showed that certain frailty indexes did have a high predictive ability of recurrent falls, although were only significant in female patients. Passmore et al. (2011) identified that a sickness impact profile (SIP) score was the most influential factor in predicting unplanned hospital admissions among elderly patients. Additionally, the number of medications prescribed was found to influence hospitalisation, suggesting opportunities for preventative measures.

These six papers demonstrate the success of applying CART to frail and elderly patients when considering both the care setting and the medical condition suffered. This research aims to build upon the previous literature in this field by considering how different hospital locations determine a

patient's LOS. Additionally, a further variety of data types which may have an impact on patient stays within hospitals, such as radiological data, will be incorporated.

## **2.2. Multi-site deterministic and stochastic healthcare modelling**

Deterministic modelling is popular within healthcare settings due to its ease of application, however, traditionally many healthcare services are stochastic in nature (Mandelbaum et al., 2020).

### **2.2.1. Deterministic models**

Hare et al. (2009) and Segall (1992) analysed deterministic modelling within healthcare settings. Hare et al. (2009) developed a deterministic multi-state Markov model to plan for services within home and community care, incorporating five age groups, three of which represent the elderly. The authors incorporated the changing age and health demographics of the patient groups to determine how service use will change over time. Segall (1992) extended a disaggregated resource allocation model to include a demand constraint, to determine spatial allocation of resources. Applying to 16 acute care hospitals and then to the entire state of Massachusetts, the author was able to determine occupancy rates and demands for hospitals, and specialties and predict the average LOS.

### **2.2.2. Stochastic models**

The works of Abdelaziz and Masmoudi (2012), Guo et al. (2021), Levis and Papageorgiou (2004), Shehadeh and Padman (2022) and Thompson et al. (2009) applied stochastic modelling techniques to healthcare settings. Abdelaziz and Masmoudi (2012) used a multi-objective stochastic program to determine optimal bed and staffing allocations in 157 public hospitals. The authors considered random demand in order to minimise costs across specialty levels. Guo et al. (2021) applied logic-based benders decomposition and binary decision diagram-based approaches to optimise surgery scheduling, incorporating uncertainties such as cancellations and procedure durations. These methods successfully generated more robust schedules, reducing cancellations and improving operating room utilisation. Levis and Papageorgiou (2004) applied stochastic capacity planning to the pharmaceutical industry, addressing multi-site, multi-period problems with uncertain clinical trial outcomes and customer demand. The authors provided five examples of planning problems which are solved with uncertain clinical trial outcomes and customer demand. Shehadeh and Padman (2022) used stochastic

optimisation for elective surgery scheduling, emphasising that ignoring uncertainties in surgery durations and postoperative recover times results in suboptimal schedules and higher costs. The paper discusses the complexity of modelling these uncertainties due to limited data, and the conflicting objectives of various stakeholders. Thompson et al. (2009) used Markov decision processes to plan short-term allocation of patients during demand surges. The authors aimed to determine the best patient assignments and reduce the cost of transferring patients.

### **2.2.3. Deterministic and stochastic models**

Mestre et al. (2015) used location-allocation models with both deterministic and stochastic approaches for the planning and designing of a network of hospitals. The authors aim to inform decision-makers on how to improve access to different healthcare services and specialties whilst minimising costs. The authors were able to successfully conclude that by including both location and allocation within the first stage, the model was more flexible in terms of hospital network planning, allowing the second stage to incorporate unsatisfied demand and extra capacities. Restrepo et al. (2020) presents a two-stage stochastic model for integrated staffing and scheduling in home healthcare, demonstrating that it reduces under-covering and over-covering costs compared to a deterministic model using average demand. The study demonstrates that accounting for demand variability through stochastic modelling leads to more robust and cost-effective staffing and scheduling decisions in home healthcare services. Dehghani et al. (2021) proposed both deterministic and stochastic programming models for proactive transshipment in blood supply chains to balance the wastage and shortage of blood units. The deterministic model provided a baseline by assuming known and fixed demand, while the stochastic model accounted for demand uncertainty across a network of hospitals. Using a two-stage stochastic approach with Quasi-Monte Carlo sampling, the study optimised blood orders and hospital transfers, demonstrating significant cost reductions, improved inventory management, and potential savings compared to current and no-transshipment policies. Galli et al. (2021) used prescriptive analytics for healthcare inventory management, proposing a novel method that combines machine learning with stochastic optimisation to optimise drug replenishment. The study demonstrates the effectiveness of this technique in providing robust and cost-efficient inventory solutions tailored to the dynamic and uncertain nature of healthcare environments.

These papers display the variety of both deterministic and stochastic modelling applications within healthcare. These models will be built upon to determine where specific specialties should be located based on demand locations when focusing on a subgroup of patients. The benefits will be determined by using either deterministic or stochastic modelling.

### 2.3. Literature summary

In conclusion, our search of the literature has revealed gaps in linking predictive and prescriptive analytics to optimise resource allocation under uncertainty in healthcare. While existing studies often use deterministic models, our approach integrates data-driven predictive modelling with stochastic optimisation. We employ CART to forecast demand in a two-stage stochastic program, offering a robust optimisation model. Unlike previous studies focused on single-site models, our work emphasises an integrated network of hospitals, specifically in elderly care. We uniquely optimise staffing and bed capacity jointly and explicitly account for the specialty and acuity mix of frail elderly patients. These aspects, including multi-site planning, predictive forecasting, stochastic optimisation, and a focus on elderly care, constitute our key contributions to the healthcare operations literature.

In terms of the deterministic and stochastic models, our approach builds upon the work of Maggioni and Wallace (2012). Our work extends this two-fold: first, by extending the formulation which provides a healthcare example and second, by incorporating CART models into the demand function to create real-world variation in the formulation. We extend this work by providing a healthcare example and incorporating CART models into the modelling formulation.

Second, we research an area of healthcare which has been under-researched, despite the increased demand and medical needs of these patients (Williams et al., 2021).

Finally, with respect to methodology, we provide a different perspective on how this could be modelled. Previous bed and staffing capacity models have focused on neural networks (Kutafina et al., 2019), queueing theoretical approaches (de Véricourt & Jennings, 2011; Ghayoomi et al., 2023; Yankovic & Green, 2011), simulation (Amelia et al., 2021; Lu et al., 2021). Our methodology builds on traditionally used deterministic methods in healthcare by demonstrating the benefits of using stochastic models by calculating the VSS.

## 3. Methods

This section will discuss the approaches used within this paper, providing the notation used to allow the

reader to apply the theories to their own work. There is a large amount of variation in terms of hospital LOS. We have carefully chosen these methods to allow easy incorporation of this variation. CART was chosen over other predictive methodologies because it offers a visual representation. This makes it possible for healthcare practitioners to comprehend and trust the model's output, with the potential for collaboration to create clinically and statistically meaningful groupings. Due to the complexity of our problem, a mixed integer programming approach is used, allowing the LOS variation to be integrated into the bed and staffing planning simultaneously.

### 3.1. Classification and regression trees

CART is a non-parametric, supervised machine learning technique that can be used for both classification and regression tasks. It works by recursively partitioning the data into smaller subsets based on predictor variable values. The partitioning process creates a tree-like model of decisions and possible consequences, which are the terminal nodes or leaf nodes of the tree.

One common criterion utilised in the regression side of CART is the mean squared error (MSE) to determine the optimal split at each node (Zanakis & Baccera-Fernandez, 2005), shown in Equation (1).

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2, \quad (1)$$

where  $n$  is the number of observations,  $Y_i$  is the actual value and  $\hat{Y}_i$  is the predicted value.

Algorithm 1 has been formulated to demonstrate the process of regression tree building within the sklearn package within Python.

---

#### Algorithm 1: Regression Tree

---

Determine stopping criteria:

$max\_depth, min\_samples\_split, min\_samples\_leaf,$

$min\_weight\_fraction\_leaf, max\_leaf\_nodes,$

$min\_impurity\_decrease$

Start with a single node  $n$  containing all points.

Calculate  $MSE^n$

**while**  $MSE^n > 0$  **or** *stopping criterion not met* **do**

$k =$  number of binary splits

**for**  $a = 1$  **to**  $k$  **do**

        Calculate  $MSE_a^n$

$x_a = MSE^n - MSE_a^n$

**end**

    Set  $MSE^n = \max(x_a)$

    Create two new nodes,  $n'$  and  $n''$ , and calculate new  $MSE^n$  for each.

**end**

---

The Gini index is a common criterion used within the classification side of CART and is used

to determine the optimal split at each node. The formulation is given in Equation (2).

$$\text{Gini Index} = 1 - \sum_{i=1}^n p_i^2, \quad (2)$$

where  $i$  is the number of classes and  $p_i$  is the probability of an object that is being classified to a particular class.

The respective algorithm for building a classification tree using the Gini index as the splitting criterion is given in Algorithm 2.

---

**Algorithm 2:** Classification Tree

---

```

Determine stopping criteria:
max_depth, min_samples_split, min_samples_leaf,
min_weight_fraction_leaf, max_leaf_nodes,
min_impurity_decrease
Start with a single node  $n$  containing all points.
Calculate Gini-Index $n$ 
while Gini-Index $n$  > 0 or stopping criterion not
met do
   $k$  = number of binary splits
  for  $a = 1$  to  $k$  do
    Calculate Gini-Index $a$  $n$ 
     $x_a$  = Gini-Index $n$  - Gini-Index $a$  $n$ 
  end
  Set Gini-Index $n$  = max( $x_a$ )
  Create two new nodes,  $n'$  and  $n''$ , and calculate
  new Gini-Index for each.
end

```

---

CART offers several advantages that make it an attractive choice for both classification and regression problems. One of the primary benefits is its interpretability. CART models produce intuitive decision trees that are easy to understand and explain, even for non-experts. The hierarchical structure of the trees provides a clear visualisation of the decision-making process, allowing for straightforward interpretation of the relationships between predictors and the target variable. Healthcare professionals, often not trained extensively in mathematics or statistics, find CART particularly appealing due to its user-friendly interpretability (Bertsimas et al., 2021). The simplicity of the decision trees generated by CART allows these professionals to grasp the underlying logic without needing advanced mathematical expertise. This accessibility empowers healthcare practitioners to not only utilise CART effectively but also to confidently communicate its findings and implications to patients and colleagues.

