Linking Predictive And Prescriptive Analytics For Modelling Healthcare Services For Frail And Elderly Patients

Elizabeth Mary Millicent Williams

School of Mathematics

Cardiff University

A thesis submitted for the degree of

Doctor of Philosophy

August 2023

Abstract

The National Health Service in the United Kingdom is under increasing pressure to provide and deliver high-quality care to an ageing population with complex health needs. This research project was funded by KESS2 in collaboration with the Aneurin Bevan University Health Board in South East Wales.

This thesis investigates the potential of using predictive and prescriptive analytics to optimise bed capacities and staffing requirements within Aneurin Bevan University Health Board. Homogeneous patient clusters are identified through the use of classification and regression trees to predict the length of stay of the frail and elderly population. Deterministic and two-stage stochastic optimisation models are developed to determine bed capacities and staffing requirements, taking into account factors such as patient acuity, length of stay, and resource constraints.

The predictive and prescriptive models are then combined by using the classification and regression tree models to determine demand values to be inputted into the deterministic and two-stage stochastic models. To determine the benefit and cost savings of using the stochastic implementation over traditional deterministic models, the value of the stochastic solution is calculated.

Through the application of scenario analysis, the methods allow various case studies to be modelled to provide insights into how the system would cope with fluctuations in resources, demand or organisational changes. The findings of this thesis have important implications for healthcare providers and policymakers, highlighting the potential for the combination of predictive and prescriptive analytics to improve the quality and efficiency of healthcare delivery. The study also provides a framework for future research in this area, including the potential for applying these techniques to other healthcare settings and populations.

Acknowledgements

Firstly, I would like to express my sincere gratitude to my supervisors, Prof. Daniel Gartner and Prof. Paul Harper for their guidance throughout my research journey. Your support and encouragement has not only helped me grow as a researcher, but also as a person. I could not have imagined having a better supervision team for my PhD.

I would also like to acknowledge the financial support provided by KESS2, along with the Aneurin Bevan University Health Board, which made this research possible. I would like to extend my thanks to the Clinical Futures Team at ABUHB, for contributing their time and knowledge in assisting with this research. I would like to especially thank Dr. Andrew Gray and Jacqui Thornton for your enthusiasm and involvement in this project.

I am grateful to the School of Mathematics staff and my fellow PhD students for creating a supportive community and for making my PhD thoroughly enjoyable. More specifically, I would like to thank Michalis for the endless support, for always having time to fix whatever I had managed to break and most importantly looking after my plant Trev. To Abby and Shauna for always providing a welcome distraction and for our weekly Tuesday lunchtime sessions. To Tim, Michela and Matthew, thank you for your willingness to give up your time to talk, play games and above all, to drink tea.

I would like to thank my friends and family for all your support over the last three and a half years. To my sisters, Sallie and Suzanne, you both inspire me every day, and I will forever be grateful for the love and encouragement you give me. Finally, to my mum; this thesis would not exist without you. Your unconditional love, belief and confidence in me, and words of wisdom is something I will treasure forever.

Dedicated to John Williams.

Dissemination

Publications

- A survey of OR/MS models on care planning for frail and elderly patients. E. Williams, D. Gartner and P. Harper. *Operations Research for Health Care*, 2021 [1]
- Prescriptive Healthcare Analytics: A Tutorial on Discrete Optimization and Simulation. D. Gartner, E. M. Williams and P. R. Harper. In *2022 IEEE 10th International Conference on Healthcare Informatics (ICHI)*, 2022 [2]
- Linking Predictive and Prescriptive Analytics of Elderly and Frail Patient Hospital Services, E. Williams, D. Gartner and P. Harper. In *2022 IEEE 10th International Conference on Healthcare Informatics (ICHI)*, 2022 [3]

Presentations

- Multi-site and Multi-service Modelling for Elderly and Frail Patients (Poster). ORAHS, Online via Video Conferencing, July 2020.
- Multi-site and Multi-service Modelling for Elderly and Frail Patients (Poster). 3^{rd} IMA and OR Society Conference, Online via Video Conferencing, April 2021.
- Multi-site and Multi-service Modelling for Elderly and Frail Patients (Poster). SIAM Student Chapter National Conference, Online via Video Conferencing, April 2021.
- Multi-site and Multi-service Modelling for Elderly and Frail Patients. INFORMS Healthcare, Online via Video Conferencing. July 2021.
- Predictive Analytics Applied to Hospital Services for Elderly and Frail Patients, University of South Wales Seminar Series, Online via Video Conferencing, January 2022.
- Using Predictive and Prescriptive Analytics Within Hospital Services for Elderly and Frail Patients. Welsh Mathematics Colloquium, Gregynog (Wales),

May 2022.

- Linking Predictive and Prescriptive Analytics for Healthcare Services: The Case of Frail and Elderly Patients. CORS/INFORMS International Conference, Vancouver (Canada), June 2022.
- Linking Predictive and Prescriptive Analytics for Healthcare Services: The Case of Frail and Elderly Patients. ICHI IEEE Doctoral Consortium, Rochester, Minnesota (USA), June 2022.
- Linking Predictive and Prescriptive Analytics for Healthcare Services: The Case of Frail and Elderly Patients. OR64, Warwick (UK), September 2022.

Contents

Abstract **i**

Acknowledgements **iii**

Dissemination **vii**

List of Figures **xiv**

List of Tables **xxiv**

List of Acronyms **xxxv**

1 Introduction **1**

1.1 Frail and Elderly Patient Demographics 1

1.2 Aneurin Bevan University Health Board 3

1.3 Bed Planning and Staff Allocation 5

1.4 Research Aims 8

1.5 Thesis Structure 9

2 Literature Review **11**

2.1 Introduction 11

2.2 A Survey of OR/MS Models on Care Planning for Frail and Elderly Patients 11

2.2.1 Methods 12

2.2.2 Results 17

2.2.3 Research Gaps 30

2.2.4 Literature Review Findings 33

2.3 Hierarchical Prediction Models for Elderly and Frail Patients' Lengths of Stay 34

2.3.1 Methods 35

2.3.2 Results 38

2.3.3 Research Gaps 47

2.3.4 Literature Review Findings 49

2.4 Overlap Between the Two Literature Reviews 50

2.5 Summary 51

3 Predictive Analytics to Support Capacity Planning for Frail and Elderly Patients 53

3.1 Introduction 53

3.2 Predictive Techniques 54

3.2.1 Differences Between Supervised and Unsupervised Learning 54

3.2.2 General Model Formulation 56

3.2.3 Evaluation Metrics 61

3.3 Classification and Regression Trees 67

3.3.1 Generalised Formulation 69

3.3.2 Extension of Random Forests 72

3.3.3 Feature Parameters 74

3.3.4 Evaluation Metrics 78

3.4 Python Development 80

3.4.1 Linear Regression 80

3.4.2 Logistic Regression 81

3.4.3 Regression Trees 83

3.4.4 Classification Trees 85

3.4.5 Random Forests - Regression 86

3.4.6 Random Forests - Classification 87

3.5 Summary 87

4 Prescriptive Analytics to Support Capacity Planning for Frail and Elderly Patients **89**

4.1 Introduction 89

4.2 Prescriptive Techniques 91

4.2.1 Basic Notation 92

4.2.2 Evaluation of Measures 94

4.3 Deterministic Model Development 96

4.3.1 Sets 97

4.3.2 Parameters 97

4.3.3 Decision Variables 97

4.3.4 Model 97

4.3.5 Worked Example 99

4.4 Two-Stage Stochastic Model Extension 101

4.4.1 Sets 102

4.4.2 Parameters 102

4.4.3 Decision Variables 103

4.4.4 Model 103

4.4.5 Worked Example 104

4.5 Evaluation of Measures 108

4.5.1 Test A 108

4.5.2 Test B 109

4.5.3 Test C 110

4.5.4 Worked Example Conclusion 111

4.6 Summary 113

5 Experimental Analysis of the Predictive and Prescriptive Models **115**

5.1 Introduction 115

5.2 Data Introduction 116

5.2.1 Data Trends 118

5.2.2 Data Insights 123

5.3 Predictive Analytics Results 124
5.3.1 Linear Regression 124
5.3.2 Logistic Regression 126
5.3.3 Classification and Regression Trees 128
5.3.4 Predictive Analytics Summary 143
5.4 Prescriptive Analytics Results 144
5.4.1 Model Data 145
5.4.2 Model Development 149
5.4.3 Results 150
5.5 Summary 167

6 Linking Predictive and Prescriptive Analytics for Healthcare Services 169
6.1 Introduction 169
6.2 Linking Paradigms 170
6.2.1 Regression Tree and Average LOS 171
6.2.2 Regression Tree and Specific LOS 173
6.2.3 Classification Tree and Average LOS 175
6.2.4 Classification Tree and Specific LOS 178
6.3 Scenario Analysis 180
6.3.1 Addition of a New Hospital 181
6.3.2 M-Penalty 183
6.3.3 Re-evaluating the Current Setup 185
6.3.4 Long-term Planning 186
6.4 Generalisability of Results 191
6.5 Summary 192

7 Decision Support Tool for Multi-Hospital Planning of Frail and Elderly Resource Capacities 195
7.1 Introduction 195
7.2 Excel Implementation 195

7.2.1 Deterministic Model 199

7.2.2 Two-Stage Stochastic Model 202

7.2.3 Test A 203

7.2.4 Test B 205

7.2.5 Test C 207

7.3 Python Implementation 209

7.3.1 Deterministic Model 209

7.3.2 Two-Stage Stochastic Model 214

7.3.3 Test A 220

7.3.4 Test B 226

7.3.5 Test C 232

8 Conclusions 239

8.1 Research Summary 239

8.2 Research Contributions 241

8.3 Limitations of the Study 242

8.4 Impact in Practice 243

8.5 Future Work 244

Appendices 245

A List of Hospitals and Specialties Within ABUHB 245

B Literature Review Supplementary Material 249

B.1 Application of OR/MS Methods to Frail and Elderly Healthcare - Figures and Tables (Chapter 2.2) 249

B.2 Hierarchical Prediction Models for Patients' Lengths of Stay - Tables (Chapter 2.3). 254

C Experimental Analysis Further Material 259

C.1 ABUHB Admission and Scan Data - Attributes 259

C.2 Regression Tree Groupings 270

C.3 Classification Tree Groupings 271

C.4 Deterministic Parameters 273

C.5 Two-Stage Stochastic Parameters 276

D Linking Predictive and Prescriptive Paradigms Further Material 283

D.1 Regression Trees - Average LOS 283

D.2 Regression Trees - Specific LOS 294

D.3 Classification Trees - Average LOS 300

D.4 Classification Trees - Specific LOS 310

D.5 Scenario Heatmaps 316

Bibliography 321

List of Figures

1.1 Observed population pyramids for the years 2011 (a) and 2021 (b), as well as the expected projected population pyramid for the year 2031 (c). The graphs were generated by data gathered from [4] and [5]. 2
1.2 Aneurin Bevan University Health Board (ABUHB). 4
1.3 Clinical Futures future healthcare plan, where specialist hospital care is provided in one central location, with emphasis being put on care closer to home and healthy lifestyle choices [6]. 5
1.4 Map of hospital locations in ABUHB. Major A&E units are shown in red, MIU's are shown in blue and acute and community hospitals are shown in purple. 6
1.5 Hospitals and specialties in ABUHB. A '1' indicates a specialty is present in a given hospital and '0' indicates a specialty is not present. 6
1.6 Conceptual overview of the patient journey and the role of bed management. Patients arrive into the system via elective and emergency routes and are assigned a bed. Patients remain in hospital for treatment, and are discharged when are classified as 'medically fit' [7]. 7
1.7 Common reasons for delayed discharge for patients with a length of stay (LOS) longer than a week within English hospitals [8, 9]. . 8

2.1 The PRISMA flow diagram for the care planning literature review detailing the Scopus searches, the number of abstracts screened, the reason for exclusion and the number of full texts retrieved. . . 15
2.2 Number of publications broken down by JCR categories and Scopus search type. 18
2.3 Bar chart of the 62 papers identified through the Scopus search, by publication year. 19
2.4 Venn diagram of the 62 papers identified through the Scopus search, by medical setting. 20

2.5 Bar chart of the 62 papers identified through the Scopus search, by 'Acute' or 'Chronic' medical conditions. 22
2.6 Cross analysis of medical setting and condition area of the 62 papers identified through the Scopus search. 23
2.7 Bar chart of the 62 papers identified through the Scopus search, by research focus. 24
2.8 Bar chart of the 62 papers identified through the Scopus search, by planning decision level. 25
2.9 Cross analysis of planning decision and medical setting of the 62 papers identified through the Scopus search. 26
2.10 Cross analysis of JCR category and planning decision level of the 62 papers identified through the Scopus search. 27
2.11 Bar chart of the 62 papers identified through the Scopus search, by mathematical method. 28
2.12 Bar chart of the 17 papers using a Markov model identified through the Scopus search, classified by 'Continuous' or 'Discrete' time. . . 30
2.13 The PRISMA flow diagram for the hierarchical prediction literature review detailing the Scopus searches, the number of abstracts screened, the reason for exclusion and the number of full texts retrieved. 37
2.14 Bar chart of the 90 papers identified through the Scopus search, by publication year. 39
2.15 Venn diagram of the 90 papers identified through the Scopus search, by medical setting. 41
2.16 Bar chart of the 90 papers identified through the Scopus search, by 'Acute', 'Chronic', and 'Surgical' medical conditions. 41
2.17 Bar chart of the 90 papers identified through the Scopus search, by research focus. 42
2.18 Bar chart of the 90 papers identified through the Scopus search, by planning decision level. 43
2.19 Cross analysis of planning decision and medical setting of the 90 papers identified through the Scopus search. 44
2.20 Bar chart of the 90 papers identified through the Scopus search, by mathematical method. 45

3.1 Worked example of linear regression model where the dependent variable y represents the Length of Stay (LOS) and the independent variable x is the age of the patient. 59
3.2 ROC curve for continuous age and LOS prediction. 67
3.3 ROC curve for categorical hospital and LOS prediction. 67

3.4 Example of a decision tree with one root node and two terminal nodes. 68
3.5 Example of a decision tree with one root node, three decision nodes and five terminal nodes. 69
3.6 Regression tree of the first worked example consisting of four terminal nodes. Note that a darker colour of the node indicates a longer LOS. 71
3.7 Classification tree for second worked example consisting of four terminal nodes. Note that blue indicates the class of ≥ 4 days, and orange indicates the class of < 4 days. 72
3.8 Generic example of a random forest consisting of n decision trees and n training data sets. 73
3.9 Excerpt of the OLS results for continuous linear regression showing the R^2 value, the coefficient values, and their corresponding p-values. 81
3.10 Excerpt of the OLS results for categorical linear regression showing the R^2 value, the coefficient values, and their corresponding p-values. 81
3.11 Excerpt of the Logit regression results for continuous logistic regression showing the pseudo R^2 value, the coefficient values, and their corresponding p-values. 83
3.12 Excerpt of the Logit regression results for categorical logistic regression showing the pseudo R^2 value, the coefficient values, and their corresponding p-values. 84