Another significant advantage of CART is its ability to handle non-linear relationships and complex interactions between variables. Unlike linear models, CART can effectively capture intricate patterns and decision boundaries, making it suitable

for a wide range of problems involving non-linear relationships. Additionally, CART algorithms perform automatic feature selection during the tree-building process, identifying the most important predictors and reducing the need for manual feature engineering.

CART models are also robust to missing data, as they can handle incomplete datasets by incorporating surrogate splits or imputation techniques during the tree-building process. Furthermore, CART does not require any assumptions about the underlying data distribution, making it a non-parametric and flexible method applicable to a variety of data types and distributions.

### 3.2. Deterministic and two-stage stochastic programming

Within this section the deterministic and stochastic mathematical programs will be discussed, providing the notation used to allow the reader to apply the theories to their own work. The aim of the model is to determine the number of beds and nursing staff required for each specialty within each hospital. The motivation for choosing these resources stems from their pivotal roles in shaping healthcare service delivery. Beds directly influence patient capacity, while nursing staff are instrumental in ensuring high-quality patient care and operational efficiency. Importantly, the allocation of these resources is inherently intertwined, as the number of beds required directly affects the staffing needs, and vice versa. This research extends on the framework used by Mestre et al. (2015) and Maggioni and Wallace (2012).

#### 3.2.1. General formulation

Let us define the two-stage stochastic problem, where a decision-maker takes the decision  $x$  of solution space  $X$  to minimise expected costs:

$$\min_{x \in X} E_{\xi} z(x, \xi) = \min_{x \in X} \{f_1(x) + E_{\xi}[h_2(x, \xi)]\}, \quad (3)$$

where  $x$  are the first-stage decision variables restricted to the set  $X \subset \mathbb{R}^n$  and  $f_1(x)$  is the value of the first stage problem.  $E_{\xi}$  indicates the expectation with respect to a random vector denoted  $\xi$  defined on the probability space  $(\Omega, \mathcal{A}, p)$ , with  $\Omega \in \mathbb{R}^n$  and probability distribution  $p$  on the  $\sigma$ -algebra  $\mathcal{A}$ .

The function  $h_2(x, \xi)$  is the value function of the second stage of the stochastic problem, defined as follows:

$$h_2(x, \xi) = \min_{y \in Y(x, \xi)} f_2(y : x, \xi), \quad (4)$$

where  $y$  is the second-stage solution which is restricted to the set  $Y \in \mathbb{R}^n$ .

Equation (4) reflects the costs associated with the information being revealed through the realisation of  $\xi$  from the random vector  $\xi$ . The term  $[h_2(x, \xi)]$ , is known as the recourse function.

The solution obtained is defined as the “here and now solution” (RP) and is the optimal value of the objective function:

$$RP = E_{\xi} z(x^*, \xi). \quad (5)$$

Equation (3) can be considered where the decision-maker replaces the random variables with their expected values and in turn, solves a deterministic model. This is also known as the expected value (EV) problem.

$$EV = \min_{x \in X} z(x, \bar{\xi}), \quad (6)$$

where  $\bar{\xi} = E(\xi)$ , which is the expected value of the random vector  $\xi$  and  $z$  is the objective value.

### 3.2.2. Sets

The sets used within the deterministic and two-stage stochastic models are pivotal for defining the parameters and variables within the model. Each set serves a distinct purpose. Firstly,  $\mathcal{B}$  represents the set of nursing bands, encompassing different skill levels and experiences of nurses. Secondly,  $\mathcal{S}$  denotes the set of specialties, where each specialty must be assigned to at least one hospital ( $\mathcal{S} \subseteq \mathcal{H}$ ). Thirdly,  $\mathcal{H}$  signifies the set of hospitals, with each hospital being associated with at least one region ( $\mathcal{H} \subseteq \mathcal{R}$ ). Moreover,  $\mathcal{R}$  stands for the set of regions, ensuring geographical coverage across the healthcare system. Regions, in this context, could be defined as counties or districts, each comprising of multiple hospitals operating within the same healthcare system organisation. Therefore  $|\mathcal{H}| \leq |\mathcal{R}|$ . Lastly,  $\mathcal{K}$  indicates the set of scenarios, capturing various potential circumstances or conditions (Table 1).

### 3.2.3. Parameters

Table 2 displays the parameters used within the deterministic and two-stage stochastic models.

### 3.2.4. Decision variables

The decision variables introduced in Table 3 determine the number of beds and nursing staff required for each specialty within each hospital.

**Table 1.** The sets used within the two-stage stochastic model where (B, S, H, R, K) represent the maximum number of nursing bands, specialties, hospitals, regions and scenarios, respectively.

Set	Range	Definition
$\mathcal{B}$	$b = 1, \dots, B$	Set of nursing bands
$\mathcal{H}$	$h = 1, \dots, H$	Set of hospitals
$\mathcal{K}$	$k = 1, \dots, K$	Set of scenarios
$\mathcal{R}$	$r = 1, \dots, R$	Set of regions
$\mathcal{S}$	$s = 1, \dots, S$	Set of specialties

### 3.2.5. Mathematical models

In this section, we present novel deterministic and two-stage stochastic models derived from the introduced sets, parameters, and decision variables. These models serve as the foundation for our analysis and provide a structured framework for addressing the problem at hand. The deterministic model can be defined as follows:

$$\min \sum_{h \in \mathcal{H}} \sum_{s \in \mathcal{S}} \left( c_{s,h}^{\text{bed}} x_{s,h}^{\text{bed}} + \sum_{b \in \mathcal{B}} c_b^{\text{staff}} x_{s,b,h}^{\text{staff}} \right), \quad (7)$$

subject to:

$$\sum_{h \in \mathcal{H}} x_{s,h}^{\text{bed}} \geq D_{s,r} \quad \forall s \in \mathcal{S}, r \in \mathcal{R}, \quad (8)$$

$$\sum_{b' \in \mathcal{B}: b' \geq b} x_{s,b',h}^{\text{staff}} \geq R_{s,b} \cdot x_{s,h}^{\text{bed}} \quad \forall s \in \mathcal{S}, b \in \mathcal{B}, h \in \mathcal{H}, \quad (9)$$

$$x_{s,h}^{\text{bed}} \leq K_{s,h} \quad \forall s \in \mathcal{S}, h \in \mathcal{H}, \quad (10)$$

$$0 \leq \sum_{s \in \mathcal{S}} x_{s,h}^{\text{bed}} \leq UB_h^{\text{max,bed}} \quad \forall h \in \mathcal{H}, \quad (11)$$

$$0 \leq \sum_{s \in \mathcal{S}} \sum_{h \in \mathcal{H}} x_{s,b,h}^{\text{staff}} \leq UB_b^{\text{max,staff}} \quad \forall b \in \mathcal{B}. \quad (12)$$

Objective function (7) minimises the cost of deploying beds and staff in each specialty and hospital. Constraints (8) ensure the number of beds deployed satisfies the demand. Constraints (9) make sure the number of staff deployed to each specialty within each hospital meets the minimum requirements. Constraints (10) ensure the number of beds deployed cannot exceed the maximum available specialty beds within each hospital. Constraints (11) and (12) denote the decision variables and their domains.

Similarly, the two-stage stochastic model can be defined as follows:

$$\min \sum_{h \in \mathcal{H}} \sum_{s \in \mathcal{S}} \left( c_{s,h}^{\text{bed,1st}} x_{s,h}^{\text{bed}} + \sum_{b \in \mathcal{B}} c_b^{\text{staff,1st}} x_{s,b,h}^{\text{staff}} \right) + \sum_{k \in \mathcal{K}} \sum_{h \in \mathcal{H}} \sum_{s \in \mathcal{S}} p_k \left( c_{s,h}^{\text{bed,2nd}} u_{s,h,k}^{\text{bed}} + \sum_{b \in \mathcal{B}} c_b^{\text{staff,2nd}} u_{s,b,k,h}^{\text{staff}} \right), \quad (13)$$

subject to:

$$\sum_{h \in \mathcal{H}} (x_{s,h}^{\text{bed}} + u_{s,h,k}^{\text{bed}}) \geq D_{s,r,k} \quad \forall s \in \mathcal{S}, r \in \mathcal{R}, k \in \mathcal{K}, \quad (14)$$

$$\sum_{b' \in \mathcal{B}: b' \geq b} x_{s,b',h}^{\text{staff}} \geq R_{s,b} \cdot x_{s,h}^{\text{bed}} \quad \forall s \in \mathcal{S}, b \in \mathcal{B}, h \in \mathcal{H}, \quad (15)$$

$$\sum_{b' \in \mathcal{B}: b' \geq b} u_{s,b',k,h}^{\text{staff}} \geq R_{s,b} \cdot u_{s,h,k}^{\text{bed}} \quad \forall s \in \mathcal{S}, b \in \mathcal{B}, h \in \mathcal{H}, k \in \mathcal{K}, \quad (16)$$

$$x_{s,h}^{\text{bed}} \leq K_{s,h} \quad \forall s \in \mathcal{S}, h \in \mathcal{H}, \quad (17)$$

$$u_{s,h,k}^{\text{bed}} \leq K_{s,h} \quad \forall s \in \mathcal{S}, h \in \mathcal{H}, k \in \mathcal{K}, \quad (18)$$

**Table 2.** The parameters used within the two-stage stochastic model where  $(b, s, h, r, k)$  represent the indices of nursing bands, specialties, hospitals, regions and scenarios, respectively.