4.1 Line graph showing the objective function for the deterministic implementation against the different daily demand levels, where -10% indicated the low daily bed demand and +10% indicates the high daily bed demand. 101
4.2 Line graph showing the objective function for the two-stage stochastic implementation against the different daily demand levels, where -10% indicates the low daily bed demand and +10% indicated the high daily bed demand. 108
4.3 Clustered bar chart showing the objective function value for the deterministic and two-stage stochastic models. Included within this is the EEV, ESSV and EIV results for the low, average and high daily bed demands. 112

5.1 Flow chart of data cleansing process resulting in 165,118 patient admissions and 15,483 patient scans within ABUHB for the period April 2017 to March 2020. 118

5.2 Line graph showing the trend of the number of patients admitted per day into ABUHB, split by fiscal year. 119
5.3 Bar chart showing the number of patients within the ABUHB data set, falling into each of the seven age groups. 120
5.4 Line graph showing the number of admissions per day by hourly interval, for the ABUHB data set. 121
5.5 Bar chart showing the number of patients undertaking each type of scan within the ABUHB data set. 123
5.6 Regression tree with an R^2 of 34.23% for predicting continuous LOS for patients within ABUHB, consisting of 30 terminal nodes. 135
5.7 Classification tree with an accuracy score of 89.89% for predicting grouped LOS for patients within ABUHB, consisting of 30 terminal nodes. 141
5.8 Heatmap of bed locations for each specialty within each hospital for the deterministic model for Experiment 1. Note that a darker colour indicates a larger number of beds are deployed. 152
5.9 Heatmap of bed locations for each specialty within each hospital for the two-stage stochastic model for Experiment 1. Note that a darker colour indicates a larger number of beds are deployed. . . . 153
5.10 Heatmaps of bed locations for each specialty within each hospital for the deterministic model for the years 2017-2018 (a), 2018-2019 (b) and 2019-2020 (c), for Experiment 2. Note that a darker colour indicates a larger number of beds are deployed. 161
5.11 Heatmaps of bed locations for each specialty within each hospital for the two-stage stochastic model for the years 2017-2018 (a), 2018-2019 (b) and 2019-2020 (c), for Experiment 2. Note that a darker colour indicates a larger number of beds are deployed. . . . 163

6.1 An updated version of the hospitals and specialties in ABUHB, with a '1' indicating a specialty is present in a given hospital. For cells with a red background, displays where specialties have opened or closed since the opening of the Grange University Hospital (GUH). 182

7.1 The data requirements for the Excel OpenSolver plugin where the user is required to enter, as a minimum, the 'Specialty', 'Hospital' and either 'Short LOS' or 'LOS_hours' for each patient. 196
7.2 The average daily bed demand matrix automatically generated by Excel, segmented by region and specialty. The user is required to enter the number of days within the data. 197

7.3 The maximum number of beds that can be deployed to each hospital location for each specialty. Additionally, the user is required to enter the total capacity for the hospitals in the first and second stages. 198
7.4 The user is required to enter the first stage cost for each hospital and specialty combination. 198
7.5 The user is required to enter the ratio of nursing band staff to each specialty. Additionally, the user is required to enter the cost per hour of the first and second stage nurses staff, and the total capacity. 199
7.6 The OpenSolver objective cell, decision variable cells and constraints for the deterministic implementation. The 'Current Solver Engine' can be changed to Gurobi if a license is available. 200
7.7 The output from the deterministic model once solved, displaying the number of beds to deploy to each hospital and specialty. The total daily cost, beds and staff are also summarised. 201
7.8 The staffing output from the deterministic model once solved, displaying the number of nursing staff to deploy to each specialty and of which band level. 202
7.9 The scenario selector for the two-stage stochastic implementation. Users are prompted to input the demand scalar and the corresponding probability of occurrence. 202
7.10 The OpenSolver objective cell, decision variables and constraints for the two-stage stochastic implementation. The 'Current Solver Engine' can be changed to Gurobi if a license is available. 203
7.11 The output from the two-stage stochastic model once solved, displayed the number of beds to deploy to each hospital and specialty. The total daily cost, first and second stage beds and nursing staff are also summarised. 204
7.12 The output from Test A model implementation once solved, displayed the number of beds to deploy to each hospital and specialty. The total daily cost, first and second stage beds and nursing staff are also summarised. 204
7.13 The additional constraint required for Test B, where the zero variables within the deterministic model are set as the lower bound for the first stage within Test B. 205
7.14 For Test B, an 'if statement' is required to be used to determine the zero bound of the decision variables from the Deterministic output. As OpenSolver is a linear programming solver, the user is required to copy and paste the values only from cells `AB3:AM31` to `O3:Z31`. 206

7.15 The output from Test B model implementation once solved, displayed the number of beds to deploy to each hospital and specialty. The total daily cost, first and second stage beds and nursing staff are also summarised. 207
7.16 The minimum values which are required to be met within the first stage of Test C. This will determine the upgradeability of the model. 208
7.17 The additional constraint required for Test C, where the deterministic values are the minimum bound for the first stage of Test C. 208
7.18 The output from the deterministic model once solved, displaying the number of beds to deploy to each hospital and specialty. The total daily cost, beds and staff are also summarised. 208

D.1 Heatmap of bed locations for each specialty within each hospital for the deterministic model using the regression tree and average LOS over three years' worth of data. 284
D.2 Heatmap of bed locations for each specialty within each hospital for the two-stage stochastic model using the regression tree and average LOS over three years' worth of data. 284
D.3 Heatmap of bed locations for each specialty within each hospital for the deterministic model using the regression tree and average LOS for 2017-2018. 286
D.4 Heatmap of bed locations for each specialty within each hospital for the two-stage stochastic model using the regression tree and average LOS for 2017-2018. 286
D.5 Heatmap of bed locations for each specialty within each hospital for the deterministic model using the regression tree and average LOS for 2018-2019. 287
D.6 Heatmap of bed locations for each specialty within each hospital for the two-stage stochastic model using the regression tree and average LOS for 2018-2019. 287
D.7 Heatmap of bed locations for each specialty within each hospital for the deterministic model using the regression tree and average LOS for 2019-2020. 287
D.8 Heatmap of bed locations for each specialty within each hospital for the two-stage stochastic model using the regression tree and average LOS for 2019-2020. 287
D.9 Heatmap of bed locations for each specialty within each hospital for the deterministic model using the regression tree and average year LOS for 2017-2018. 290

D.10 Heatmap of bed locations for each specialty within each hospital for the two-stage stochastic model using the regression tree and average year LOS for 2017-2018. 290
D.11 Heatmap of bed locations for each specialty within each hospital for the deterministic model using the regression tree and average year LOS for 2018-2019. 291
D.12 Heatmap of bed locations for each specialty within each hospital for the two-stage stochastic model using the regression tree and average year LOS for 2018-2019. 291
D.13 Heatmap of bed locations for each specialty within each hospital for the deterministic model using the regression tree and average year LOS for 2019-2020. 291
D.14 Heatmap of bed locations for each specialty within each hospital for the two-stage stochastic model using the regression tree and average year LOS for 2019-2020. 291
D.15 Heatmap of bed locations for each specialty within each hospital for the deterministic model using the regression tree and specific LOS over three years' worth of data. 295
D.16 Heatmap of bed locations for each specialty within each hospital for the two-stage stochastic model using the regression tree and specific LOS over three years' worth of data. 295
D.17 Heatmap of bed locations for each specialty within each hospital for the deterministic model using the regression tree and specific LOS for 2017-2018. 298
D.18 Heatmap of bed locations for each specialty within each hospital for the two-stage stochastic model using the regression tree and specific LOS for 2017-2018. 298
D.19 Heatmap of bed locations for each specialty within each hospital for the deterministic model using the regression tree and specific LOS for 2018-2019. 299
D.20 Heatmap of bed locations for each specialty within each hospital for the two-stage stochastic model using the regression tree and specific LOS for 2018-2019. 299
D.21 Heatmap of bed locations for each specialty within each hospital for the deterministic model using the regression tree and specific LOS for 2019-2020. 299
D.22 Heatmap of bed locations for each specialty within each hospital for the two-stage stochastic model using the regression tree and specific LOS for 2019-2020. 299

D.23 Heatmap of bed locations for each specialty within each hospital for the deterministic model using the classification tree and average LOS over three years' worth of data. 301
D.24 Heatmap of bed locations for each specialty within each hospital for the two-stage stochastic model using the classification tree and average LOS over three years' worth of data. 301
D.25 Heatmap of bed locations for each specialty within each hospital for the deterministic model using the classification tree and average LOS for 2017-2018. 304
D.26 Heatmap of bed locations for each specialty within each hospital for the two-stage stochastic model using the classification tree and average LOS for 2017-2018. 304
D.27 Heatmap of bed locations for each specialty within each hospital for the deterministic model using the classification tree and average LOS for 2018-2019. 305
D.28 Heatmap of bed locations for each specialty within each hospital for the two-stage stochastic model using the classification tree and average LOS for 2018-2019. 305
D.29 Heatmap of bed locations for each specialty within each hospital for the deterministic model using the classification tree and average LOS for 2019-2020. 305
D.30 Heatmap of bed locations for each specialty within each hospital for the two-stage stochastic model using the classification tree and average LOS for 2019-2020. 305
D.31 Heatmap of bed locations for each specialty within each hospital for the deterministic model using the classification tree and average year LOS for 2017-2018. 308
D.32 Heatmap of bed locations for each specialty within each hospital for the two-stage stochastic model using the classification tree and average year LOS for 2017-2018. 308
D.33 Heatmap of bed locations for each specialty within each hospital for the deterministic model using the classification tree and average year LOS for 2018-2019. 309
D.34 Heatmap of bed locations for each specialty within each hospital for the two-stage stochastic model using the classification tree and average year LOS for 2018-2019. 309
D.35 Heatmap of bed locations for each specialty within each hospital for the deterministic model using the classification tree and average year LOS for 2019-2020. 309

D.36 Heatmap of bed locations for each specialty within each hospital for the two-stage stochastic model using the classification tree and average year LOS for 2019-2020. 309
D.37 Heatmap of bed locations for each specialty within each hospital for the deterministic model using the classification tree and specific LOS over three years' worth of data. 311
D.38 Heatmap of bed locations for each specialty within each hospital for the two-stage stochastic model using the classification tree and specific LOS over three years' worth of data. 311
D.39 Heatmap of bed locations for each specialty within each hospital for the deterministic model using the classification tree and specific LOS for 2017-2018. 314
D.40 Heatmap of bed locations for each specialty within each hospital for the two-stage stochastic model using the classification tree and specific LOS for 2017-2018. 314
D.41 Heatmap of bed locations for each specialty within each hospital for the deterministic model using the classification tree and specific LOS for 2018-2019. 315
D.42 Heatmap of bed locations for each specialty within each hospital for the two-stage stochastic model using the classification tree and specific LOS for 2018-2019. 315
D.43 Heatmap of bed locations for each specialty within each hospital for the deterministic model using the classification tree and specific LOS for 2019-2020. 315
D.44 Heatmap of bed locations for each specialty within each hospital for the two-stage stochastic model using the classification tree and specific LOS for 2019-2020. 315
D.45 Heatmap of bed locations for each specialty within each hospital for the deterministic model for Scenario 1 where GUH is added. . 316
D.46 Heatmap of bed locations for each specialty within each hospital for the two-stage stochastic model for Scenario 1 where GUH is added. 316
D.47 Heatmap of bed locations for each specialty within each hospital for the deterministic model for Scenario 2 where the M-penalty method is added. 317
D.48 Heatmap of bed locations for each specialty within each hospital for the two-stage stochastic model for Scenario 2 where the M-penalty method is added. 317

D.49 Heatmap of bed locations for each specialty within each hospital for the deterministic model for Scenario 3 where the hospital setup is re-evaluated. 317
D.50 Heatmap of bed locations for each specialty within each hospital for the two-stage stochastic model for Scenario 3 where the hospital setup is re-evaluated. 317
D.51 Heatmap of bed locations for each specialty within each hospital for the deterministic model for Scenario 4 where the nursing capacity is reduced. 318
D.52 Heatmap of bed locations for each specialty within each hospital for the two-stage stochastic model for Scenario 4 where the nursing capacity is reduced. 318
D.53 Heatmap of bed locations for each specialty within each hospital for the deterministic model for Scenario 5 with the introduction of virtual wards. 318
D.54 Heatmap of bed locations for each specialty within each hospital for the two-stage stochastic model for Scenario 5 with the introduction of virtual wards. 318
D.55 Heatmap of bed locations for each specialty within each hospital for the deterministic model for Scenario 6 with the sudden increase in demand. 319
D.56 Heatmap of bed locations for each specialty within each hospital for the two-stage model for Scenario 6 with the sudden increase in demand. 319
D.57 Heatmap of bed locations for each specialty within each hospital for the deterministic model for Scenario 7 with an overall increase in T&O services of 10%. 319
D.58 Heatmap of bed locations for each specialty within each hospital for the two-stage stochastic model for Scenario 7 with an overall increase in T&O services of 10%. 319
D.59 Heatmap of bed locations for each specialty within each hospital for the deterministic model for Scenario 7 with targetting T&O specific nodes in the regression tree with a 10% increase. 320
D.60 Heatmap of bed locations for each specialty within each hospital for the two-stage stochastic model for Scenario 7 with targetting T&O specific nodes in the regression tree with a 10% increase. . . 320

List of Tables

1.1 S42.201A - Example ICD10 code that shows the categorisation of each sublevel of ICD10. The first three characters provide a high level description of the category, where each additional character increases the diagnostic complexity [10]. 3

1.2 Type and names of the 11 hospitals located in ABUHB. 5

2.1 Scopus search string terms for the care planning literature review to include at least one 'OR Method', one 'Patient Flow' term and one 'Classification of People' term. 13

2.2 Number of papers which fall into each JCR category and the location of where the research was conducted. 18

2.3 Number of papers which fall into each medical setting and OR/MS method within the published research. 29

2.4 Scopus search terms for hierarchical prediction literature search to include at least one 'Prediction Method', one 'Length of Stay' term and one 'Classification of People' term 36

2.5 Number of papers which fall into each JCR category and the location of the where the research was conducted. 39

2.6 Number of papers which fall into each medical setting and OR/MS method within the published research. 46

3.1 Supervised learning methods and their applications across healthcare. 55

3.2 Unsupervised learning methods and their applications across healthcare. 55

3.3 Data set of 10 patients that will be used for the first worked example. 58

3.4 Data set of 15 patients that will be used for the second worked example. 60

3.5 General form of a confusion matrix consisting of actual values against predicted values. 64

3.6 Confusion matrix for the logistic regression model used in the second worked example with the data set of 15 patients. 65