Parameter	Definition
$C_{s,h}^{\text{bed,1st}}$	Cost of the first stage beds for specialty $s \in \mathcal{S}$ , in hospital $h \in \mathcal{H}$
$C_{s,h}^{\text{bed,2nd}}$	Cost of the second stage bed per day for specialty $s \in \mathcal{S}$ , in hospital $h \in \mathcal{H}$
$C_b^{\text{staff,1st}}$	Cost of the first stage staff of band $b \in \mathcal{B}$
$C_b^{\text{staff,2nd}}$	Cost of the second stage staff of band $b \in \mathcal{B}$
$D_{s,r,k}$	Demand for each specialty $s \in \mathcal{S}$ , arriving from region $r \in \mathcal{R}$ , for scenario $k \in \mathcal{K}$
$K_{s,h}$	Number of beds available to open in each specialty $s \in \mathcal{S}$ , in hospital $h \in \mathcal{H}$
$p_k$	Probability of scenario $k \in \mathcal{K}$
$R_{s,b}$	Ratio of nursing staff of band $b \in \mathcal{B}$ to patient for each specialty $s \in \mathcal{S}$
$UB_h^{\text{max,bed,1st}}$	Upper bound of the number of beds that are able to be deployed in hospital $h \in \mathcal{H}$ in the 1 <sup>st</sup> stage
$UB_h^{\text{max,bed,2nd}}$	Upper bound of the number of beds that are able to be deployed in hospital $h \in \mathcal{H}$ in the 2 <sup>nd</sup> stage
$UB_b^{\text{max,staff,1st}}$	Upper bound of the number of staff of band $b \in \mathcal{B}$ that can be deployed in the 1 <sup>st</sup> stage
$UB_b^{\text{max,staff,2nd}}$	Upper bound of the number of staff of band $b \in \mathcal{B}$ that can be deployed in the 2 <sup>nd</sup> stage

**Table 3.** The decision variables used within the two-stage stochastic model where  $(b, s, h, r, k)$  represent the indices of nursing bands, specialties, hospitals, regions and scenarios, respectively.

Decision variable	Definition
$x_{s,h}^{\text{bed}} \in \mathbb{N}$	Number of beds planned in the 1 <sup>st</sup> stage for specialty $s \in \mathcal{S}$ , in hospital $h \in \mathcal{H}$
$x_{s,b,h}^{\text{staff}} \in \mathbb{N}$	Number of staff planned in the 1 <sup>st</sup> stage for specialty $s \in \mathcal{S}$ , of band $b \in \mathcal{B}$ , in hospital $h \in \mathcal{H}$
$u_{s,r,h,k}^{\text{bed}} \in \mathbb{N}$	Number of beds needed in the 2 <sup>nd</sup> stage for specialty $s \in \mathcal{S}$ , for patients arriving from region $r \in \mathcal{R}$ in hospital $h \in \mathcal{H}$ , for scenario $k \in \mathcal{K}$
$u_{s,b,h,k}^{\text{staff}} \in \mathbb{N}$	Number of staff needed in the 2 <sup>nd</sup> stage for specialty $s \in \mathcal{S}$ , of band $b \in \mathcal{B}$ , in hospital $h \in \mathcal{H}$ , for scenario $k \in \mathcal{K}$

$$0 \leq \sum_{s \in \mathcal{S}} x_{s,h}^{\text{bed}} \leq UB_h^{\text{max,bed,1st}} \quad \forall h \in \mathcal{H}, \quad (19)$$

$$0 \leq \sum_{s \in \mathcal{S}} \sum_{h \in \mathcal{H}} x_{s,b,h}^{\text{staff}} \leq UB_b^{\text{max,staff,1st}} \quad \forall b \in \mathcal{B}, \quad (20)$$

$$0 \leq \sum_{s \in \mathcal{S}} u_{s,h,k}^{\text{bed}} \leq UB_h^{\text{max,bed,2nd}} \quad \forall h \in \mathcal{H}, k \in \mathcal{K}, \quad (21)$$

$$0 \leq \sum_{s \in \mathcal{S}} \sum_{h \in \mathcal{H}} u_{s,b,h,k}^{\text{staff}} \leq UB_b^{\text{max,staff,2nd}} \quad \forall b \in \mathcal{B}, k \in \mathcal{K}. \quad (22)$$

The first sum in the objective function (13) is the cost of deploying both beds and staff to specialties within each hospital. The second sum represents the additional beds and staff within the same hospital or a different hospital in the same region. The first constraint, (14), assures the demand for each specialty and region is met by the number of hospital beds deployed. The demand is dependent on the scenario parameter. Constraints (15) ensures the number of staff deployed meets the minimum requirements for staff on each specialty ward in the first stage, whilst Constraints (16) ensures this requirement is met in the second stage. Constraints (17) and (18) ensure the beds deployed do not exceed the maximum number of beds available for each specialty within each hospital. Constraints (19)–(22) denote the decision variables and their domains.

### 3.2.6. Evaluation measures

Within prescriptive analytics, it is widely recognised that the EV solution can behave poorly in the stochastic domain. Traditional evaluation tests can be carried out in order to determine how each of the EV, RP and EEV performs and determine their robustness. The EEV can be defined as the expected cost when using the solution  $\bar{x}(\bar{\xi})$ . Maggioni and Wallace (2012) discussed four tests to determine the success of stochastic models. For this research, the first method of determining the VSS will be used.

If we let  $\bar{x}(\bar{\xi})$  be the optimal solution to Equation (6), we can take values and fix these as the first stage, and then allow the second stage of the stochastic model to be performed.

$$EEV = E_{\xi}(z(\bar{x}(\bar{\xi}), \xi)). \quad (23)$$

To determine the VSS, the difference between the EEV and RP can be calculated, measuring the expected increase in value from solving the stochastic solution to the simple deterministic one:

$$VSS = EEV - RP. \quad (24)$$

The VSS measures expected loss when using the deterministic solution. If we have hard constraints, the expected cost of the deterministic solution is often  $\infty$ . Whereas if we use soft constraints, we can make the expected cost arbitrarily bad by setting penalties high using the deterministic solution. If the VSS is large, this could mean the wrong choice of variables has been chosen or the wrong values have been entered.

### 3.2.7. Illustrative example

To demonstrate the applicability of our proposed model, a small example with fictional numerical values illustrates the optimisation models and key outputs. A small number of scenarios have been included within the second stage to provide a simple illustration of the stochastic programming approach while retaining computational traceability. A dataset of 15 patients demonstrates how the models perform in the application of elderly and frail patient care (Table 4).

Within the dataset there are two hospitals within the same region, each serving the same two specialties: care of the elderly (COTE) and trauma and orthopaedics (T&O). We assume there are two nursing staff band levels required on the wards, with differing staff/bed ratios depending on the specialty. Table 5 shows the parameters and their corresponding values for the deterministic and stochastic models.

The demand from Table 4 can be calculated as follows to determine the average daily bed demand (ADBD):

$$ADBD_{s,h} = ALOS_{s,h} \times ADNA_{s,h}, \quad (25)$$

where ALOS is equal to the average LOS and ADNA is equal to the average daily number of admissions. The demand in terms of scenarios and regions can then be calculated:

$$D_{s,r} = ADBD_{s,r} = \sum_{h \in \mathcal{R}} ADBD_{s,h}. \quad (26)$$

This produces a value of 16.67 and 19.01 for the  $D_{0,0}$  and  $D_{1,0}$  parameters, respectively. For the two-stage stochastic model, a number of scenarios are required. For this example and to demonstrate the model's functionality, we introduce three scenarios which average to the same deterministic demand: Average demand with a probability of 40%, demand increasing by 20% with a probability of 30%, demand decreasing by 20% with a probability of 30%.

Therefore the demand matrix  $D_{s,r,k}$  can be represented as:  $\begin{bmatrix} [16.66, 19.99, 13.33] \\ [19.01, 22.80, 15.20] \end{bmatrix}$ , where the first index refers to the row and the second index refers to the column. In this instance we only have one region, therefore only one column matrix is shown. The third index refers to the column inside the sub-matrix. Table 6 displays the optimal decision variables and objective function values for the illustrative example.

The results reveal that the deterministic model deploys fewer beds and nursing staff than the stochastic model, leading to an EV solution that costs approximately two-thirds of the RP. The EEV value is calculated by fixing the first-stage decision variables in the two-stage stochastic model with the optimal values from the deterministic model. This then calculates the values for the second-stage variables and the objective function value. The VSS is then calculated to be the difference between the EEV and RP, resulting in a value of £56, a saving of 2.5%, if

**Table 5.** The parameter values that were used within the deterministic and two-stage stochastic model specifically for the illustrative example.

Parameters	Values
1 <sup>st</sup> Stage Bed Costs ( $c_{s,h}^{bed,1st}$ )	$\begin{bmatrix} 20 & 30 \\ 30 & 40 \end{bmatrix}$
2 <sup>nd</sup> Stage Bed Costs ( $c_{s,h}^{bed,2nd}$ )	$\begin{bmatrix} 22 & 33 \\ 33 & 44 \end{bmatrix}$
Ratio ( $R_{s,b}$ )	$\begin{bmatrix} 0.29 & 0.14 \\ 0.14 & 0.29 \end{bmatrix}$
Maximum Specialty Capacity ( $K_{s,h}$ )	$\begin{bmatrix} 20 & 25 \\ 20 & 25 \end{bmatrix}$
1 <sup>st</sup> Stage Staff Costs ( $c_b^{staff,1st}$ )	[£50, £60]
2 <sup>nd</sup> Stage Staff Costs ( $c_b^{staff,2nd}$ )	[£55, £66]
Upper 2 <sup>nd</sup> bed limit ( $UB_{h,k}^{max,bed,2nd}$ )	$\begin{bmatrix} 20 & 20 & 20 \\ 25 & 25 & 25 \end{bmatrix}$
Upper 1 <sup>st</sup> staff limit ( $UB_b^{max,staff,1st}$ )	[15, 25]
Upper 2 <sup>nd</sup> staff limit ( $UB_{b,k}^{max,staff,2nd}$ )	$\begin{bmatrix} 15 & 15 \\ 25 & 25 \end{bmatrix}$
Probability of Scenarios ( $p_k$ )	[0.4, 0.3, 0.3]

**Table 4.** Illustrative example patient dataset with 15 patients.

Patient number	Age	Hospital	LOS	Specialty	Admission method	Admission source	Frailty source
Patient 1	95	1	5	COTE	Emergency	Own Home	3
Patient 2	82	1	3	COTE	Emergency	Own Home	2
Patient 3	89	1	4	T&O	Emergency	Own Home	2
Patient 4	87	1	4	T&O	Elective	Own Home	2
Patient 5	85	2	3	COTE	Elective	Transferred	1
Patient 6	76	2	1	COTE	Elective	Transferred	1
Patient 7	71	2	1	T&O	Emergency	Transferred	1
Patient 8	96	1	5	T&O	Emergency	Own Home	3
Patient 9	70	2	1	COTE	Emergency	Transferred	1
Patient 10	67	2	1	T&O	Elective	Own Home	1
Patient 11	89	1	4	COTE	Elective	Transferred	3
Patient 12	70	2	1	COTE	Elective	Own Home	2
Patient 13	75	2	4	T&O	Elective	Transferred	3
Patient 14	72	2	2	COTE	Elective	Transferred	3
Patient 15	87	1	5	COTE	Emergency	Own Home	2

**Table 6.** Deterministic and two-stage stochastic results for the illustrative example – results are recorded [(beds), (staff)].