3.7 'DecisionTreeRegressor' and 'DecisionTreeClassifier' associated parameters and attributes the user can select. The default variable is highlighted in blue. For variables which are an integer, float or there is no default, this is listed in brackets. 76
3.8 'RandomForestRegressor' and 'RandomForestClassifier' associated parameters and attributes the user can select. The default variable is highlighted in blue. For variables which are an integer, float or there is no default, this is listed in brackets. 77

4.1 Data set of 10 patients that will be used for the worked example. 91
4.2 The sets used within the deterministic model where (B, S, H, R) represent the maximum number of nursing bands, specialties, hospitals and regions, respectively. 97
4.3 The parameters used within the deterministic model where b, s, h, r represent the variable number of nursing bands, specialties, hospitals and regions, respectively. 98
4.4 The decision variables used within the deterministic model where b, s, h represent the variable number of nursing bands, specialties and hospitals, respectively. 98
4.5 The parameter values that will be used within the deterministic model specifically for the worked example. 99
4.6 The daily bed demand figures for each region and specialty ($D_{s,r}$), categorised by low, average and high. 100
4.7 The deterministic results for each of the decision variables x^{bed} and x^{staff} categorised by the low, average and high daily bed demands. 100
4.8 The sets used within the two-stage stochastic model where (B, S, H, R, K) rep- resent the maximum number of nursing bands, specialties, hospitals, regions and scenarios, respectively 102
4.9 The parameters 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 102
4.10 The sets used within the two-stage stochastic model where (b, s, h, r, k) represent the value number of nursing bands, specialties, hospitals, regions and scenarios, respectively 103
4.11 The parameter values that will be used within the two-stage stochastic model specifically for the worked example 105
4.12 The daily bed demand figures for each region and specialty $D_{s,r,k}$, categorised by low, average and high. 105

4.13 The two-stage stochastic results for each of the decision variables x^{bed}, x^{staff}, u^{bed} and x^{staff}, categorised by low, average and high daily bed demands. 107
4.14 Test A results for the worked example using low daily bed demand values, with results recorded in the form [(beds), (staff)]. 109
4.15 Test A results for the worked example using average daily bed demand values, with results recorded in the form [(beds), (staff)]. 109
4.16 Test A results for the worked example using high daily bed demand values, with results recorded in the form [(beds), (staff)]. 109
4.17 Test B results for the worked example using low daily bed demand values, with results recorded in the form [(beds), (staff)]. 110
4.18 Test B results for the worked example using average daily bed demand values, with results recorded in the form [(beds), (staff)]. 110
4.19 Test B results for the worked example using high daily bed demand values, with results recorded in the form [(beds), (staff)]. 111
4.20 Test C results for the worked example using low daily bed demand values, with results recorded in the form [(beds), (staff)]. 111
4.21 Test C results for the worked example using average daily bed demand values, with results recorded in the form [(beds), (staff)]. 111
4.22 Test C results for the worked example using high daily bed demand values, with results recorded in the form [(beds), (staff)]. 112

5.1 Definitions for each of the data items within the merged ABUHB data set. 117
5.2 The total number of readmissions/transfers for the 9,332 care spell episodes within the ABUHB data set. Number of patient transfers in ABUHB. 119
5.3 The top 11 admitting specialties by count within the ABUHB data set. 120
5.4 The top six admitting methods by count within the ABUHB data set. 121
5.5 The top seven admission methods by count within the ABUHB data set. 121
5.6 The mean and standard deviation of patient LOS in hospital by the admission day. 122
5.7 The total number of scans for patients who had a scan during their inpatient admission within the ABUHB data set. 123
5.8 The results for the linear regression when run against the continuous and categorical variables within the ABUHB data set. The R^2 and Adjusted R^2 values are given. 124

5.9 The linear regression results for the specialty variable, where each x has a corresponding LOS value. 126
5.10 The results for the logistic regression when run against the continuous and categorical variables within the ABUHB data set. The accuracy, precision, recall and F1 scores are given. 127
5.11 The log odds ratios for age group category after running the logistic regression model. The odds compared to the 65-69 age group are also provided. 128
5.12 The parameters used within the regression tree model using the 'DecisionTreeRegressor' algorithm within Python. 129
5.13 The regression tree results for the R^2 (a) and computational time (b) for continuous age and no frailty. 131
5.14 The regression tree results for the R^2 (a) and computational time (b) for continuous age and continuous frailty. 131
5.15 The regression tree results for the R^2 (a) and computational time (b) for continuous age and grouped frailty. 132
5.16 The regression tree results for the R^2 (a) and computational time (b) for grouped age and no frailty. 132
5.17 The regression tree results for the R^2 (a) and computational time (b) for grouped age and continuous frailty. 133
5.18 The regression tree results for the R^2 (a) and computational time (b) for grouped age and grouped frailty. 133
5.19 The parameters used within the classification tree using the 'DecisionTreeClassifier' algorithm within Python. 136
5.20 The classification tree results for accuracy score (a), precision score (b), recall score (c) and computational time (d) for continuous age and no frailty. 137
5.21 The classification tree results for accuracy score (a), precision score (b), recall score (c) and computational time (d) for continuous age and continuous frailty. 138
5.22 The classification tree results for accuracy score (a), precision score (b), recall score (c) and computational time (d) for continuous age and grouped frailty. 138
5.23 The classification tree results for accuracy score (a), precision score (b), recall score (c) and computational time (d) for grouped age and no frailty. 139
5.24 The classification tree results for accuracy score (a), precision score (b), recall score (c) and computational time (d) for grouped age and continuous frailty. 139

5.25 The classification tree results for accuracy score (a), precision score (b), recall score (c) and computational time (d) for grouped age and grouped frailty. 139
5.26 The parameters used within the random forests using the 'RandomForestRegressor' and 'RandomForestClassifier' algorithms within Python. 142
5.27 The accuracy and R^2 score for the regression and classification random forest models, measured against the n_estimator parameter, respectively. Each model's computational runtime was additionally provided. 142
5.28 List of the 14 care locations within ABUHB and their associated regions. 145
5.29 Specialty cost data displaying the minimum, weighted average and the maximum daily bed cost per specialty, taken from NHS Scotland [11]. 148
5.30 Hourly nursing staffing costs per nursing band level within the NHS [12, 13]. 148
5.31 The daily bed demands for each specialty grouped by regions within ABUHB for three years' worth of patient admissions, rounded to four decimal places. 151
5.32 The three scenarios and their associated probabilities of occurring which will be used within the two-stage stochastic model. 152
5.33 The EV and RP values for the x^{bed}, x^{staff}, u^{bed} and u^{staff} decision variables and the objective function value for Experiment 1. . . . 153
5.34 The EV, RP and EEV values for the x^{bed}, x^{staff}, u^{bed} and u^{staff} decision variables and the objective function value for Experiment 1. 154
5.35 The EV, RP, EEV and ESSV values for the x^{bed}, x^{staff}, u^{bed} and u^{staff} decision variables and the objective function value for Experiment 1. 154
5.36 The EV, RP, EEV, ESSV and EIV values for the x^{bed}, x^{staff}, u^{bed} and u^{staff} decision variables and the objective function value for Experiment 1. 155
5.37 The daily bed demands for each specialty within regions one, two and three of ABUHB for three individual years' worth of patient admissions, rounded to four decimal places. 157
5.38 The daily bed demands for each specialty within regions four, five and six of ABUHB for three individual years' worth of patient admissions, rounded to four decimal places. 158
5.39 The EV values for the x^{bed}, u^{bed}, x^{staff} and u^{staff} decision variables and the objective function value per year for Experiment 2. 159

5.40 The EV and RP values for the x^{bed}, u^{bed}, x^{staff} and u^{staff} decision variables and the objective function value per year for Experiment 2. 161
5.41 The EV, RP and EEV values for the x^{bed}, u^{bed}, x^{staff} and u^{staff} decision variables and the objective function value per year for Experiment 2. 163
5.42 The EV, RP, EEV and ESSV values for the x^{bed}, u^{bed}, x^{staff} and u^{staff} decision variables and the objective function value per year for Experiment 2. 164
5.43 The EV, RP, EEV, ESSV and EIV values for the x^{bed}, u^{bed}, x^{staff} and u^{staff} decision variables and the objective function value per year for Experiment 2. 165
5.44 The total yearly VSS values for each of the three years by experiment. Note that year 3 was a leap year and therefore contains 366 days which accounts for the variation between years. 166

6.1 The count of admissions and the associated average LOS for each hospital and specialty within the second node of the regression tree. The average daily bed demand has additionally been calculated. . 171
6.2 The EV, RP and EEV values for the x^{bed}, u^{bed}, x^{staff}, u^{staff} decision variables and objective function using the regression tree and the average LOS across all three years. 172
6.3 The EV, RP and EEV values for the x^{bed}, u^{bed}, x^{staff}, u^{staff} decision variables and objective function using the regression tree and the yearly average LOS. 173
6.4 The EV, RP and EEV values for the x^{bed}, u^{bed}, x^{staff}, u^{staff} decision variables and objective function using the regression tree and the yearly average LOS for each hospital and specialty. 173
6.5 The count of admissions and the associated specific LOS for each hospital and specialty within the second node of the regression tree. The average daily bed demand has additionally been calculated. . 174
6.6 The EV, RP and EEV values for the x^{bed}, u^{bed}, x^{staff}, u^{staff} decision variables and objective function using the regression tree and the specific LOS across all three years. 174
6.7 The EV, RP and EEV values for the x^{bed}, u^{bed}, x^{staff}, u^{staff} decision variables and objective function using the regression tree and the yearly specific LOS. 175
6.8 The count of admissions and the associated average LOS for each hospital and specialty within the ninth node of the classification tree. The average daily bed demand has additionally been calculated. 176

6.9 The EV, RP and EEV values for the x^{bed}, u^{bed}, x^{staff}, u^{staff} decision variables and objective function using the classification tree and the average LOS across all three years. 176
6.10 The EV, RP and EEV values for the x^{bed}, u^{bed}, x^{staff}, u^{staff} decision variables and objective function using the classification tree and the yearly average LOS. 177
6.11 The EV, RP and EEV values for the x^{bed}, u^{bed}, x^{staff}, u^{staff} decision variables and objective function using the classification tree and the yearly average LOS for each hospital and specialty. 178
6.12 The count of admissions and the associated specific LOS for each hospital and specialty within the ninth node of the classification tree. The average daily bed demand has additionally been calculated. 179
6.13 The EV, RP and EEV values for the x^{bed}, u^{bed}, x^{staff}, u^{staff} decision variables and objective function using the classification tree and the specific LOS across all three years. 179
6.14 The EV, RP and EEV values for the x^{bed}, u^{bed}, x^{staff}, u^{staff} decision variables and objective function using the classification tree and the yearly specific LOS. 180
6.15 The EV, RP and EEV values for the x^{bed}, u^{bed}, x^{staff}, u^{staff} decision variables and objective function for Scenario 1 where the hospital GUH is added. 183
6.16 The EV, RP and EEV values for the x^{bed}, u^{bed}, x^{staff}, u^{staff} decision variables and objective function for Scenario 2 where the M-penalty method is added. 185
6.17 The EV, RP and EEV values for the x^{bed}, u^{bed}, x^{staff}, u^{staff} decision variables and objective function for Scenario 3 where the hospital setup is re-evaluated. 186
6.18 The EV, RP and EEV values for the x^{bed}, u^{bed}, x^{staff}, u^{staff} decision variables and objective function for Scenario 4 where the nursing capacity is reduced. 187
6.19 The EV, RP and EEV values for the x^{bed}, u^{bed}, x^{staff}, u^{staff} decision variables and objective function for Scenario 5 with the introduction of virtual wards. 189
6.20 The EV, RP and EEV values for the x^{bed}, u^{bed}, x^{staff}, u^{staff} decision variables and objective function for Scenario 6 with the sudden increase in demand. 189

6.21 The EV, RP and EEV values for the x^{bed}, u^{bed}, x^{staff}, u^{staff} decision variables and objective function for Scenario 7 with a 10% increase in demand for T&O services. Example 1 is the case where the overall average demand is increased by 10% and Example 2 is targeting T&O only nodes within the regression tree and increasing the demand by 10%. 191

A.1 Hospital and Specialty Locations in ABUHB 247

B.1a Summary of papers identified through the Scopus search in the first literature review, with a total of 62 papers. 250
B.1b Summary of papers identified through the Scopus search in the first literature review, with a total of 62 papers. 251
B.2 Number of papers that fall into each JCR category for the first literature review. 252
B.3 Number of papers that fall into each hospital setting for the first literature review. 252
B.4 Number of papers that fall into each condition area for the first literature review. 252
B.5 Number of Markov model papers identified in the first literature review. 252
B.6 Number of papers that fall into each research aim for the first literature review. 253
B.7 Number of papers that fall into each planning decision for the first literature review. 253
B.8a Summary of papers identified through the Scopus search in the first literature review, with a total of 90 papers. 255
B.8b Summary of papers identified through the Scopus search in the first literature review, with a total of 90 papers. 256
B.9 Number of papers that fall into each JCR category for the second literature review. 257
B.10 Number of papers that fall into each hospital setting for the second literature review. 257
B.11 Number of papers that fall into each condition area for the second literature review. 257
B.12 Number of papers that fall into each research aim for the second literature review. 258
B.13 Number of papers that fall into each planning decision for the second literature review. 258

C.1 List of attributes used for the CART analysis. 262

C.2 Linear regression results for continuous variables. 263
C.3 Linear regression results for categorical variables. 266
C.4 Logistic regression results for continuous variables. 267
C.5 Logistic regression results for categorical variables. 270
C.6 Data and parameters used within the deterministic model. 276
C.7 Data and parameters used within two-stage stochastic model. . . 277

D.1 The daily bed demands for each specialty grouped by regions within ABUHB for three years' worth of patient admissions, using the regression tree and average LOS. 285
D.2 The daily bed demands for each specialty for regions one, two and three within ABUHB for three individual years' worth of patient admissions, using the regression tree and average LOS. 288
D.3 The daily bed demands for each specialty for regions four, five and six within ABUHB for three individual years' worth of patient admissions, using the regression tree and average LOS. 289
D.4 The daily bed demands for each specialty for regions one, two and three within ABUHB for three individual years' worth of patient admissions, using the regression tree and the year specific average LOS. 292
D.5 The daily bed demands for each specialty for regions four, five and six within ABUHB for three individual years' worth of patient admissions, using the regression tree and the year specific average LOS. 293
D.6 The daily bed demands for each specialty grouped by regions within ABUHB for three years' worth of patient admissions, using the regression tree and specific LOS. 294
D.7 The daily bed demands for each specialty for regions one, two and three within ABUHB for three individual years' worth of patient admissions, using the regression tree and the year specific LOS. . 296
D.8 The daily bed demands for each specialty for regions four, five and six within ABUHB for three individual years' worth of patient admissions, using the regression tree and the year specific LOS. . 297
D.9 The daily bed demands for each specialty grouped by regions within ABUHB for three years' worth of patient admissions, using the classification tree and average LOS. 300
D.10 The daily bed demands for each specialty for regions one, two and three within ABUHB for three individual years' worth of patient admissions, using the classification tree and the node average LOS. 302