	$s = 0, h = 0$	$s = 0, h = 1$	$s = 1, h = 0$	$s = 1, h = 1$	Objective Function Value (£)
EV	[(0), (0, 0)]	[(17), (3, 5)]	[(20), (3, 6)]	[(0), (0, 0)]	2,050
RP	[(20), (3, 6)]	[(1), (1, 1)]	[(24), (4, 8)]	[(0), (0, 0)]	2,185
EEV	[(4), (1, 2)]	[(17), (3, 5)]	[(24), (4, 8)]	[(0), (0, 0)]	2,241

the stochastic solution were to be implemented. The results also show that the deterministic model is not robust, as the EEV is greater than the RP. This is due to the deterministic model not being able to account for the uncertainty in the demand.

#### 4. Case study of a network of hospitals in the U.K

In this section, we apply the predictive and prescriptive analytical methodologies to a case study involving a network of hospitals. We aim to demonstrate the practical application and benefits of our approach in a real-world setting. Three years' worth of data covering the period from April 2017 to March 2020 was provided by a large NHS trust in the U.K. To ensure data integrity and avoid skewing from the COVID-19 pandemic, only data preceding the pandemic was utilised. The dataset comprises 165,188 patient records of individuals aged 65 and over, encompassing 29 different specialties and spanning 11 hospital sites. This comprehensive dataset forms the basis of our analysis and modelling efforts, enabling us to derive insights and optimise resource allocation effectively.

##### 4.1. Results of CART analysis

This section discusses the development and results of the CART models. This section will be split into two subsections, the first for regression trees (Section 4.1.1) and the second for classification trees (Section 4.1.2). The results of the CART analysis will be used to inform the deterministic and stochastic models.

###### 4.1.1. Regression trees

Regression trees were developed to predict the LOS of patients. Admission method, admission source, age (as a continuous and grouped measure), day, diagnosis, frailty (as a continuous and grouped), hospital, number of scans and specialty were the nine variables used within the regression trees.

Age and frailty were included as continuous and grouped variables to determine which would be more appropriate for the regression trees. The models were trained on 80% of the data with a 20% test set being used to validate the models.

The “Scikit-learn” package within Python was used to generate the regression trees utilising the

“DecisionTreeRegressor” function. The default parameters were utilised within the model with the exception of the “max\_leaf\_nodes” and the “min\_samples\_leaf” parameters. The default parameters have been successfully employed in other studies (Hancock & Khoshgoftaar, 2022; Heyburn et al., 2018; Kilincer et al., 2023). These “max\_leaf\_nodes” and the “min\_samples\_leaf” parameters underwent parameter optimisation through a step wise approach. This was to ensure underfitting and overfitting were avoided (Bramer, 2007). Additionally, we wanted to set the upper limit of the “max\_leaf\_nodes” to a manageable amount of groupings. The parameters used have been provided within Table 2 in the Supplementary Material and were selected based on the default values and parameter optimisation to determine the largest  $R^2$  score.

The largest  $R^2$  score was achieved by grouping age into five-year intervals and using frailty as a continuous measure. This value was calculated to be 34.2%. Whilst this is a low  $R^2$  score, it is expected due to the large variation in LOS within the data (417 days). This result shows the model correctly assigns patients to the correct node 34.2% of the time.

The achieved  $R^2$  score of 34.2% by the regression trees in predicting the LOS of patients may appear low at first glance. However, several factors contribute to this result.

Firstly, the dataset exhibits a wide range of LOS values, with a maximum LOS of 417 days. Such variability in LOS can make accurate prediction challenging, as numerous factors, beyond those included in the model, may influence the length of a patient's stay.

Secondly, despite including relevant variables such as admission method, source, diagnosis, and frailty, the predictive power of the model may be limited by the complexity of the LOS prediction task. Some nuances of patient stays may not be fully captured by the chosen variables, leading to inherent limitations in the model's accuracy.

Moreover, while regression trees offer flexibility and interpretability, they may struggle to capture complex relationships or interactions between variables, especially in datasets with high variability like ours. To mitigate this, we carefully optimised model parameters, including “max\_leaf\_nodes” and “min\_samples\_leaf”, to balance model complexity and performance.

Furthermore, it's important to highlight that the utility of decision trees extends beyond predictive accuracy alone. Decision trees offer a transparent and interpretable framework for understanding the factors influencing patient LOS. Healthcare practitioners can readily interpret decision trees, making them valuable tools for clinical decision-making and care management.

While the achieved  $R^2$  score may seem modest, the visual representation provided by decision trees empowers healthcare practitioners to identify critical decision points and understand the underlying factors contributing to patient LOS. This interpretability can facilitate targeted interventions and resource allocation, ultimately improving patient outcomes and healthcare delivery efficiency.

Therefore, despite the relatively low  $R^2$  score, the inherent interpretability and actionable insights offered by decision trees make them a valuable asset for healthcare practitioners seeking to optimise patient care pathways and resource utilisation.

#### 4.1.2. Classification trees

Classification trees were also developed to predict patients who were discharged on the same day or admitted overnight. In addition to the nine variables used within the regression trees, month was also included as a predictor, since it was found to be a significant variable within logistic regression analysis. Similar to the regression trees, the model was performed with both age and frailty as a continuous and grouped measure to determine which would be more appropriate. The models were trained on 80% of the data with a 20% test set being used to validate the models. The "DecisionTreeClassifier" function was used to develop the classification trees using the "Scikit-learn" package within Python. The parameters used for the "DecisionTreeClassifier" function are shown in Table 2 in the [Supplementary Material](#). The accuracy score was calculated to be 89.9% when age was grouped and frailty was used as a continuous measure whilst using one minimum sample per leaf and 30 maximum leaf nodes. The groupings determined by the end nodes of the CART models can be used to determine the demand for each specialty within each hospital.

#### 4.2. Aggregate results of the deterministic vs. stochastic multi-site model

This subsection will utilise the average demand for each specialty and each region from the three years' worth of data, and feed this into the deterministic and two-stage stochastic model. All optimisation models were solved using both Microsoft Excel with OpenSolver and Python with the PuLP optimisation

library. Python provided significantly faster computational times, demonstrating the efficiency and practical feasibility of the approach for real-world implementation and rapid decision support. In all cases, models were solved to optimality. The models have been provided open source *via* GitHub (Williams, 2023).

The 11 healthcare locations can be grouped into six different regions, offering 29 specialties across all hospitals. In practice, there are 90 combinations of where specialties can be located, with not all hospitals offering every specialty. Within the NHS, there are different levels of nursing staff, ranging from band two to band eight. The higher the band, the more senior the staff member. Two levels of nursing bands will be used to develop the models (bands five and six). In order to gather costing figures, open source data from Public Health Scotland (2021) was used. The costing data and parameter values are shown in Table 1 in the [Supplementary Material](#) for reference.

In order to meet the demand and satisfy the constraints, the deterministic model utilised the first-stage variables only. The results yielded a yearly cost of £904,281. In total, 1,027 beds across different hospitals were deployed with Figure 1 displaying the precise locations of these beds. In order to satisfy demand a total of 414 NHS nurses across a 24-h period were required.

The two-stage stochastic model was considered with three different years of data in each scenario, with weightings based on the number of admissions per year. This approach aims to ascertain, drawing insights from past performance, the most effective strategies for future planning and decision-making. Compared to the deterministic model, the stochastic model deployed a higher total number of 1,278 beds across the two stages. The stochastic model also allocated a number of more beds to specialties such as COTE, General Medicine and Rehabilitation. Additionally, the distribution of beds across hospitals differed for certain specialties between the two models, providing more flexibility and adaptability to varying demand patterns in the stochastic approach.

- Year one (2017–2018) with a probability of 32.2%
- Year two (2018–2019) with a probability of 34.0%
- Year three (2019–2020) with a probability of 33.8%

The average of all scenarios is equal to the deterministic daily demands.