D.11 The daily bed demands for each specialty for regions four, five and six within ABUHB for three individual years' worth of patient admissions, using the classification tree and the node average LOS. 303
D.12 The daily bed demands for each specialty for regions one, two and three within ABUHB for three individual years' worth of patient admissions, using the classification tree and the yearly average LOS. 306
D.13 The daily bed demands for each specialty for regions four, five and six within ABUHB for three individual years' worth of patient admissions, using the classification tree and the yearly average LOS. 307
D.14 The daily bed demands for each specialty grouped by regions within ABUHB for three years' worth of patient admissions, using the classification tree and specific LOS. 310
D.15 The daily bed demands for each specialty for regions one, two and three within ABUHB for three individual years' worth of patient admissions, using the classification tree and the yearly specific LOS. 312
D.16 The daily bed demands for each specialty for regions four, five and six within ABUHB for three individual years' worth of patient admissions, using the classification tree and the yearly specific LOS. 313

List of Acronyms

A&E = Accident and Emergency

ABCi = Aneurin Bevan Continuous Improvement Team

ABUHB = Aneurin Bevan University Health Board

CART = Classification and Regression Trees

CCH = Chepstow Community Hospital

CH = County Hospital

COIN-OR = Computational Infrastructure for Operations Research

COIN-OR CBC = COIN-OR Branch and Cut

COIN-OR CLP = COIN-OR Linear Programming

COTE = Care of the Elderly

DRG = Diagnosis-related group

EEV = Expected Cost when using the solution $\bar{}(\bar{\xi})$

ED = Emergency Department

EIV = Expected Input Value

ESSV = Expected Skeleton Solution Value

EV = Expected Values

FN = False Negative

FP = False Positive

GG = Geriatrics and Gerontology

GP = General Practitioner

GUH = Grange University Hospital

HPS = Health Policy and Services

ICD10 Codes = International Statistical Classification of Diseases and Related Health Problems - 10^{th} revision

IE = Industrial Engineering

JCR = Clarivate Journal Citation Report

LOS = Length of Stay

LUDS = Loss of Upgrading the Deterministic Solution

LUSS = Loss Using Skeleton Solution

MI = Medical Informatics

MILP = Mixed Integer Linear Program

MIP = Mixed Integer Program
MIU = Minor Injury Unit
MSE = Mean Square Error
MVHSCF = Monnow Vale Health and Social Care Facility
NHH = Nevill Hall Hospital
NHS = National Health Service
OLS = Ordinary Least Squares
OR/MS = Operational Research and Management Sciences
PAS = Patient Admission System
RadIS = Welsh Radiological Information System or WRIS
RGH = Royal Gwent Hospital
RIHSC = Rhymney Integrated Health and Social Care
RP = Here and Now Solution
STWAH = St Woolos Acute Hospital
STWCH = St Woolos Community Hospital
T&O = Trauma and Orthopaedics
TN = True Negative
TP = True Positive
VSS = Value of the Stochastic Solution
YAB = Ysbyty Aneurin Bevan
YYF = Ysbyty Ystrad Fawr

Chapter 1

Introduction

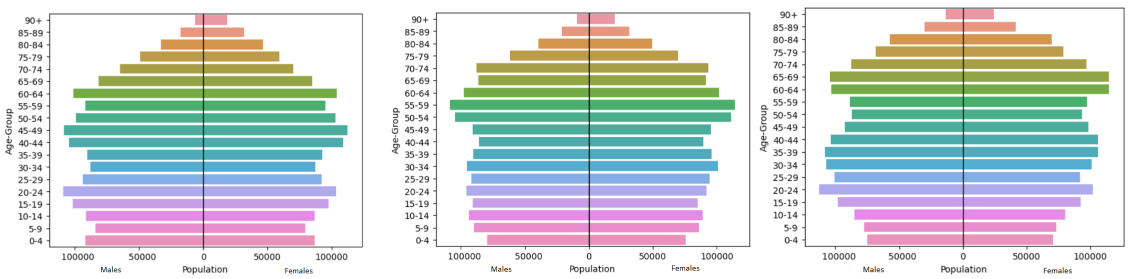
This research project has been conducted in collaboration with the Clinical Futures team within the Aneurin Bevan University Health Board (ABUHB) [14] at NHS Wales, UK, and is funded jointly by the Welsh Government's European Social Fund (ESF) convergence programme for East Wales and the Knowledge Economy Skills Scholarship (KESS2) [15]. Clinical Futures is the health board's plan for sustainable health and care services for the NHS across South East Wales [16]. The organisational structure of bed capacity and personnel resource planning within ABUHB is examined in this thesis, specifically for the frail and elderly patient demography.

In the United Kingdom (UK), the National Health Service (NHS) is a publicly funded healthcare system [17]. Founded in 1948, the principle was that services should be comprehensive, universal and free at the point of delivery, with the basis for health care being clinical need rather than financial capability. It is the second largest single-payer healthcare system in the world and in 2021 accounted for £229 billion of the UK Government's annual spending [18].

1.1 Frail and Elderly Patient Demographics

Due to declining fertility and mortality rates as well as increased life expectancy, many countries are dealing with an ageing population. Additionally, those who were born during the post World War Two baby boom, are now aged 60 or older. These factors are among the key causes behind the United Nations (UN) prediction that by 2050, one in six people worldwide, and one in four in Europe, will be aged over 65. According to the UN [19], in 2018 there were more people aged 65 and older than children aged under five worldwide for the first time in history. In healthcare research, elderly patients are commonly defined as individuals who are aged 65 and older [20, 21, 22, 23]. This definition will be applied throughout this thesis.

Within the UK in 2022, 18.65% of residents are aged 65 and over (approximately, 12.5 million), of which 8.63% are aged 75 and over [24]. The population's share of those aged 90 and older is thought to be around 0.91%. With 21.09% of the population over the age of 65, Wales has a proportionally larger elderly population than the rest of the UK. Those aged 75 and older account for 9.68% with 0.97% of the population being aged 90 and over. Figure 1.1 displays the population pyramids for the Welsh population from the 2001 and 2021 censuses, together with the estimated population for 2031. The 2011 pyramid demonstrates stationary growth, which is defined as the population remaining constant in different age groups. The 2021 and 2031 pyramids depict the start of a constrictive pyramid, in which the number of new births are low, and the population is ageing. Between 2020 and 2030, the Welsh Government projects a 16.1% growth in the population aged 65 and over. This rises to an increase of 23.9% for those aged 75 and over between 2020 and 2030 [25].



(a) 2011 (b) 2021 (c) 2031

Figure 1.1: Observed population pyramids for the years 2011 (a) and 2021 (b), as well as the expected projected population pyramid for the year 2031 (c). The graphs were generated by data gathered from [4] and [5].

Frailty is described as having a high risk of falling into dependency as a result of a negative event, such as an accident, fall or disability. Despite the fact that frailty is more common as people get older, it develops independently from ageing [26]. Since there is no international standard of measurement for frailty, this is challenging to categorise and monitor [27]. This means that determining whether or not a person is frail frequently depends on clinical judgement. Even in the UK, there are various ways to determine frailty, such as the clinical frailty score [28] and the electronic frailty score [29]. According to estimates from Age UK [30], 10% of those over 65 live with frailty, and the percentage rises to between 25% and 50% for those over 85.

Within this research, a frailty score was created using a combination of two methods ([31, 32]), on the hospital disease codes, also known as the International Statistical Classification of Diseases and Related Health Problems - 10^{th} revision (ICD10) [10, 33]. The ICD10 code, which was first mandated for use in the UK in 1995, provides descriptions of all recognised diseases and injuries. Each condition is detailed with diagnostic characteristics and given a unique identifier that is used to code morbidity data from patient and clinician records as well as mortality data on death certificates. The ICD10 has a minimum of three characters, and a maximum of four characters can be added to provide more information about the diagnosis. An example of the breakdown of an example ICD10 code is shown in Table 1.1:

	Category				Subcategory			Extension
Description	1^{st}	2^{nd}	3^{rd}		4^{th}	5^{th}	6^{th}	7^{th}
Fracture of shoulder and upper arm	S	4	2					
Fracture of upper end of humerus	S	4	2	.	2			
Unspecified fracture of upper end of humerus	S	4	2	.	2	0		
Unspecified fracture of upper end of right humerus	S	4	2	.	2	0	1	
Unspecified fracture of upper end of right humerus - initial encounter for closed fracture	S	4	2	.	2	0	1	A

Table 1.1: S42.201A - Example ICD10 code that shows the categorisation of each sublevel of ICD10. The first three characters provide a high level description of the category, where each additional character increases the diagnostic complexity [10].

Gilbert et al. [31] developed a hospital frailty risk score, where certain ICD10 codes were more than twice as likely to be present in a frail patient compared to a non-frail patient. This scoring system has a range of 0.1 to 7.1, where a higher score suggests a higher proportion of frail patients within the groupings. Soong et al. [32] generated a list of syndromes which were more common within frailty patients, where if a syndrome is present then a score of one is given. To assure coverage of all frailty syndromes, the two score measures were combined because they each contained different ICD10 codes. This yields a range from 0 (where there is no frailty syndrome present) to 8.1, the highest scoring frailty syndrome.

1.2 Aneurin Bevan University Health Board

The Aneurin Bevan University Health Board (ABUHB) was established in 2009 and serves 650,000 residents located in five counties: Blaenau Gwent, Caerphilly, Monmouthshire, Newport and Torfaen, as well as some areas of South Powys (Figure 1.2). Two-thirds of the health board's 14,000 personnel work directly with patients [34]. The health board has direct links with applying Operational Research (OR) and modelling to improve healthcare decisions with a team of analysts embedded within the organisation forming the Aneurin Bevan Continuous Improvement team (ABCi) [35].

Clinical Futures is the health board's plan for sustainable health and care services for the NHS across the Gwent area. General practitioners (GP) practices, emergency

Figure 1.2: Aneurin Bevan University Health Board (ABUHB).

rooms, and longer wait times for services that a large number of people require, such as orthopaedic and ophthalmology care, are dramatically putting greater strain on the whole healthcare system. Their aim is to deliver effective and efficient care in hospitals while planning across all services to keep people out of hospital. Clinical Futures is reforming the organisation to provide more centralised hospital treatment for individuals in need, as well as care closer to home, in order to satisfy these demands and succeed in 'The Wellbeing of Future Generations Act (2015)' [36] (Figure 1.3).

The health board comprises of 11 hospitals providing different levels of care. A specialised critical care centre called The Grange University Hospital (GUH) opened earlier than expected in November 2020 to offer more assistance during the Covid-19 pandemic. Inpatient stays at GUH will not have been included because the primary emphasis of this study is on data collected before the Covid-19 epidemic. The model developed within this thesis will have the ability to be adapted to changing services throughout the health board in the future, including the addition of GUH. There are four minor injury units (MIU's), one acute hospital and five community hospitals (Table 1.2).

These hospitals are spread throughout the health board, with one major accident and emergency (A&E) unit or MIU in each county (Figure 1.4 shows major A&E unit in red and MIU's in blue). Figure 1.4 depicts the community hospitals, which are dispersed around the health board and are shown in purple.

Each hospital offers patients different inpatient services. There are a total of 29

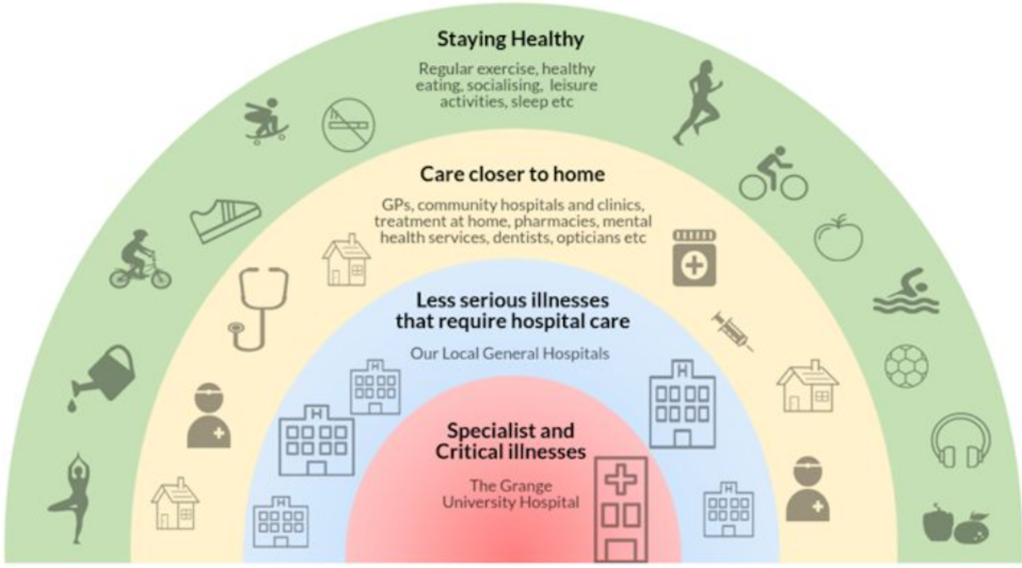


Figure 1.3: Clinical Futures future healthcare plan, where specialist hospital care is provided in one central location, with emphasis being put on care closer to home and healthy lifestyle choices [6].

Hospital Type	**Hospital Name**
Major A&E Unit	The Grange University Hospital (GUH)
Minor Injury Unit (MIU)	Nevill Hall Hospital (NHH), Royal Gwent Hospital (RGH), Ysbyty Aneurin Bevan (YAB), Ysbyty Ystrad Fawr (YYF)
Acute Hospitals	St. Woolos Acute Hospital (STWAH)
Community Hospitals	Chepstow Community Hospital (CCH), County Hospital (CH), Monnow Vale Health & Social Care Facility (MVHSCF), St. Woolos Community Hospital (STWCH), Rhymney Integrated Health & Social Care Centre (RIHSC)

Table 1.2: Type and names of the 11 hospitals located in ABUHB.

specialisations offered by the area. How specialties are currently organised in each of these hospitals is shown in Figure 1.5 (as of March 2020). There are 98 unique hospital and specialty pairings in total. More specialties are found in larger, more acute hospitals, such as RGH, which has 25, as opposed to fewer specialties in smaller hospitals like County Hospital (CH). Appendix A has a complete list of hospitals, along with a summary of the specialties that each one offers.

1.3 Bed Planning and Staff Allocation

Figure 1.6 shows a patient's journey after being admitted to an acute ward. Patients may be admitted as elective or emergency patients. Patients who are unexpectedly and quickly admitted to the hospital are considered emergency admissions. These patients may be accepted through GPs or consultants in ambulatory clinics, as well as A&E units. According to Steventon et al. [37], the number of emergency admissions is rising at an average of 3.2% year. Elective patients receive scheduled

Figure 1.4: Map of hospital locations in ABUHB. Major A&E units are shown in red, MIU's are shown in blue and acute and community hospitals are shown in purple.