The results of the stochastic model yielded a yearly cost of £923,828. In total, 941 beds are deployed

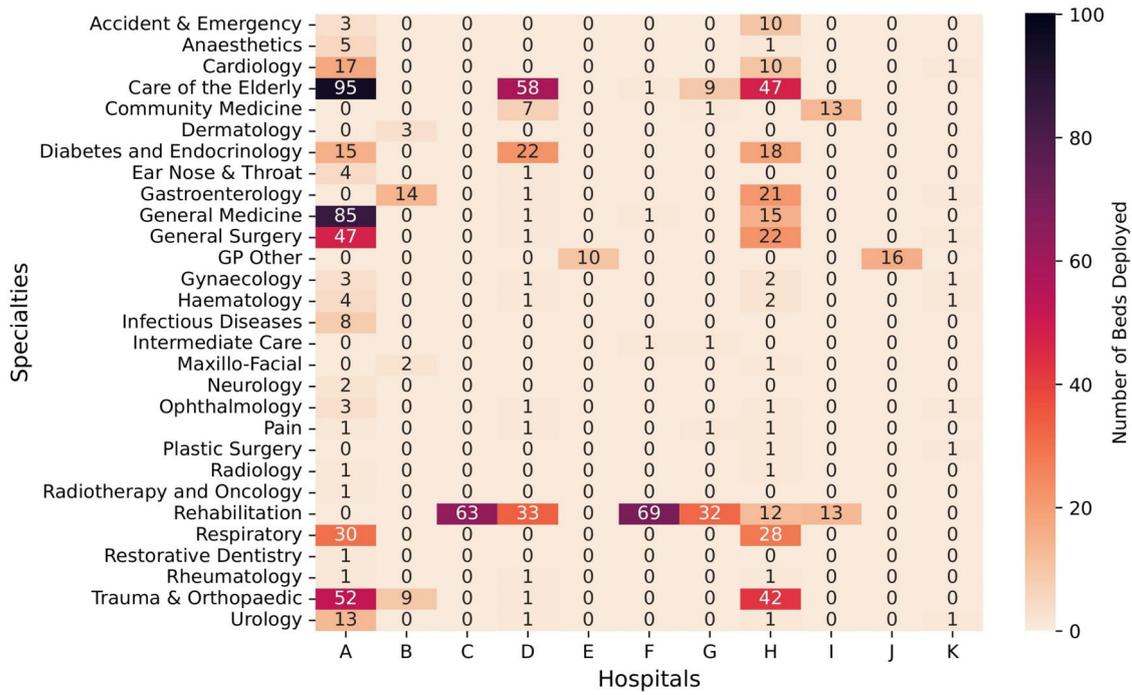


Figure 1. Heatmap of bed locations for each specialty within each hospital for the deterministic model.

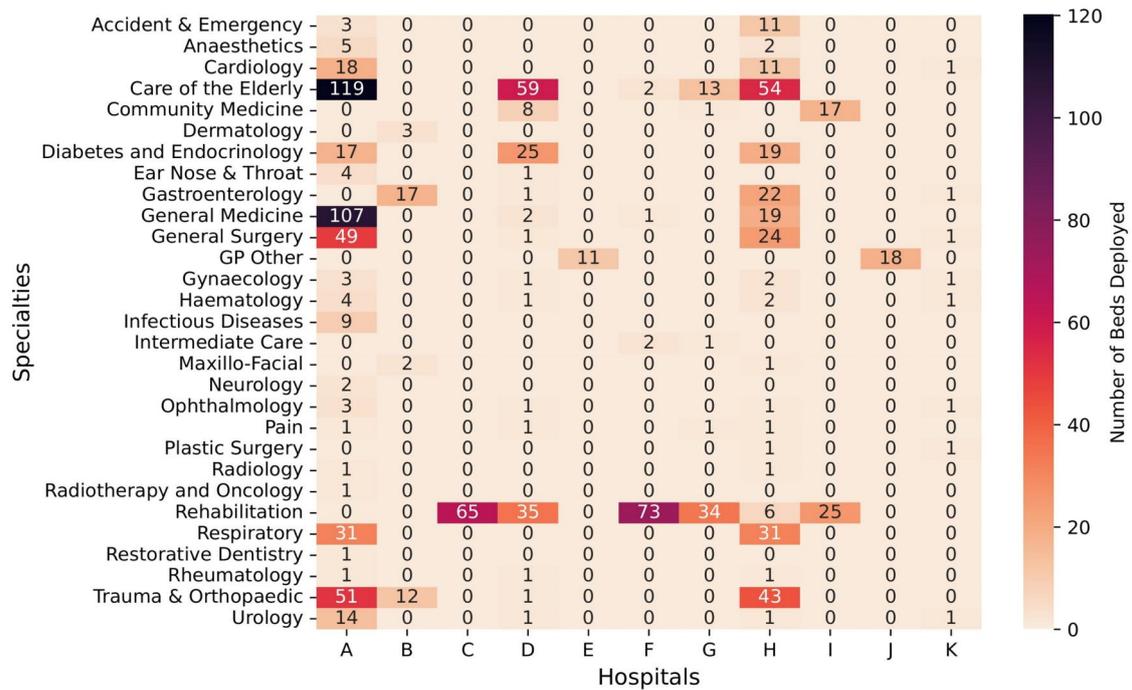


Figure 2. Heatmap of bed locations for each specialty within each hospital for the two-stage stochastic model.

within the first stage, with a maximum of 337 in the second stage. Similarly, 86 additional nursing staff are deployed across both stages of the model compared to the deterministic counterpart. The location of the 1,278 beds can be seen within Figure 2.

To provide an understanding of why both models come to different values, we now investigate the robustness of the deterministic model. The VSS was calculated by fixing the results from the deterministic model as the first stage within the two-stage stochastic model. The VSS was calculated to be £35,439 over a year period, a saving of  $\approx 3.8\%$  (Table 7). This

Table 7. The EV, RP and EEV values for the  $x^{bed}$ ,  $x^{staff}$ ,  $u^{bed}$  and  $u^{staff}$  decision variables and the objective function value.

	Total beds		Total staff		Objective function value (£)
	$x^{bed}$	$u^{bed}$	$x^{staff}$	$u^{staff}$	
EV	1,026	–	414	–	904,281
RP	941	337	346	154	923,828
EEV	1,026	186	414	104	959,267

indicates that the deterministic model is not robust, as the EEV is greater than the RP. The primary reason for this difference is that the deterministic model

operates under the assumption of a fixed, known demand, which simplifies the decision making process but fails to incorporate the variability and uncertainty inherent in real-world scenarios.

In contrast, the two-stage stochastic model explicitly considers the uncertainty in demand by allowing for multiple possible future scenarios. This enables more flexible and adaptive decision-making that can better respond to actual demand fluctuations. Consequently, the stochastic model can optimise decisions more effectively, leading to potential cost savings and improved performance.

### 4.3. Results broken down by specialty

In this subsection, we break down the results by specialty to understand why the deterministic and stochastic models yield different outcomes.

The variation in demand, especially for high-impact specialties such as COTE, plays a significant role in these differences. The COTE specialty often has high variability in patient demand and longer LOS, making it a critical factor in resource allocation.

To illustrate, consider the allocation of beds for the COTE specialty across all hospitals. The deterministic model, which does not account for variability in demand, assigns a fixed number of 210 beds for COTE (Figure 1). This fixed allocation is based on average demand and does not adjust for peak periods, potentially leading to shortages during times of high demand and under-utilisation during low-demand periods. In contrast, the stochastic model incorporates variability and dynamically adjusts resource allocation to better match real-time demand fluctuations. As shown in Figure 2, the stochastic model allocates up to 247 beds for COTE when necessary, reflecting its flexibility in responding to peaks in demand. This dynamic adjustment capability allows the stochastic model to efficiently manage resources, reducing the likelihood of shortages and improving overall service delivery.

The stochastic model's ability to handle variability results in several benefits. Although it requires a higher initial deployment of resources, such as additional beds, this approach ultimately leads to cost savings. By preventing bed shortages and ensuring that patient care needs are met during peak times, the stochastic model minimises the need for expensive emergency measures and improves patient outcomes. This proactive allocation strategy exemplifies the model's superiority in handling high-impact specialties with significant demand variability.

Overall, the comparison between deterministic and stochastic models highlights the importance of incorporating demand variability into resource planning, particularly for specialties with high impact and variability like COTE. By dynamically adjusting

resource allocation, the stochastic model demonstrates its effectiveness in optimising resource use and enhancing the healthcare system's responsiveness to patient needs.

### 4.4. Results of linking the two paradigms

To explore the relationship between patient characteristics and LOS within the context of decision trees, we investigated multiple methods that can be employed to derive insights from both classification and regression tree results. These methods offer different perspectives on how patient attributes and healthcare facility factors contribute to LOS prediction.

The first method involved calculating the number of patients of each specialty and determining the overall average LOS for each end node. This approach provides a broad overview of LOS patterns across different patient populations and allows for the identification of specialties with particularly high or low LOS averages.

The second method utilised each end node and considered the specific LOS for each specialty and hospital within the node. By aggregating LOS data at this granular level, we obtained a more detailed understanding of the factors influencing LOS within specific patient groups and healthcare facilities.

Moreover, our analysis demonstrated that employing 30 end node groups, as opposed to traditional methods, offers significant benefits in capturing nuanced relationships between patient characteristics, healthcare settings, and LOS prediction. This approach provides a finer granularity of analysis, enabling more precise identification of factors influencing LOS and facilitating targeted interventions for improved patient outcomes.

Moving forward, we will focus on exploring utilising the classification tree with specific LOS for each node, in more detail to demonstrate its applicability in uncovering nuanced relationships within the British healthcare context. This deeper exploration will shed light on the potential insights that can be gleaned from decision tree analyses in healthcare research and clinical practice.

The process for generating demands based on these specific LOS values is depicted in Equation (27), which accounts for both the number of patients and their LOS within each specialty and hospital.

$$D_{s,r} = \sum_{h \in r} D_{s,h} = \frac{\text{Number of Patients}_{s,h} \times \text{Specific LOS}_{s,h}}{\text{Total Number of Days in Dataset}}. \quad (27)$$

These refined demand generation processes were used to understand how CART can be linked to

**Table 8.** The EV, RP and EEV values for the  $x^{\text{bed}}$ ,  $u^{\text{bed}}$ ,  $x^{\text{staff}}$ ,  $u^{\text{staff}}$  decision variables and objective function using the classification tree and the specific LOS across all three years.