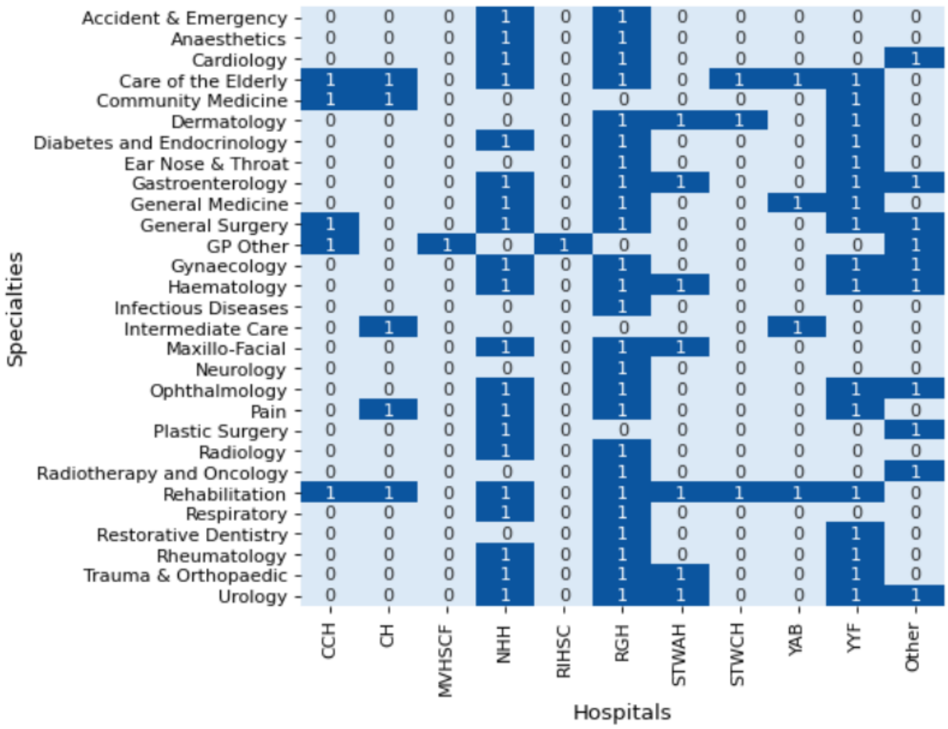


Figure 1.5: Hospitals and specialties in ABUHB. A '1' indicates a specialty is present in a given hospital and '0' indicates a specialty is not present.

care, frequently involving expert clinical treatment or surgery, and are typically referred by a GP or other community health provider.

Patients who are admitted to the hospital, whether on an elective or emergency basis, remain there until they are deemed to be 'medically fit to be discharged'. Patient discharges can frequently be delayed for a variety of reasons, such as waiting for home care, waiting for a permanent bed in a nursing or care facility or awaiting a medical decision and full discharge summary. The home first approach [38], strives to discharge patients with long-term care needs to an appropriate setting, where assessments of what care they require and how it will be financed take place (Care and nursing homes in the UK are typically funded by the patient). This approach is crucial for maximising health outcomes. Most patients leave hospital and go home without any more assistance. However, the majority of patients who encounter delays to discharge require community care and typically, these patients are older adults.

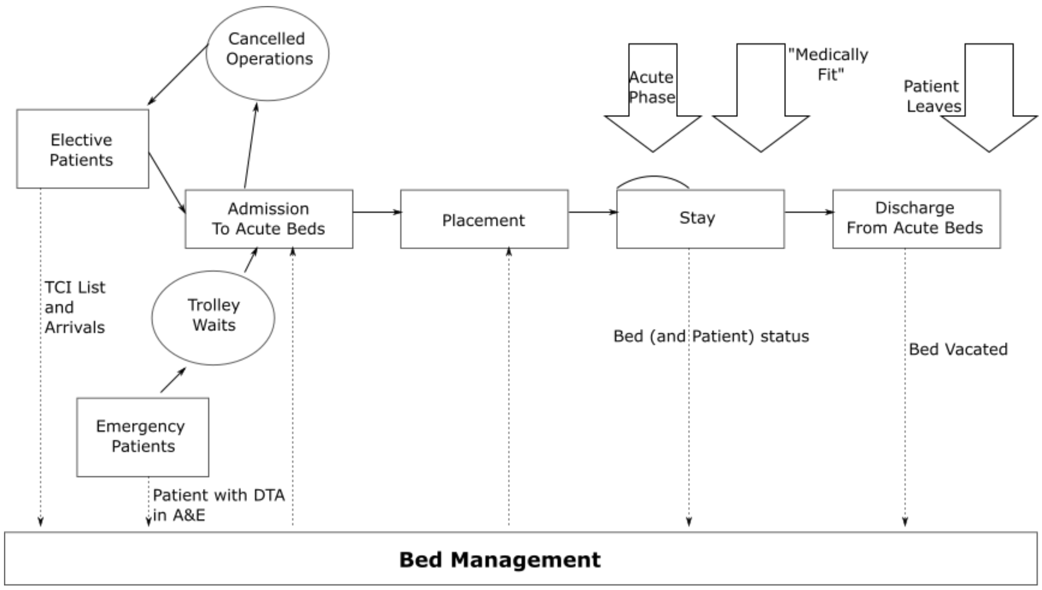


Figure 1.6: Conceptual overview of the patient journey and the role of bed management. Patients arrive into the system via elective and emergency routes and are assigned a bed. Patients remain in hospital for treatment, and are discharged when are classified as 'medically fit' [7].

The Nuffield Trust analysed the most common causes for patients to be bed blocked across hospitals within England [8, 9]. Figure 1.7 illustrates the primary reasons causing delays in patient discharges, prominently centring on limited access to essential services. Notably, home care contributes to 24% of delays, followed closely by short-term rehabilitation at 22%, and care/nursing homes at 15%. These figures underscore the critical constraint faced by community services, impeding their ability to accommodate new patients while occupied beds await discharge, leading to the phenomenon known as "bed blocking". Consequently, this pressing issue calls for attention and intervention to alleviate the strain on healthcare facilities and ensure efficient patient flow.

Elective patient operations are cancelled to make way for priority emergency ad-

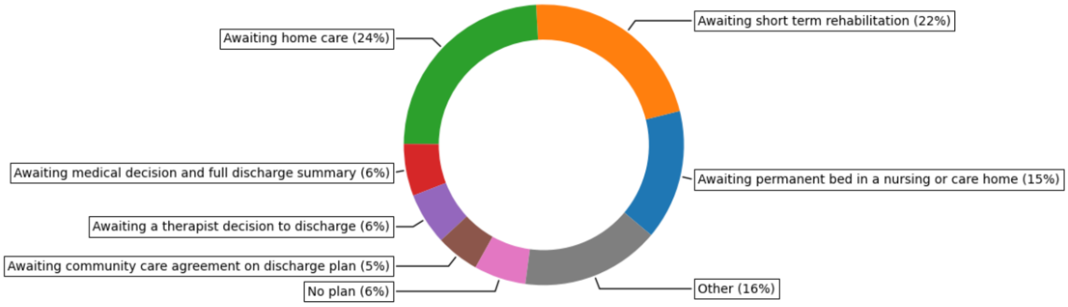


Figure 1.7: Common reasons for delayed discharge for patients with a length of stay (LOS) longer than a week within English hospitals [8, 9].

missions when the capacity is insufficient and the beds are full. Similarly, planned admissions are frequently cancelled and wards are altered in order to securely meet ward criteria if there are insufficient nursing staff. The Nurse Staffing Levels (Wales) Act 2016 [39], was the first law passed in the UK requiring that health boards provide sufficient nursing staff so all patients receive compassionate care. This requires a designated individual to determine the number of nurses necessary to provide care to patients that satisfies all reasonable requirements in each situation. Additionally, they must take all responsible steps to maintain the nurse staff level. This relies on one individual to utilise expert judgement and can be problematic when deciding on the final personnel levels. Staff planning must take into account uplift, and capacity should accommodate anticipated and planned changes in the number of nursing staff members available (e.g., annual leave, training, study leave).

Equation 1.1 displays the method of allocating ward nursing staff in order to calculate the typical nursing staff demand for a 24 hour period [40].

$$\begin{aligned}\text{Average Requirement} =& (\text{Average Hours Per Patient} \times \text{Average Daily Bed Utilisation}) \\ &+ \text{Additional Workload in Nursing Hours Per Day} \end{aligned} \tag{1.1}$$

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. 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 nature of the healthcare industry [41]. By utilising OR approaches, this thesis seeks to enhance planning for both beds and staff.

1.4 Research Aims

To address the organisational needs within the NHS, four research questions were established in partnership with senior staff from Clinical Futures and ABUHB.

1. How do the clinical and demographical attributes of frail and elderly patients effect their length of stay within hospital?
2. How best can specialties be organised among a network of hospitals to ensure staffing and bed costs are minimised, whilst still meeting the demand for frail and elderly patients?
3. Can linking predictive and prescriptive analytics provide improvements for decision making for frail and elderly services?
4. How can deterministic and two-stage stochastic models be used to plan hospital services for frail and elderly patients within Aneurin Bevan University Health Board

1.5 Thesis Structure

This thesis consists of eight chapters, together seeking to answer the four research questions described in Section 1.4. The structure of the thesis is as follows:

- Chapter 1 has contextualised four main research questions of this thesis. Background into the ABUHB has been discussed, with the major problems in the planning of beds and staffing for frail and elderly patients.
- Chapter 2 contains two literature reviews: The first comprising of [42] discusses the application of OR techniques within frail and elderly patient care and the second discusses hierarchical prediction models for patients' length of stay (LOS). These literature reviews provide an insight into the type of OR models that are currently used in this field of study and the opportunities available for researchers to conduct future study.
- Chapter 3 discusses the theory on the predictive modelling techniques that will be applied to health board data in Chapter 5. The concept of classification and regression trees are introduced with an explanation into how these methods can be implemented within Python. Discussion into the application of bed and staff planning is also given.
- Chapter 4 considers the theory of prescriptive modelling. Deterministic and two-stage stochastic models are developed which aims to minimise overall costs by optimally planning bed and nursing resources. This theory will be applied to a case study of frail and elderly in ABUHB within Chapter 5.
- Chapter 5 discuses the current trends within ABUHB. This chapter will also determine the results of the predictive and prescriptive algorithms when applied to the case study of ABUHB. The results will discuss the benefits of

using more sophisticated modelling techniques over ones traditionally used within healthcare modelling.

- Chapter 6 will link the predictive and prescriptive paradigms discussed in previous chapters together. The predictive models will be used as a demand input for the prescriptive models. Various scenario analysis will take place to investigate the impact on the healthcare system and the requirements needed to put in place for the health board in the future.
- Chapter 7 provides a tutorial on how to use the deterministic and two-stage stochastic modelling tools that were developed in both Microsoft Excel's Open-Solver add-in and Python's PuLP package.
- Chapter 8 concludes the thesis with a summary of the research, discussing how each research aim has been met and outlining future research possibilities.

Chapter 2

Literature Review

2.1 Introduction

This chapter aims to provide insight into the existing operational research and management science (OR/MS) literature and its application to planning patient care for frail and elderly patients. By identifying gaps within the literature, this will enable avenues for future research. Additionally, it will also make it possible to define the motivation for this research and establish the significance of the issues currently being faced within healthcare. Two main literature reviews are provided within this chapter. The first literature review, presented in Section 2.2, comprises of the research paper, 'A survey of OR/MS models on care planning for frail and elderly patients' which was published within Operations Research for Health Care [1]. This initial review provided a broad and foundational understanding of the area, determining various strategies and methodologies utilised in healthcare planning for the elderly and frail population. The focus of the review identifies different methods to examine, forecast or improve the current care planning. A subsequent and more targeted literature review, detailed in Section 2.3, has been conducted focusing on LOS prediction through the utilisation of hierarchical prediction models within healthcare settings. Section 2.4 will discuss the overlap between the two reviews. Finally, Section 2.5 will summarise the important findings from both literature reviews and discuss common research gaps.

2.2 A Survey of OR/MS Models on Care Planning for Frail and Elderly Patients

This section focuses on the research, 'A survey of OR/MS models on care planning for frail and elderly patients' [1]. The section is structured as follows: Section 2.2.1

introduces the methods used to identify the papers and discusses related literature reviews identified through this search. Section 2.2.2 analyses and provides a classification of the results. Section 2.2.3 considers gaps within the research, with Section 2.2.4 discussing the findings of the review. For the figures which discuss a classification result, a respective table within the Appendix has been included detailing the reference numbers for each paper. Table B.1 within the Appendix provides a comprehensive list of the 62 papers including each classification category.

2.2.1 Methods

2.2.1.1 Data Sources

To identify major research streams in the literature, a structured search was performed following Webster and Watson's methodology [43]. The search engine, Scopus, was used to identify relevant journal articles and conference proceedings papers, from January 2000 to December 2020 restricting the search to English results.

2.2.1.2 Inclusion Criteria

Webster and Watson [43] highlighted that a literature search should not be confined to one research methodology, one journal or one region. To provide a complete search, the search string contained at least one of the following terms found within each column of Table 2.1: One OR/MS method phrase, one patient flow term and one age category, mentioned in the article title, abstract or given keywords. The Boolean operators: 'AND' and 'OR' were used to concatenate different terms among different categories. For each category, the terms within the string were concatenated with an 'OR' command, whilst the overall categories were connected with an 'AND' command. For terms such as 'Integer program*' an * was used to signify multiple endings, e.g., 'integer program' or 'integer programming'. For phrases with multiple endings with only one character, such as 'Heuristic$', a $ sign was used to indicate this, i.e., 'Heuristic' or 'Heuristics'. This is similar to the methods Hulshof et al. [44] performed within their taxonomy.

To allow multiple OR methods to be investigated, the search terms were identified within Hulshof et al. [44] and Palmer et al's. [45] review of OR methods for modelling patient flow and outcomes. Soft OR methods were investigated including systems thinking, problem structuring and Delphi methods, however, these did not increase the number of publications. This suggests that soft OR methods are under-represented within the field and highlights potential future research. To include a range of techniques, overall classification terms were used such as 'Metaheuristic$' as well as common methods encompassed within this technique such as 'Tabu search' and 'Genetic algorithm'.

OR Method		Patient Flow	Classification of People
Agent based model*	Network analysis	Appointment	Elderly
Branch and bound	Neural network$	Capacity allocation	Elderly care
Branch and price	Optimi*	Capacity management	Frail*
Clustering	Quadratic program*	Capacity planning	Geriatr*
Column generation	Queuing	Care access	Home care
Computer simulation	Queueing	Care pathway	Long term care
Constraint program*	SCA	Clinical pathway	Nurs* care
Discrete event simulation	Scatter search	Critical pathway	Old people
Discrete optimi*	Scheduling	Demand forecasting	Older people
Dynamic program*	Simulation	Demand management	Palliative care
Genetic algorithm	SSM	Demand prediction	>65
Goal program*	Strategic Choice Analysis	Flow of care	
Heuristics$	Strategic Options Development and Analysis	Flow of patients	
Integer program*	Stochastic analysis	Integrated pathway	
Linear program*	Stochastic modelling	Patient flow	
Logistics	Stochastic processes	Patient pathway	
Markov chain	Stochastic program*	Patient process	
Markov decision	SODA	Patient route	
Markov model	Soft OR	Patient throughput	
Mathematical model	Soft Systems Methodology	Process flow	
Mathematical program*	System dynamics	Scheduling	
Metaheuristic$	Tabu search	Whole-system$	
Mixed integer program*			

Table 2.1: Scopus search string terms for the care planning literature review to include at least one 'OR Method', one 'Patient Flow' term and one 'Classification of People' term.

Patient flow terms were identified through multiple sources (Table 2.1). Firstly, Palmer et al.'s [45] review on patient flow within community care allowed demand and capacity terms such as "demand management" and "capacity allocation" to be incorporated. Secondly, De Luc et al. [46] found 17 phrases which encompassed pathways of patients with the most prominent terms within the literature: "integrated care pathway" and "critical pathway". These terms all loosely follow the same three main stages: the development process to design the pathway; the application and use of the pathway; and the ongoing review of the pathway to learn from the practical experience and to continuously apply improvements [46, 47].

To ensure a variety of journal sets were used, the search focused on five categories in the Clarivate Journal Citation Report (JCR). The five categories were as follows: Geriatrics and Gerontology (GG), Health Policy and Services (HPS), Industrial Engineering (IE), Medical Informatics (MI) and Operations Research and Management Sciences (OR/MS). The rationale to select these five journal categories was because they contain journals in which OR/MS methods are applied to healthcare. A number of upcoming journals were also incorporated as they do not belong to a JCR category and these were appropriately assigned to one of the five categories. These journals were as follows: Health Systems, IISE Transactions on Healthcare Systems Engineering (formally known as IIE Transactions on Healthcare Systems Engineering), Operations Research for Health Care and Proceedings of the Winter Simulation Conference. A brief description of each journal category is as follows with the four additional journals added to the most appropriate JCR category:

- Geriatrics and Gerontology (GG) - Captures a subgroup of medical journals which focus on clinical problems in the treatment of elderly patients (e.g., Age and Ageing).
- Health Policy and Services (HPS) - Captures journals covering policy and service improvements within healthcare systems (e.g., Health Care Management Science and Journal of Health, Organisation and Management).
- Industrial Engineering (IE) - Includes papers that focus on systems that integrate people, materials and equipment to provide a service (e.g., International Journal of Simulation Modelling and IISE Transactions on Healthcare Systems Engineering).
- Medical Informatics (MI) - Captures papers which focus on healthcare information in clinical studies and medical research (e.g., Health Information Management Journal).
- Operations Research and Management Sciences (OR/MS) - Includes papers focusing on advanced analytical methods to solve complex problems (e.g., Journal of Operations Management, Health Systems, Operations Research for Health Care and Proceedings of the Winter Simulation Conference).

2.2.1.3 Study Selection and Data Extraction

The initial search resulted in 437 papers being identified and these underwent analysis by abstract to determine the papers which met the inclusion criteria. A publication was excluded if the abstract was not relevant to frail and elderly patients and their planning of care. These exclusions reduced the number of papers to 39. As advised within Webster and Watson's paper [43], a forward and backward search was conducted after the initial analysis to ensure related papers that had not met all the key search criteria were included. In total 65 publications were found to be relevant, including three literature reviews. The three reviews are discussed separately within Section 2.2.1.5, with the remainder of this paper focusing on the other 62 papers. The PRISMA diagram showing a visual representation of this process is shown within Figure 2.1.

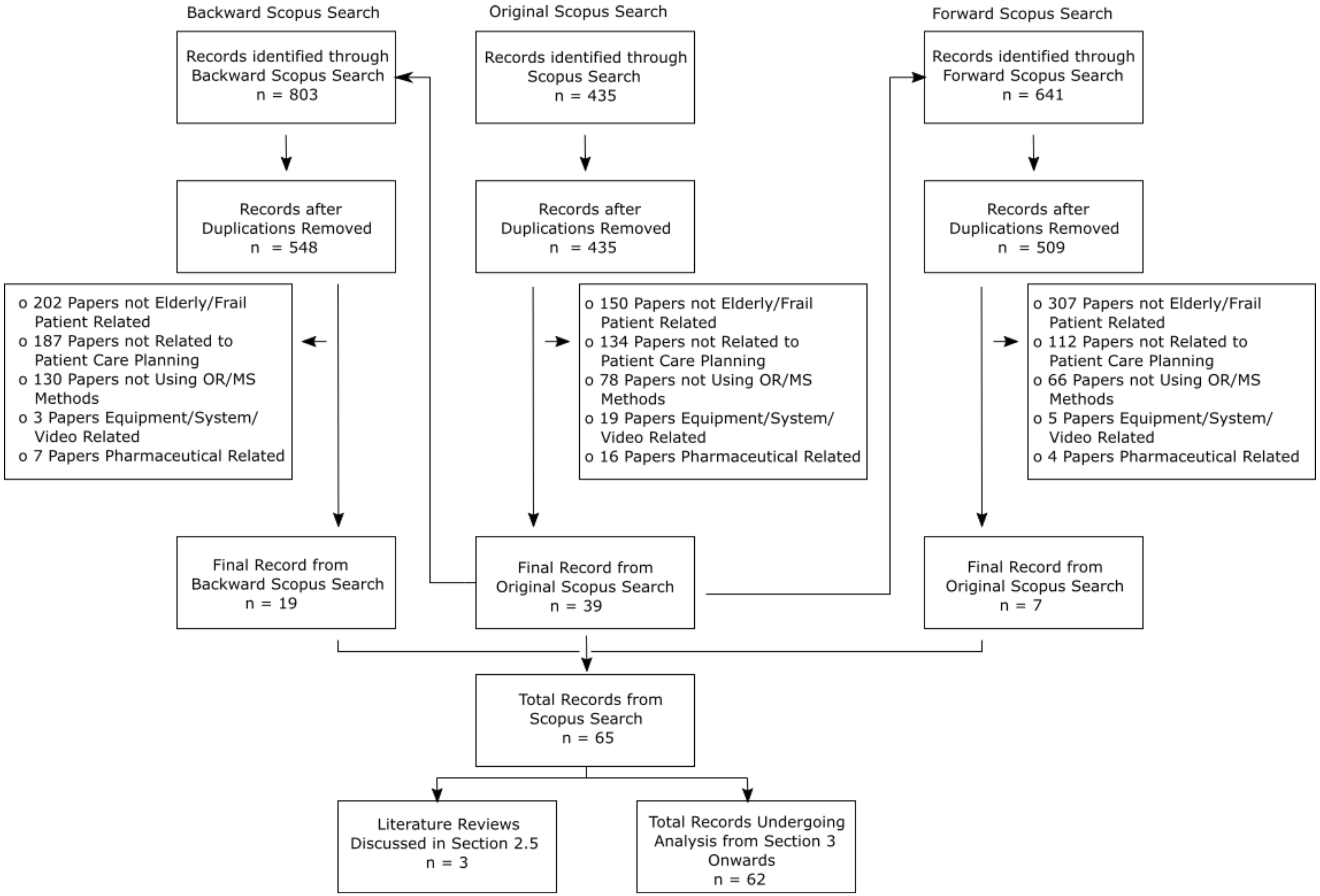


Figure 2.1: The PRISMA flow diagram for the care planning literature review detailing the Scopus searches, the number of abstracts screened, the reason for exclusion and the number of full texts retrieved.

2.2.1.4 Study Protocol

To classify and analyse the papers the following protocol was set up to ensure the objectives of the literature review were met. Firstly, a general classification is provided demonstrating the characteristics of the papers including geographical location, JCR category and publication year. Secondly, the papers' medical context is established with discussion around the care locations and diseases suffered by the frail and elderly. Finally, the research aims, the planning decisions and the types of OR/MS methods utilised within the papers are discussed.

2.2.1.5 Previous Literature Reviews

This subsection aims to provide a brief overview of the three literature reviews identified through the Scopus search, followed by a discussion of how our review aims to fill the gaps in the literature not covered by these reviews.

Firstly, Berntsen et al. [48] used their research to provide evidence for a patient pathway for the frail and elderly to be generated using Digi-PIP (digitally support person-centred, integrated and proactive care) methods. Through a systematic search, 10 papers were identified as focusing on Digi-PIP care on population health, patient experience and cost-effectiveness. The results showed that despite the belief that a Digi-PIP approach was the key to sustainable care, research has not been able to provide sufficient evidence.

Secondly, Freeman et al. [49] focused on patients aged over 65 and the factors affecting the transition from long-term care facilities (LTCFs) to the community. LTCFs were distinguished as care institutions that provided 24-hour nursing care, personal care or other services, whereas community was defined as home care programmes, retirement homes, assisted living facilities or patient's own home, where 24-hour care is not provided. They identified 36 articles and recommended that further understanding was needed due to the complexity of the discharge process with more evidence in the factors and barriers that influence the discharge. The authors concluded that it was unclear of the combination of multidisciplinary team members and institutional factors that best support discharge planning.

The third and final review identified was Gaugler et al.'s [50] paper on the research focusing on admission predictors of community care specifically within nursing homes in the USA. The review identified 77 papers which encompassed 12 data sources. After analysing different methods, such as logistic regression and Cox regression models, on a variety of different care factors, including gender and medical condition, their results identified a number of predictors, e.g., cognitive impairment. The work highlights the opportunity for future research to develop tools using the strongest predictors to estimate nursing home admissions, potentially adapting and applying these methods to predict demand for nursing homes and other long-term care facilities.

The three reviews either focused on specific locations within the pathway; [49] with the movement from LTCFs to the community, [50] with nursing home care or they focused on analysing specific OR/MS methods [48]. Whilst these reviews provide beneficial contributions to their areas, we aim to consolidate literature on a wider scale. Instead of analysing one OR/MS technique, [48], 44 different OR/MS methods have been incorporated into the search criteria to cover a wider range of methods. There has also been expansion across different patient groups and treatment settings, (e.g., nursing homes and palliative care), to ensure each aspect of the pathway and its care planning can be investigated. As multiple settings were analysed, this allowed further investigation into how different settings were applying different OR/MS methods. Additionally, there was analysis on how different settings are working collaboratively to ensure successful care planning. This review will serve as a guide on how to conduct further research on the future challenges in frail and elderly care planning.

2.2.2 Results

After highlighting the focus of previous literature reviews, it was identified that there was a need for the research on OR/MS methods for frail and elderly care planning to be summarised. This would then allow for gaps within the present literature to be determined and a research agenda to be developed enabling these gaps to be filled. The following results analysed the findings of the initial, forward and backward searches and classified the literature by general, medical and methodological contents. Each section provides summary statistics discussing the results. Research gaps and discussion of results will take place in Sections 2.2.3 and 2.2.4, respectively.

2.2.2.1 General Classification

Table 2.2 highlights the divide between the location of the research conducted, with the majority of papers being published within Europe and North America. Kerpershoek et al.'s [51] study focused on eight different European countries analysing access to dementia care and is denoted as 'Multi-national' within Table 2.2. It is worth highlighting that no other papers were found to be multi-national.

English only papers were analysed which may explain why mainly European and North American publications met the inclusion criteria. Further categorisation shows there was a disparity between these continents and the work within this field that is being published. Table 2.2 also displays the JCR categories for each

Country	UK	Canada	USA	Italy	Australia	France	Hong Kong	Ireland	
GG	2	1	2	2	2	0	0	0	
HPS	6	3	3	1	0	1	0	0	
IE	0	1	1	1	0	1	1	0	
MI	2	1	0	1	0	0	0	0	
OR/MS	5	3	3	0	0	1	1	1	
Other	3	1	1	1	1	0	0	1	
Total	18	10	10	6	3	3	2	2	
Country	**China**	**Japan**	**Netherlands**	**Norway**	**Poland**	**Sweden**	**Spain**	**Multi-national**	**Total**
GG	0	1	0	0	1	0	0	1	12
HPS	0	0	1	1	0	0	0	0	16
IE	0	0	0	0	0	0	0	0	5
MI	0	0	0	0	0	0	0	0	4
OR/MS	0	0	0	0	0	1	0	0	15
Other	1	0	0	0	0	0	1	0	10
Total	1	1	1	1	1	1	1	1	62

Table 2.2: Number of papers which fall into each JCR category and the location of where the research was conducted.

country.

The final column in Table 2.2 shows the quantity of papers published within each of the JCR categories as discussed within Section 2.2.1. Within the backward and forward searches, there were 10 papers which did not have ISSNs related to the five JCR categories (Figure 2.2), so these papers have been attributed to the 'Other' category. HPS and OR/MS were the leading journal categories with 16 and 15 papers respectively.

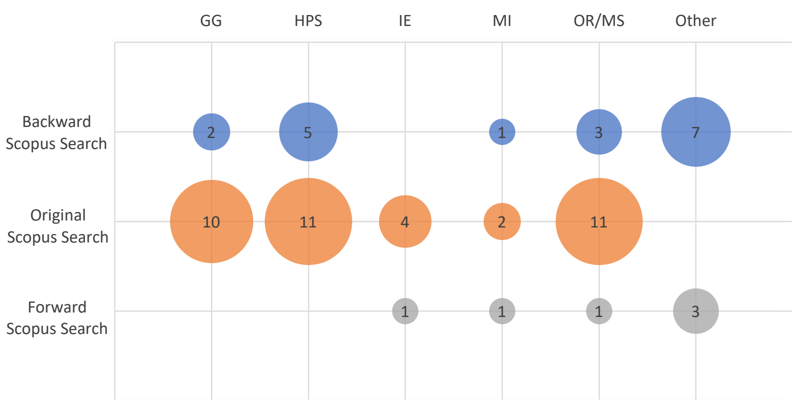


Figure 2.2: Number of publications broken down by JCR categories and Scopus search type.

Although there were only four and five papers within the MI and IE categories respectively, it is important to include these within the analysis as they are under-represented areas and provide a different journal focus on patient pathways and their application.

To show the general trend of the research in this field, Figure 2.3 displays the quantity of papers published every three years. Between 2000 and 2014, the number of papers published remained fairly stable with an upward trend from 2015. Within the last three years, 27% of papers were published, highlighting that this area is becoming more widely researched.