	Total beds		Total staff		Objective function value (£)
	$x^{\text{bed}}$	$u^{\text{bed}}$	$x^{\text{staff}}$	$u^{\text{staff}}$	
EV	1,002	–	422	–	840,245
RP	911	343	348	162	862,155
EEV	1,002	185	422	98	894,198

both deterministic and two-stage stochastic models, with the results presented in Table 8. Additionally, the EEV was computed to assess the VSS within the context of the models' performance.

Calculating the VSS from the results in Table 8 produces a saving of £32,0423 per day (3.7%). Although the calculated VSS is a smaller value compared to the VSS from average based planning, the specific LOS model results in a lower overall EEV, with a 7% cost difference between the two EEV values. The model suggests deploying fewer beds and nursing staff in the first stage when comparing the RP to the EEV results, however, given the uncertainties of the second stage, it is more cost-efficient to maintain flexibility with fewer resources initially, resulting in long-term savings and improved cost efficiency.

The comparison between the deterministic and CART-linked models reveals significant advantages associated with integrating CART methods into healthcare planning processes. Firstly, the CART approach offers enhanced granularity and insight into LOS prediction compared to traditional deterministic models. By considering specific LOS values for each specialty and hospital within each node, the CART model provides a more nuanced understanding of the factors influencing LOS. This finer level of detail enables healthcare planners to tailor interventions and allocate resources more effectively, addressing the diverse needs of different patient populations and healthcare facilities.

Moreover, the CART-linked models present opportunities for cost savings through more robust decision-making. By leveraging predictive analytics and historical data, the CART model can identify patterns and trends in LOS that may not be apparent in traditional deterministic approaches. This enables healthcare planners to make more informed decisions about resource allocation, optimising bed capacity, staffing levels, and other resources to meet patient demand more efficiently. As a result, healthcare organisations can potentially reduce operational costs while improving patient outcomes and satisfaction.

The VSS savings observed in the CART-linked models further emphasise the potential value of leveraging CART techniques in healthcare resource management. The VSS analysis quantifies the

potential cost savings associated with incorporating uncertainty into decision-making processes, highlighting the importance of considering both historical data and predictive analytics in strategic planning. By accounting for variability and uncertainty in patient demand and LOS, the CART-linked models enable healthcare organisations to make more resilient and adaptable decisions, better positioning them to respond to changing healthcare needs and environmental factors.

Overall, the comparison of results between the deterministic and CART-linked models underscores the transformative potential of CART methods in healthcare planning. By providing enhanced granularity, insight, and cost savings opportunities, CART techniques offer a powerful tool for optimising resource allocation, improving operational efficiency, and ultimately enhancing patient care delivery in healthcare settings.

## 5. Discussion

In this section, we synthesise the key findings of our research and outline their implications for healthcare resource planning. By linking predictive and prescriptive analytics, our research offers a comprehensive approach to optimising hospital bed and staffing allocations, particularly for frail and elderly patients. We will discuss the practical applications of our models, examine their limitations, and suggest directions for future research. Our goal is to provide actionable insights and recommendations that can enhance decision making processes in healthcare management.

The integration of predictive and prescriptive analytics in the context of healthcare resource planning, specifically for frail and elderly patients, presents a robust approach to enhancing decision-making processes. Our research demonstrates the effectiveness of combining CART analysis with deterministic and stochastic programming models. By analysing over 165,000 patient records, CART analysis provided insights into patient groupings based on clinical and demographic attributes affecting the length of stay (LOS). These insights were crucial in developing our predictive models, which in turn informed the resource allocation strategies in our optimisation models.

The deterministic and stochastic models we developed address both the certainty and uncertainty inherent in healthcare demand. The deterministic model, while straightforward, operates under the assumption of fixed future demand, which is rarely the case in real-world settings. In contrast, the stochastic model incorporates multiple scenarios to account for the variability in patient admissions

and LOS, offering a more flexible and realistic planning tool. The integration of CART-based predictions into these models ensures that variations and uncertainties are considered, leading to more accurate and practical solutions.

Our findings underscore the value of the stochastic approach. The VSS analysis revealed that incorporating uncertainty into the optimisation process significantly improves resource allocation, reducing costs while maintaining or enhancing patient care quality. This is particularly relevant in the healthcare sector, where demand is highly unpredictable, and the cost of over- or under-estimating resources can be substantial.

### **5.1. Bridging the literature gap**

The literature search of [Section 2.3](#) has revealed gaps in linking predictive and prescriptive analytics to optimise resource allocation under uncertainty in healthcare. Current literature relies heavily on deterministic optimisation models for healthcare planning, which fail to capture inherent variability. We bridge this gap by integrating data-driven predictive modelling with stochastic optimisation. Specifically, we use CART to generate demand inputs for a two-stage stochastic program, creating a robust optimisation model. This combination of predictive and prescriptive analytics differentiates our work from existing literature. Additionally, our focus on an integrated network of hospitals providing elderly care contrasts with single-site models that dominate previous studies. Furthermore, our joint optimisation of staffing and bed capacity is unique compared to other papers that optimise these decisions separately. Finally, the explicit incorporation of specialty and acuity mix for frail elderly patients differs from previous studies.

Moreover, our contribution includes the development of a novel model and a comprehensive process/workflow that evaluates the quality of different models. This process ensures rigorous assessment of model performance, including predictive accuracy and optimisation effectiveness, by employing relevant metrics and validation techniques. This systematic approach enhances the reliability and applicability of our findings in real-world healthcare settings, marking a significant advancement in the field.

### **5.2. Limitations**

Despite the promising results, our research has several limitations that warrant consideration. Firstly, the CART analysis, while powerful, has inherent limitations. The method can be prone to overfitting,

especially when dealing with highly complex datasets. Although we employed techniques to mitigate this risk, such as pruning and cross-validation, the potential for overfitting remains a concern.

Furthermore, the stochastic model's accuracy is heavily dependent on the quality of the input data and the scenarios generated. Inaccurate or biased data can lead to suboptimal resource allocation decisions. While our research utilised robust data and scenario generation techniques, the ever-changing nature of healthcare demand means that continuous updates and validations of the model are necessary.

Lastly, the computational complexity of the stochastic model can be a barrier to its practical implementation. Solving such models requires significant computational resources and time, which may not be feasible for all healthcare institutions, particularly those with limited resources. Future research could focus on developing more efficient algorithms and exploring the use of advanced computing techniques, such as parallel processing, to enhance the model's practicality.

In conclusion, the results from our innovative approach bridging the world of predictive analytics with the prescriptive analytics paradigm provides valuable insights into the application of predictive and prescriptive analytics in healthcare resource planning, these limitations highlight the need for cautious interpretation of the results and suggest directions for future research. Addressing these limitations will be crucial in refining the models and ensuring their broader applicability and effectiveness in diverse healthcare settings.

### **5.3. Managerial insights and generalisability of the results**

The results discussed can provide guidance and recommendations for enhancing healthcare bed and staffing allocations. One key takeaway from the models is the clear demonstration of the drawbacks of planning solely based on averages. Specifically, the model outputs reveal that using average LOS-based demand predictions yields an EEV that is 7% higher than when using patient-specific LOS predictions. This cost difference amounts to a potential saving of £34,425 per day by accounting for patient-level variability. Additionally, the VSS for the specific LOS model demonstrates that flexibility and scenario-based planning reduce daily operational costs by £32,042, further underscoring the financial and resource management benefits of avoiding average-based approaches. This eye-opening insight has underscored the importance of adopting more sophisticated and dynamic approaches to resource planning, steering the health board away from

potential pitfalls in their decision-making process. Perhaps the most impactful aspect of this project lies in its utilisation of predictive modelling, specifically within the healthcare domain. For a healthcare provider who is accustomed to simpler average models, this research has showcased the true potential of mathematical modelling, revealing its power in unravelling complexities, optimising operations, and delivering data-driven insights into healthcare planning. The predictive models developed in this research were trained on data collected under the existing bed and staffing allocations, aiming to capture patterns in patient LOS and demand. While acknowledging the potential for endogeneity in demand generation, particularly in elective specialties, we argue that the observed demand patterns are largely exogenous to the staffing and bedding decisions. This consideration regarding the exogeneity of demand in relation to staffing and bedding decisions should be taken into account when applying the developed predictive models to other healthcare settings or similar problems.

For practitioners to adopt these models effectively, several steps are necessary. First, robust data collection systems must be in place to capture real-time patient information, including clinical and demographic data for predictive modelling. Second, user-friendly decision-support tools need to integrate predictive outputs with optimisation algorithms, allowing healthcare managers to simulate scenarios and adjust resource plans dynamically. Third, training for healthcare decision-makers is essential to interpret model outputs and apply them in policy and operational contexts. Lastly, institutional support for integrating predictive and prescriptive analytics into routine planning processes can facilitate sustained improvements in efficiency and patient care.

Generalising results is a critical aspect of research that helps to ensure that the findings of a study are relevant and applicable beyond the specific context in which they were obtained. This makes it possible to guarantee that the research will be beneficial and instructive for other academics, professionals, and policymakers who could be working in other locations or with various populations. The deterministic and two-stage stochastic equations are able to be applied to any healthcare scenario. Whilst this research particularly focused on frail and elderly patients, due to the changing population demographics within the area, the equations can be extended to other age groupings. The benefit of using CART models is that researchers and clinicians can apply the theory to their own patient types and identify distinctive homogeneous clusters of patient features. As time passes and the

demographic of patients changes, these models can be rerun to determine new patient clusters. The user can choose the number of hospitals in each region and the range of specialties they may provide because of the equations' structure, which allows the models to be adjusted to fit any size region. Whilst these models were run with three levels of nursing bands, these can be increased or decreased to suit the user. Additionally, if decision-makers wanted to determine the needs for other hospital resources such as ventilators, these could be easily added to the model. The models are adaptable and reliable to suit a variety of healthcare situations.