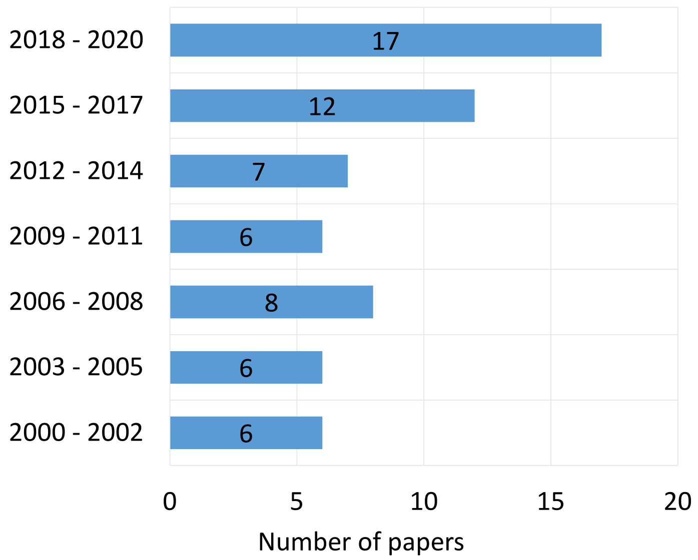


Figure 2.3: Bar chart of the 62 papers identified through the Scopus search, by publication year.

2.2.2.2 Medical Context

This subsection will analyse the medical context of the papers broken down into the medical setting and the condition area. The medical setting of a patient is the location where their care takes place, e.g., a hospital ward or nursing home. The condition area focuses on the medical condition of the patient in the study and whether this was long-term such as dementia or acute, e.g., heart attack. Successful care planning should consider the entire pathway of a patient across multiple settings. It is often necessary to consider the next steps a patient will take to ensure appropriate resources and available capacity to avoid delays in discharge. It is also important to know the condition type of the patient as this will likely affect their discharge destination or the time required to stay within the care setting.

2.2.2.2.1 Medical Setting

The medical setting of the paper was important to understand how care settings can work together for the planning of care for frail and elderly patients. The research focused on three main areas:

- Single Hospital
- Multiple Hospitals

- Community Care

The community care grouping encompassed: Home care, long-term care, nursing care and hospice care, to meet the wide range of healthcare services that do not take place in a hospital setting. Figure 2.4 shows a Venn diagram which breaks down the publications into the type of care settings. The numbers reveal there was a clear focus on both community care and single hospital settings.

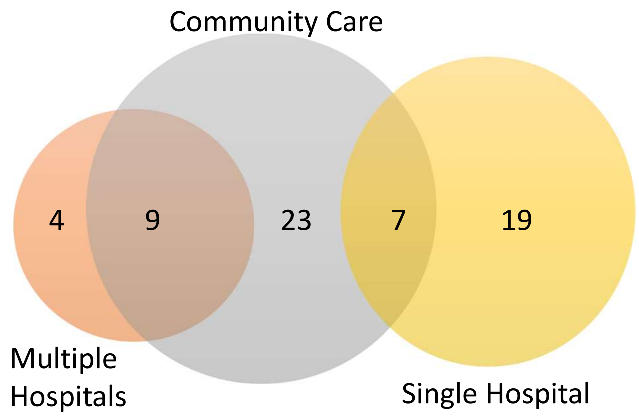


Figure 2.4: Venn diagram of the 62 papers identified through the Scopus search, by medical setting.

In total, there were 16 papers (26%), which used care planning in a holistic manner. These papers were particularly interesting as they focused on the cross over between community care and either single or multiple hospitals. Papers [52, 53, 54, 55, 56, 57, 58, 59, 60] focused on the intersection between community care and multiple hospitals whilst papers [61, 62, 63, 64, 65, 66, 67] focused on single hospitals and community care.

A brief overview of Patrick [57] and Taylor et al. [67] is provided as these works are examples of community care with multiple hospitals and single hospital settings.

Patrick [57] developed a Markov decision model that determined the required bed numbers in long-term care facilities in order to keep demand below a given threshold. Patrick also developed a simulation model to incorporate both hospital and community care demand to predict the impact of policy implementations. These models have aided future capacity planning by comparing current practices against proposed alternative models.

The work conducted by Taylor et al. [67] involved modelling the time geriatric patients spent in hospital and in community care. The authors generated a stochastic compartmental Markov model with three hospital components: acute care; rehabilitative and long stay; two community components and an absorbing state. They were able to successfully provide short-term estimates and a better understanding of future bed usage within geriatric hospital settings.

2.2.2.2.2 Illness/Disease Focused Papers

There were eight papers which focused on a specific illness/disease often suffered by the frail and elderly. These illnesses were as follows: dementia [68], falls [69], gastrointestinal [70], heart failure [71, 72, 73] and hip fractures [74, 75]. Seven of the eight papers focused on a single hospital setting and the remaining paper focused on community care [68]. There were a further eight papers which focused on an inpatient department: an emergency department [57, 76, 77, 78] or a geriatric ward [79, 80, 81, 82]. Interestingly, none of these 16 papers used care planning across multiple settings. The recovery times for frail and elderly patients is usually longer than the general population. Often, they will require further care within the community once they are ready to be discharged. If there are insufficient resources or a lack of availability within the community, then these patients may have to remain in the hospital, causing bed blocking.

2.2.2.2.3 Community Care Focused Papers

There were 23 papers which had community care as the only setting. Six of these papers concentrated on the overlap and movement between settings in community [83, 84, 85, 86, 87, 88]. Lin et al. [89] focused on day care whilst [90, 91, 92, 93] studied home care. The remaining papers focused on either nursing care [94, 95, 96] or long-term/aged related care facilities [51, 68, 97, 98, 99, 100, 101, 102, 103]. This showed that within the grouping of community care, there was a wide range of different settings being analysed.

For frail and elderly care mapping to be successful, the journey of a patient should be documented, which will depend on the type of illness or condition they are suffering from. Therefore it is important for more research to be conducted into specific medical conditions and healthcare settings. Monitoring the journey of a patient from admittance through to discharge, may become a valuable tool in order to predict long-term demands and capacity planning.

2.2.2.2.4 Condition Area

A classification which has been commonly used within healthcare literature reviews is the condition area of the patient [104, 105]. These condition areas are often categorised as either: acute or chronic.

- Acute - Medical conditions that are brought on unexpectedly, e.g., heart failure, or patients undergoing or recovering from a surgical procedure, e.g., femur fracture.
- Chronic - Medical conditions that are prolonged and rarely cured, e.g., dementia.

Often chronic conditions can develop and cause an acute condition, and likewise, if untreated an acute condition can often become chronic. It is therefore important for research to be focused on both strands of conditions, especially when considering frail and elderly patients. These patients often have many chronic conditions which require long-term care, however, they can easily become more serious conditions requiring immediate care.

Figure 2.5 displays the quantity of papers in each condition category.

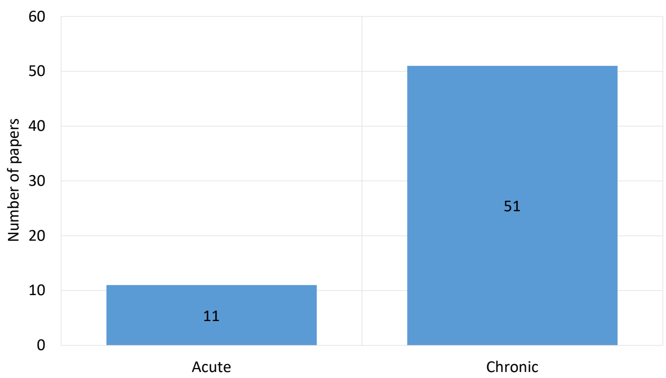


Figure 2.5: Bar chart of the 62 papers identified through the Scopus search, by 'Acute' or 'Chronic' medical conditions.

Within the elderly population, there are many people who have multiple long-term conditions (MLTC), which may explain why there were several papers that focus on chronic conditions. There were 11 papers focused on the acute care setting [69, 70, 71, 72, 73, 75, 76, 77, 78, 106, 107]. This was surprising given frail and elderly patients are more likely to suffer from acute conditions as a result of chronic illness. The 11 papers were all based in a single hospital setting with no overlap between community care (Figure 2.6).

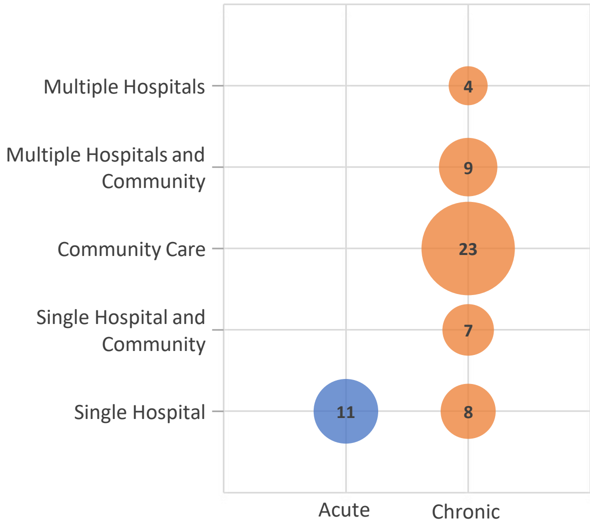


Figure 2.6: Cross analysis of medical setting and condition area of the 62 papers identified through the Scopus search.

2.2.2.3 Methodological Content

This subsection will analyse the technical side of the research. Firstly, discussing the research aims, then moving on to the planning decision levels discussed in Hulshof et al.'s taxonomy [44]. Finally, the different OR/MS methods used within frail and elderly care planning are identified.

2.2.2.3.1 Research Aims

The literature, in terms of care planning, can be grouped into three main aims: examining, forecasting and improving. These categories indicate the direction of the research and the main interest to the authors.

- Examining - Using OR methods to determine how a care plan was performing, e.g., characteristics of patients who move within community care [84], hospital outcomes following an updated care pathway [75].
- Forecasting - Predicting future scenarios with the current care plan in place, e.g., forecasting LOS in hospital and community care [65], capacity planning in community care settings [86].
- Improving - Improvements were made or suggested to enhance the quality of care planning, e.g., improve elderly care in hospital [58], improving quality and efficiency in home care [90].

Figure 2.7 displays the quantity of papers in each research aim category.

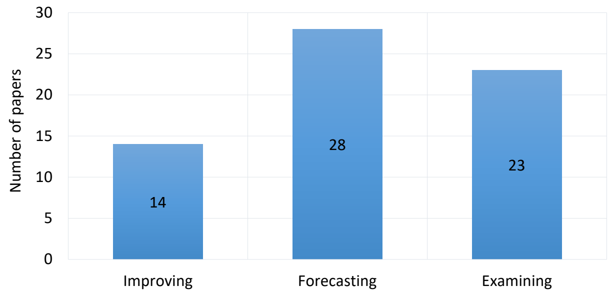


Figure 2.7: Bar chart of the 62 papers identified through the Scopus search, by research focus.

There were three papers which considered a multiple combination of the research aims. Abe et al. [70] examined how polypharmacy affected gastrointestinal surgery patients, whilst also identifying the effects on LOS if measures reducing polypharmacy were implemented. Garg et al. [63] and Patrick [57] both have improving and forecasting aims: Garg et al. focused on improving admission scheduling with resource forecasting and Patrick focused on improving waiting times along with capacity planning. All three of these papers, focused on a single hospital setting.

The results showed most papers aim to forecast frail and elderly patients in the system. Out of the 28 papers, 16 focused on predicting future demands and how the corresponding departments would be required to adapt to this change [52, 54, 57, 66, 79, 81, 83, 85, 86, 88, 91, 95, 97, 99, 101, 108]. A further nine papers aimed to predict the LOS of patients in hospitals or within community care [65, 67, 70, 73, 82, 103, 109, 110, 111]. Surprisingly, of these nine, one author was co-author on five of these papers, suggesting that the research within this field is limited to a few research teams [65, 73, 82, 110, 111].

There were 14 papers focused on making improvements to an aspect of the pathway. Five of these aimed to improve the flow of patients [76, 78, 107, 112, 113]. Only three papers had the primary focus on improving patient care [58, 77, 90]. Their results highlighted the importance of appropriate care to the elderly, which in Rossille et al.'s [77] paper can be achieved by successfully scheduling patients in an emergency department and not categorising these patients by their symptoms. Ragab et al. [58] used simulation modelling to improve the management of frail patients by introducing intermediate care beds for those admitted to acute hospitals. Eveborn et al. [90] used the vehicle routing problem to improve quality for patients receiving home care.

2.2.2.3.2 Planning Decisions

Hulshof et al's. [44] research on taxonomic classification in healthcare systems highlighted a hierarchy of decision-making techniques. There were three different decision levels discussed: strategic, tactical and operational. A brief description is given as follows:

- Strategic planning focuses on structural decision-making such as determining the locations of facilities or resource capacities, these often have a long planning time.
- Tactical planning addresses the execution of strategic plans on the mid-horizon planning time, e.g., staffing levels.
- Operational planning analyses short-term decisions and focuses on the individual patient and resources. Patient appointment scheduling would be an example of this.

Figure 2.8 displays the breakdown of publications by planning decision level.

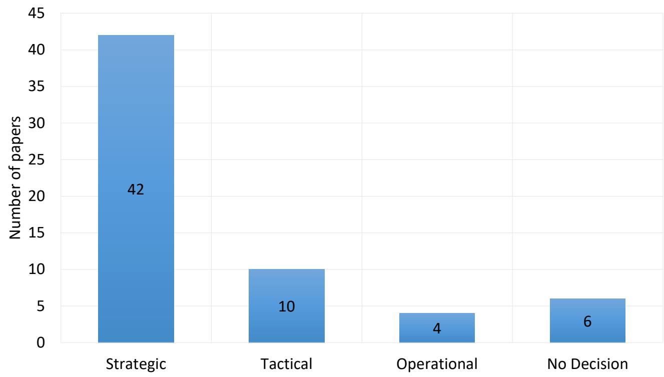


Figure 2.8: Bar chart of the 62 papers identified through the Scopus search, by planning decision level.

The majority of papers focused on strategic planning, with a high concentration on capacity planning and placement policy. Xie et al. [103] used strategic planning and created a Markov model to represent the LOS of the elderly moving within and between residential and nursing homes. By developing this model, it aimed to assist planning authorities to fully understand the pattern of resource usage within their local area. Further work included an extension of their model to incorporate particular attributes of patients, e.g., age, gender and physical conditions, to predict differences in survival by treatment locations.

The number of papers (two) on the operational planning level was smaller than the number of papers (six) that had no planning decision level. The results showed that care planning for frail and elderly pathways was being addressed on some scale across all three decision levels; day-to-day; mid-level planning; long-term, wider policy decisions. However, as there were substantially fewer papers in the tactical and operational decision levels it would suggest these areas are more difficult to plan in frail and elderly healthcare.

Two of the four operational planning papers focused on staff scheduling [90, 91] and the other two focused on readmission of patients [64] and treatment outcomes [72]. Kul et al. [72] evaluated the effect of the heart failure care pathway on geriatric patients. Logistic regression showed positive results supporting the use of care pathways, highlighting reduced mortality and readmission rates along with no increase in hospital costs.

Figure 2.9 displays the cross analysis between the planning decision and the medical setting. To some degree, tactical planning levels were addressed in each setting, although the operational papers had only been addressed within community care and single hospital settings.