#### **5.4. Future work**

This work could be extended in several ways based on the required application. One avenue could involve delving into causal models to establish deeper analytical frameworks for understanding the relationships between patient attributes and LOS. Additionally, exploring online algorithms for real-time resource management offers a promising avenue for dynamic allocation strategies, addressing day-to-day or specific time-to-time decisions in healthcare settings, especially considering factors such as hospital utilisation and seasonal demand changes. Implementing reinforcement learning strategies based on Markov decision processes could further enhance resource allocation approaches, particularly in managing dynamic patient demands and optimising resource utilisation. Furthermore, characterising the trade-offs between different methodological approaches, such as the two-stage stochastic program and alternative models like infinite-horizon dynamic programming, will provide deeper insights into methodological choices and their implications for healthcare resource management. Additionally, conducting a comprehensive sensitivity analysis to identify which parameters and modelling assumptions have the biggest potential impact on cost savings and resource optimisation will allow for the prioritisation of refinements aimed at further improving the predictive and prescriptive models.

#### **6. Conclusion**

By linking predictive and prescriptive analytics, decision-makers can obtain a comprehensive view of their data and use it to make better decisions. For example, if predictive analytics indicates there is a high likelihood of a certain event occurring in the future, prescriptive analytics can recommend specific actions that can be taken to mitigate the risk or take advantage of the opportunity. Furthermore, this

integration can also allow decision-makers to continuously improve their decision-making processes over time. By tracking the effectiveness of their decisions and making adjustments based on new data and insights, they can optimise their operations and achieve better outcomes.

This research has discussed how predictive and prescriptive analytics could be used in combination for efficiently planning hospital specialty beds and staffing requirements for a network of hospitals in the U.K. By comparing the regression tree and classification results to the averages, it allowed differences to be determined and validation of the linked methods to take place. The results showed regression trees produced closer results to the averages. The validation of these regression trees paves the way for more complex scenario analysis to be able to take place.

### Disclosure statement

The authors have no competing interests to declare that are relevant to the content of this article.

### Funding

The authors would like to acknowledge the generous support and contributions from NHS staff in the Aneurin Bevan University Health Board as well as the financial contribution from the Knowledge Economy Skills Scholarship (KESS2) and the Leadership Engagement Acceleration & Partnership (LEAP). KESS2 is a pan-Wales higher level skills initiative led by Bangor University on behalf of the Higher Education sector in Wales. It is part funded by the Welsh Government's European Social Fund (ESF) convergence programme for East Wales. LEAP is an EPSRC Digital Health Hub with award reference number [EP/X031349/1].

### ORCID

Elizabeth Williams  <http://orcid.org/0000-0003-4515-441X>

Daniel Gartner  <http://orcid.org/0000-0003-4361-8559>

Paul Harper  <http://orcid.org/0000-0001-7894-4907>

### References

- Abdelaziz, F. B., & Masmoudi, M. (2012). A multiobjective stochastic program for hospital bed planning. *Journal of the Operational Research Society*, 63(4), 530–538. <https://doi.org/10.1057/jors.2011.39>
- Amelia, P., Lathifah, A., Haq, M. D., Reimann, C. L., & Setiawan, Y. (2021). Optimising outpatient pharmacy staffing to minimise patients queue time using discrete event simulation. *Journal of Information Systems Engineering and Business Intelligence*, 7(2), 102–111. <https://doi.org/10.20473/jisebi.7.2.102-111>
- Ardakani, E. S., Larimi, N. G., Nejad, M. O., Hosseini, M. M., & Zargoush, M. (2023). A resilient, robust transformation of healthcare systems to cope with covid-19 through alternative resources. *Omega*, 114, 102750. <https://doi.org/10.1016/j.omega.2022.102750>
- Bertsimas, D., Pauphilet, J., Stevens, J., & Tandon, M. (2021). Predicting inpatient flow at a major hospital using interpretable analytics. *Manufacturing & Service Operations Management*, 24(6), 2797–3306. <https://doi.org/10.1287/msom.2021.0971>
- Best, T. J., Sandıkçı, B., Eisenstein, D. D., & Meltzer, D. O. (2015). Managing hospital inpatient bed capacity through partitioning care into focused wings. *Manufacturing & Service Operations Management*, 17(2), 157–176. <https://doi.org/10.1287/msom.2014.0516>
- Bhattacharjee, P., & Ray, P. K. (2014). Patient flow modelling and performance analysis of healthcare delivery processes in hospitals: A review and reflections. *Computers and Industrial Engineering*, 78, 299–312. <https://doi.org/10.1016/j.cie.2014.04.016>
- Bramer, M. (2007). Avoiding overfitting of decision trees. In *Principles of data mining* (p. 119–134). Springer.
- Byeon, H. (2015). Development of prediction model for endocrine disorders in the Korean elderly using CART algorithm. *International Journal of Advanced Computer Science and Applications*, 6(9), 125–129. <https://doi.org/10.14569/IJACSA.2015.060916>
- Chanchaichujit, J., Tan, A., Meng, F., & Eaimkhong, S. (2019). Optimization, simulation and predictive analytics in healthcare. In *Healthcare 4.0: Next generation processes with the latest technologies* (pp. 95–121). Springer Singapore.
- Charles, V., Mousavi, S. M. H., Gherman, T., & Mosavi, S. M. H. (2024). From data to action: Empowering covid-19 monitoring and forecasting with intelligent algorithms. *Journal of the Operational Research Society*, 75(7), 1261–1278. <https://doi.org/10.1080/01605682.2023.2240354>
- Costa, A. X., Ridley, S. A., Shahani, A. K., Harper, P. R., Senna, V. D., & Nielsen, M. S. (2003). Mathematical modelling and simulation for planning critical care capacity. *Anaesthesia*, 58(4), 320–327. <https://doi.org/10.1046/j.1365-2044.2003.03042.x>
- Dash, S., Shakyawar, S. K., Sharma, M., & Kaushik, S. (2019). Big data in healthcare: Management, analysis and future prospects. *Journal of Big Data*, 6, 54. <https://doi.org/10.1186/s40537-019-0217-0>
- Dehghani, M., Abbasi, B., & Oliveira, F. (2021). Proactive transshipment in the blood supply chain: A stochastic programming approach. *Omega*, 98, 102112. <https://doi.org/10.1016/j.omega.2019.102112>
- de Véricourt, F., & Jennings, O. B. (2011). Nurse staffing in medical units: A queueing perspective. *Operations Research*, 59(6), 1320–1331. <https://doi.org/10.1287/opre.1110.0968>
- Erhard, M., Schoenfelder, J., Fügner, A., & Brunner, J. O. (2018). State of the art in physician scheduling. *European Journal of Operational Research*, 265(1), 1–18. <https://doi.org/10.1016/j.ejor.2017.06.037>
- Fildes, R., Nikolopoulos, K., Crone, S. F., & Syntetos, A. A. (2008). Forecasting and operational research: A review. *Journal of the Operational Research Society*, 59(9), 1150–1172. <https://doi.org/10.1057/palgrave.jors.2602597>
- Galli, L., Levato, T., Schoen, F., & Tigli, L. (2021). Prescriptive analytics for inventory management in health care. *Journal of the Operational Research Society*, 72(10), 2211–2224. <https://doi.org/10.1080/01605682.2020.1776167>

- Ghayoomi, H., Miller-Hooks, E., Tariverdi, M., Shortle, J., & Kirsch, T. D. (2023). Maximizing hospital capacity to serve pandemic patient surge in hot spots via queuing theory and microsimulation. *IIEE Transactions on Healthcare Systems Engineering*, 13(4), 314–332. <https://doi.org/10.1080/24725579.2022.2149936>
- Grange, H., Johns, G., Ahuja, A., Harper, P., Williams, E., & Gartner, D. (2024). Capacity planning of virtual wards for frail and elderly patients. *Healthcare*, 12(5), 533. <https://doi.org/10.3390/healthcare12050533>
- Guo, C., Bodur, M., Aleman, D. M., & Urbach, D. R. (2021). Logic-based benders decomposition and binary decision diagram based approaches for stochastic distributed operating room scheduling. *INFORMS Journal of Computing*, 33(4), 1551–1569. <https://doi.org/10.1287/ijoc.2020.1036>
- Hancock, J., Khoshgoftaar, T. M. (2022). *Optimizing ensemble trees for big data healthcare fraud detection* [Paper presentation]. 2022 IEEE 23rd International Conference on Information Reuse and Integration for Data Science (IRI) (pp. 243–249).
- Hare, W., Alimadad, A., Dodd, H., Ferguson, R., & Rutherford, A. (2009). A deterministic model of home and community care client counts in British Columbia. *Health Care Management Science*, 12(1), 80–98. <https://doi.org/10.1007/s10729-008-9082-7>
- Harper, P. (2002). A framework for operational modelling of hospital resources. *Health Care Management Science*, 5(3), 165–173. <https://doi.org/10.1023/a:1019767900627>
- Harper, P. R., & Shahani, A. K. (2002). Modelling for the planning and management of bed capacities in hospitals. *Journal of the Operational Research Society*, 53(1), 11–18. <https://doi.org/10.1057/palgrave/jors/2601278>
- Heyburn, R., Bond, R. R., Black, M., Mulvenna, M., Wallace, J., Rankin, D., & Cleland, B. (2018). Machine learning using synthetic and real data: Similarity of evaluation metrics for different healthcare datasets and for different algorithms. *Data Science and Knowledge Engineering for Sensing Decision Support*, 11, 1281–1291. [https://doi.org/10.1142/9789813273238\\_0160](https://doi.org/10.1142/9789813273238_0160)
- Heydari, A., Sharifi, M., & Moghaddam, A. B. (2019). Challenges and barriers to providing care to older adult patients in the intensive care unit: A qualitative research. *Open Access Macedonian Journal of Medical Sciences*, 7(21), 3682–3690. <https://doi.org/10.3889/oamjms.2019.846>
- Humphreys, P., Spratt, B., Tariverdi, M., Burdett, R. L., Cook, D., Yarlagadda, P. K. D. V., & Corry, P. (2022). An overview of hospital capacity planning and optimisation. *Healthcare*, 10(5), 826. <https://doi.org/10.3390/healthcare10050826>
- Ius, T., Somma, T., Altieri, R., Angileri, F., Barbagallo, G., Cappabianca, P., Certo, F., Cofano, F., D'Elia, A., Della Pepa, G. M., Esposito, V., Fontanella, M. M., Germanò, A., Garbossa, D., Isola, M., La Rocca, G., Maiuri, F., Olivi, A., Panciani, P. P., Pignotti, F., ... Sabatino, A. (2020). Is age an additional factor in the treatment of elderly patients with glioblastoma? A new stratification model: An Italian multicenter study. *Neurosurgical Focus*, 49(4), E13. <https://doi.org/10.3171/2020.7.FOCUS20420>
- Kilincer, I. F., Ertam, F., Sengur, A., Tan, R.-S., & Acharya, U. R. (2023). Automated detection of cybersecurity attacks in healthcare systems with recursive feature elimination and multilayer perceptron optimization. *Biocybernetics and Biomedical Engineering*, 43(1), 30–41.
- Kuo, K.-M., Talley, P. C., Kuzuya, M., & Chi-Hsien, H. (2019). Development of a clinical support system for identifying social frailty. *International Journal of Medical Informatics*, 132, 103979. <https://doi.org/10.1016/j.ijmedinf.2019.103979>
- Kutafina, E., Bechtold, I., Kabino, K., & Jonas, S. M. (2019). Recursive neural networks in hospital bed occupancy forecasting. *BMC Medical Informatics and Decision Making*, 19(1), 39. <https://doi.org/10.1186/s12911-019-0776-1>
- Lam, F., Leung, J., & Kwok, T. (2019). The clinical potential of frailty indicators on identifying recurrent fallers in the community: The Mr. Os and Ms. OS cohort study in Hong Kong. *Journal of the American Medical Directors Association*, 20(12), 1605–1610. <https://doi.org/10.1016/j.jamda.2019.06.019>
- Levis, A. A., & Papageorgiou, L. G. (2004). A hierarchical solution approach for multi-site capacity planning under uncertainty in the pharmaceutical industry. *Computers & Chemical Engineering*, 28(5), 707–725. <https://doi.org/10.1016/j.compchemeng.2004.02.012>
- Liping, Z., Na, G., Zhibin, J., & Xiuxian, W. (2018). Multi-objective capacity allocation of hospital wards combining revenue and equity. *Omega*, 81, 220–233. <https://doi.org/10.1016/j.omega.2017.11.005>
- Lu, Y., Guan, Y., Zhong, X., Fische, J. N., Hogan, T. (2021). *Hospital beds planning and admission control policies for covid-19 pandemic: A hybrid computer simulation approach* [Paper presentation]. 2021 IEEE 17th International Conference on Automation Science and Engineering (Case) (pp. 956–961). <https://doi.org/10.1109/CASE49439.2021.9551589>
- Macnee, W., Rabinovich, R. A., & Choudhury, G. (2014). Ageing and the border between health and disease. *European Respiratory Journal*, 44(5), 1332–1352. <https://doi.org/10.1183/09031936.00134014>
- Maggioni, F., & Wallace, S. W. (2012). Analyzing the quality of the expected value solution in stochastic programming. *Annals of Operations Research*, 200(1), 37–54. <https://doi.org/10.1007/s10479-010-0807-x>
- Mandelbaum, A., Momčilović, P., Trichakis, N., Kadish, S., Leib, R., & Bunnell, C. A. (2020). Data-driven appointment-scheduling under uncertainty: The case of an infusion unit in a cancer center. *Management Science*, 66(1), 243–270. <https://doi.org/10.1287/mnsc.2018.3218>
- Marynissen, J., & Demeulemeester, E. (2019). Literature review on multi-appointment scheduling problems in hospitals. *European Journal of Operational Research*, 272(2), 407–419. <https://doi.org/10.1016/j.ejor.2018.03.001>
- Mestre, A. M., Oliveira, M. D., & Barbosa-Póvoa, A. P. (2015). Location-allocation approaches for hospital network planning under uncertainty. *European Journal of Operational Research*, 240(3), 791–806. <https://doi.org/10.1016/j.ejor.2014.07.024>
- Nasrabadi, A. M., Najafi, M., & Zolfagharinia, H. (2020). Considering short-term and long-term uncertainties in location and capacity planning of public healthcare facilities. *European Journal of Operational Research*, 281(1), 152–173. <https://doi.org/10.1016/j.ejor.2019.08.014>
- Passmore, P., Kailis, S., & Marshall, G. (2011). Identifying prediction factors for unplanned hospitalisation of an elderly population using classification and regression

tree (CART) analysis. *International Journal of Pharmacy Practice*, 2(2), 71–76. <https://doi.org/10.1111/j.2042-7174.1993.tb00728.x>

Proudlove, N. C., Black, S., & Fletcher, A. (2007). Or and the challenge to improve the NHS: Modelling for insight and improvement in in-patient flows. *Journal of the Operational Research Society*, 58(2), 145–158. <https://doi.org/10.1057/palgrave.jors.2602252>

Public Health Scotland. (2021). *Reports for financial year 2019 to 2020*. <https://beta.isdscotland.org/topics/finance/file-listings-fy-2019-to-2020/>

Restrepo, M. L., Rousseau, L.-M., & Vallée, J. (2020). Home healthcare integrated staffing and scheduling. *Omega*, 95, 102057. <https://doi.org/10.1016/j.omega.2019.03.015>

Segall, R. S. (1992). Deterministic mathematical modelling for the spatial allocation of multi-categorical resources: With an application to real health data. *Journal of the Operational Research Society*, 43(6), 579–589. <https://doi.org/10.1038/sj/jors/0430604>

Shehadeh, K. S., & Padman, R. (2022). Stochastic optimization approaches for elective surgery scheduling with downstream capacity constraints: Models, challenges, and opportunities. *Computers & Operations Research*, 137, 105523. <https://doi.org/10.1016/j.cor.2021.105523>

Thompson, S., Nunez, M., Garfinkel, R., & Dean, M. D. (2009). Or practice-efficient short-term allocation and reallocation of patients to floors of a hospital during demand surges. *Operations Research*, 57(2), 261–273. <https://doi.org/10.1287/opre.1080.0584>

United Nations. (2019). *World population ageing 2019*. <https://www.un.org/en/development/desa/population/publications/pdf/ageing/WorldPopulationAgeing2019-Highlights.pdf>

Watanabe, A., Ono, Q., Nishigami, T., & Takahiko Hirooka, H. M. (2018). Differences in risk factors for rotator cuff tears between elderly patients and young patients. *Acta Medica Okayama*, 72(1), 67–72. <https://doi.org/10.18926/AMO/55665>

Williams, E. (2023). *Bed and staff capacities planning models*. <https://github.com/Williamsem/Bed-and-Staff-Capacities-Planning-Models>

Williams, E., Gartner, D., & Harper, P. (2021). A survey of OR/MS models on care planning for frail and elderly patients. *Operations Research for Health Care*, 31, 100325. <https://doi.org/10.1016/j.orhc.2021.100325>

Xue, Q.-L. (2011). The frailty syndrome: Definition and natural history. *Clinics in Geriatric Medicine*, 27(1), 1–15. <https://doi.org/10.1016/j.cger.2010.08.009>

Yankovic, N., & Green, L. V. (2011). Identifying good nursing levels: A queuing approach. *Operations Research*, 59(4), 942–955. <https://doi.org/10.1287/opre.1110.0943>

Zanakis, S. H., & Becerra-Fernandez, I. (2005). Competitiveness of nations: A knowledge discovery examination. *European Journal of Operational Research*, 166(1), 185–211. <https://doi.org/10.1016/j.ejor.2004.03.028>

## Appendix A: List of patient attributes

**Table A1.** List of patient attributes within the dataset used for analysis.

Attribute	Data type	Distinct attribute values or bins	Documentation
Admission Date	Ordinal	1,096 (e.g., 01/04/2017)	Upon admission
Admission Method	Nominal	17 (e.g., Elective waiting list)	Upon admission
Admission Source	Nominal	26 (e.g., Usual place of residence)	Upon admission
Admission Time	Continuous	1,440 (hh:mm)}	Upon admission
Borough	Nominal	174 (e.g., Newport LHB, Monmouthshire LHB)	Upon admission
Date of Birth	Ordinal	12,037 (e.g., 01/01/1940)	Upon admission
Diagnosis	Nominal	2,758 (e.g., Fracture of neck of femur, Congestive heart failure)	Upon admission
Discharge Date	Ordinal	1,154 (e.g., 01/04/2017)	On discharge
Discharge Destination	Nominal	26 (e.g., Death, Own home, Patient transfer within same health board/trust)	On discharge
Discharge Time	Continuous	1,306 (hh:mm:ss)}	On discharge
Hospital	Nominal	14 (e.g., Chepstow Community Hospital)	Upon admission
NHS Number	Nominal	66,251 (e.g., 4900000000)	Upon admission
Postcode	Nominal	13,819 (e.g., CF72 8XR)	Upon admission
Registered GP	Nominal	1,313 (e.g., G9041668)	Upon admission
Registered GP Practice	Nominal	618 (e.g., W93012)	Upon admission
Scan Attendance Date	Ordinal	1,097 (e.g., 01/04/2017)	While admitted
Scan Attendance Time	Continuous	11,417 (hh:mm:ss)}	While admitted
Scan Exam	Nominal	293 (e.g., CT Neck and thorax)	While admitted
Scan Exam Code	Nominal	295 (e.g., XCHES, XABDO, CSKUH)	While admitted
Scan Procedure Code	Nominal	16 (e.g., R, CT, MR)	While admitted
Scan Requested Date	Ordinal	1,090 (e.g., 01/04/2017)	While admitted
Scan Specialty Code	Nominal	37 (e.g., Gastro, Neuro)	While admitted
Specialty	Nominal	30 (e.g., Care of the Elderly, Neurology)	Upon admission