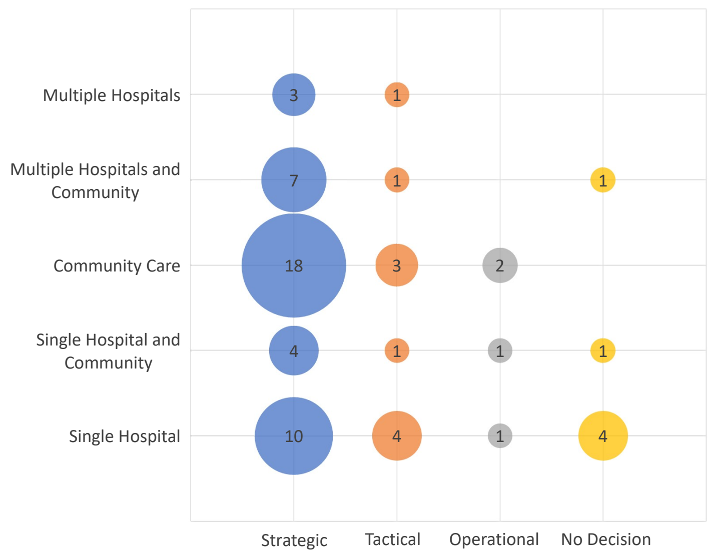


Figure 2.9: Cross analysis of planning decision and medical setting of the 62 papers identified through the Scopus search.

Hulshof et al. [44] analysed the taxonomy for papers within the OR/MS JCR category. This has been further extended to include four additional JCR categories. The cross analysis can be seen within Figure 2.10.

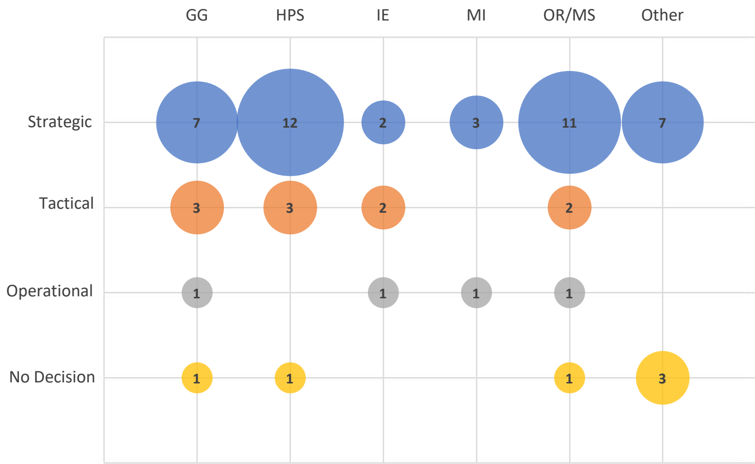


Figure 2.10: Cross analysis of JCR category and planning decision level of the 62 papers identified through the Scopus search.

Figure 2.10 shows there is a spread of decision levels against each JCR category. It demonstrates that the decision level taxonomy discussed within Hulshof et al. [44] can be successfully applied to JCR categories other than OR/MS.

2.2.2.3.3 OR/MS Methods

The final area for analysis was investigating the OR/MS methods that have been utilised within these studies. There has been a variety of different OR/MS techniques which have been used to demonstrate the effectiveness of care planning designed specifically for the frail and elderly. Figure 2.11 demonstrates the quantity of each of these methods, with statistical analysis encompassing a wide range of traditional statistical/operational analysis techniques, including Cox's regression analysis [109], mixed exponential distributions [114] and time survival analysis [96]. Optimisation included mixed integer programming [52] and quadratic programming [102]. Simulation included discrete event simulation [88] and system dynamics [68]. There were two papers which focused on multiple methods.

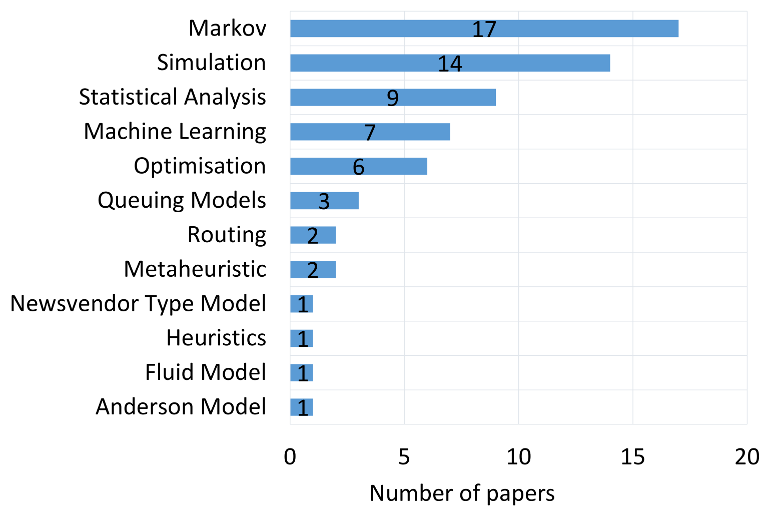


Figure 2.11: Bar chart of the 62 papers identified through the Scopus search, by mathematical method.

Patrick [57] as discussed in Section 2.2.2.2.1 used discrete event simulation along with a Markov decision process model to predict demand for long-term care.

Mohammadi Bidhandi et al. [86] used both simulation and queueing theory methods to plan demand capacities within community care. They focused on six services and ran their optimised queueing model through a simulation to determine transient behaviours of the system. By combining these two methods, capacities were optimised over the entire network at one time, instead of considering them as separate isolated units.

Table 2.3 shows the breakdown of each OR/MS method and its corresponding setting. The seven papers which looked at the overlap between single hospitals and community care all used Markov methods. This may suggest that Markov methods are the most applicable to this setting, particularly for frail and elderly individuals. The data available may also lend itself well to fit into a Markov model. Expanding upon this, when analysing both single and multiple hospitals with community care, only Markov, simulation and statistical analysis methods were used, suggesting these methods were useful when applied to multiple services at the same time. This leaves room in these settings for further research using alternative OR/MS methods to the Markov model.

Community care settings used the widest range of methods, 13 in total, for care planning. This showed the data available within community care can accommodate a variety of methods. There is potential for further investigation and expansion into the methods utilised as an area for future research.

	Single Hospital	Single Hospital and Community	Community Care	Multiple Hospitals and Community	Multiple Hospitals	Total
Markov	[69, 73, 79, 110, 111, 113]	[61, 62, 63, 64, 65, 66, 67]	[103]	[56, 57]	[82]	17
Simulation	[76, 106, 108]		[68, 83, 86, 88, 99, 100, 101]	[57, 58, 59]	[80]	14
Statistical Analysis	[71, 74, 75, 107]		[92, 95, 96]		[109, 114]	9
Machine Learning	[70, 72, 77, 78]		[84, 87, 97]			7
Optimisation			[94, 98, 102]	[52, 53, 54]		6
Queueing models	[81, 112]		[86]			3
Metaheuristic			[89, 91]			2
Routing			[93]	[55]		2
Anderson Model			[51]			1
Fluid model				[60]		1
Heuristics			[90]			1
Newsvendor type model			[85]			1
Total	19	7	24	10	4	64

Table 2.3: Number of papers which fall into each medical setting and OR/MS method within the published research.

Note: Mohammadi Bidhandi et al. [86] and Patrick [57] utilise two methods and therefore appear twice in the table. This resulted in a total of 64 publications.

Figure 2.11 and Table 2.3 highlight that Markov models were the most frequent method used, followed by simulation and statistical analysis. Within these 17 papers, there were a variety of Markovian methods utilised, however these were subgrouped into continuous and discrete time models. Figure 2.12 shows the breakdown of the Markov category from Figure 2.11. Continuous time Markov models were more often used with a total of 14 papers [56, 61, 62, 64, 65, 67, 69, 73, 79, 82, 103, 110, 111, 113]. There were a variety of different types of continuous time models with many focusing on Coxian phase-type Markov models. There were three papers which used discrete time [57, 63, 66] to model frail and elderly patients.

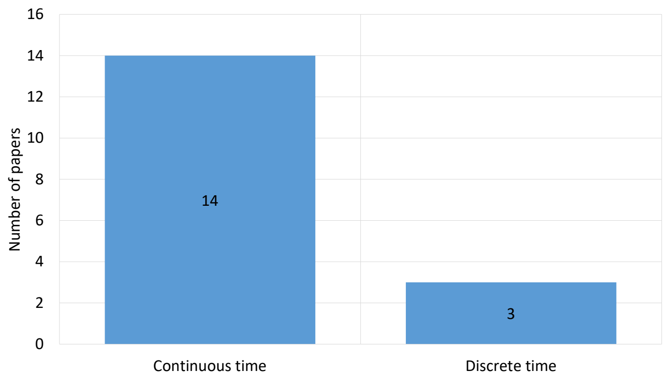


Figure 2.12: Bar chart of the 17 papers using a Markov model identified through the Scopus search, classified by ‘Continuous’ or ‘Discrete’ time.

2.2.2.4 Common Themes

Section 2.2.2 has provided an overview of the work on care planning for frail and elderly. Markov models were the most common method applied to healthcare settings. In more recent years, there has been the emergence of newer techniques being applied to healthcare, i.e., metaheuristics, the Anderson Model and fluid models. Strategic planning remained the most common planning decision level across the OR/MS methods, showing research was being accomplished in longer term care planning. There has been a wider spread of research aims across the papers, although, the majority tend to focus on forecasting future scenarios rather than improving the current systems in place. Finally, the emphasis of these papers has been on single settings, whether this be within a hospital or the community. The current research provides a wide range of different techniques for readers to apply to their own hospital or community care facility. However, there remains scope for future research to be conducted in the frail and elderly patient setting.

2.2.3 Research Gaps

The literature found within this review covered a wide range of facilities, locations and patient types within frail and elderly care planning. Despite this variety, the majority of papers fit into a few groupings as discussed previously, with a heavy focus on certain methods and locations.

2.2.3.1 Gaps in terms of Methodology

Across 62 papers, there were 12 different methods utilised, with a heavy focus on Markov and simulation. Research conducted between 1990 and 2015 has shown that the most common OR methods used in hospital applications were discrete event simulation and deterministic modelling (optimisation) [115, 116]. It was interesting to see the disparities between methods within frail and elderly care and general hospital applications. In the simulation method category (Table 2.3), only five papers used discrete event as their simulation method [80, 88, 101, 106, 108] and papers focusing on optimisation techniques were embedded within other methods [62, 83, 86, 89, 93, 102, 107, 108].

Within Abe et al.'s paper [116], the statement was made that the introduction of the Patient Protection and Affordable Care Act in 2010 (USA), has led to hospitals being required to improve quality of care and alongside this, there has also been an increase in the demand of services in USA hospitals. The seven papers that were based on data from the USA post 2010, focused on capacity planning and improving outcomes [60, 75, 83, 84, 85, 92, 94]. This suggests that the implementation of Government policy provides another avenue for research topics and should be closely monitored when identifying new areas to study.

Although healthcare data may lend itself well to some OR/MS methods explaining their higher frequency of papers (Markov, simulation, statistical analysis and machine learning), the remaining eight methods highlight the potential for further exploration into these fewer applied techniques. Soft OR methods, even though included in the search criteria, did not result in any papers being identified. This leaves potential for research using these techniques, such as soft systems methodology, and applying this to frail and elderly healthcare.

2.2.3.2 Gaps on the Intersection between Research Aims and Decision Levels

Successful care planning should consider long-term and day-to-day planning. The work has highlighted a large number of papers with strategic decision levels (long-term), with only 14 papers analysing tactical and operational approaches, 10 and four combined. Reviewing the 23 community care papers, three focused on tactical planning [93, 95, 100], and two on operational planning [90, 91]. Despite being long-term care facilities, the day-to-day planning of staff, resources and occupancy demands should be investigated for improving care planning. Such investigation could provide an interesting avenue to explore further, by comparing how factors vary day-to-day for private care companies compared to government funded elderly care services. Operational planning levels were addressed in single and not multiple hospital settings. One potential reason for this could be that when addressing multiple hospitals, the authors are more interested in strategic developments within care planning.

Another area examined was the research aims, which were able to be grouped into three streams for care planning: examining (23 papers), forecasting (28 papers) and improving (14 papers). The most popular aim, forecasting, mainly used simulation and Markov methods. Potentially the data required for forecasting techniques lends itself well to these methods, however, there were 11 papers that showed forecasting techniques can be used alongside different methods and therefore should be further explored [52, 54, 55, 70, 81, 86, 91, 93, 94, 95, 99].

2.2.3.3 Shortcomings on the Intersections between Medical Settings

Within the community care setting, there were only six papers which focused on the overlap between settings in the community [83, 84, 85, 86, 87, 88]. The remaining papers from this setting focused their research on a variety of settings including long-term care facilities, nursing homes, home care and day care.

When studying the application of the methods, the common research focus has been on capacity planning for community care services. This is important, as bed blocking often occurs when patients are medically fit to be discharged from hospital but there are insufficient places available within community care settings [101]. However, there has been little research into capacity planning for care of the elderly wards within hospitals. Patient flows [112], occupancy levels [79] and the LOS [82] are the main focus of research based in geriatric wards. These papers do not address future demands or predictions. This leaves the potential for capacity planning research within the care of the elderly wards, incorporating both short-term and long-term predicted demands. If capacities within these wards remain constant over time, with increased demand, these wards are likely to reach maximum capacity at a quicker rate. It is therefore important for research to focus on hospitals and how they feed into community care services. This overlap, between hospital and community care, will allow a successful integrated care system, similar to these nine papers [52, 53, 54, 55, 56, 57, 58, 59, 60].

2.2.3.4 Limitations of the Study

This review has provided a consolidation of care planning for frail and elderly across the care pathway, identifying gaps in research. A reproducible approach was given in relation to the search strategy in Table 2.1. We acknowledge that searching papers by keyword criteria can fail to identify relevant papers, which could be identified through other approaches. The keywords used were not an exhaustive list of all OR/MS approaches, patient flow terms or classification of patients, however, they provided a broad range of terms. To mitigate the number of papers excluded, reference lists and forward references of the initial 39 papers were included in the search. Similarly, only one search database was used (Scopus) to allow the results to be reproducible, however, this may have resulted in a small number of papers being excluded. To only include recent developments in OR/MS methods, this review restricted the literature to papers published in English between 2000 to 2020. The quality of the papers was not a factor in whether they should be included within the analysis. Despite the limitations of the study, the results have yielded some valuable findings which will be beneficial to both researchers and healthcare managers.

2.2.4 Literature Review Findings

This section has provided a categorisation framework for general, medical and methodological aspects of frail and elderly care planning literature, classifying 21 years' worth of research accordingly. The importance of bridging the gap between care of the elderly journals (GG) with HPS, IE, MI and OR/MS journals to consolidate papers with the focus of care planning for frail and elderly has been highlighted. As a result, we identified three overarching research possibilities:

1. When analysing the 12 methods utilised, nearly half of the papers focused on either Markov or simulation models. Although there were variations within the type of Markov model used, such as non-homogeneous discrete time and Coxian phase-type distributions, this leaves potential for further research to expand on the other methods discussed such as queueing models or routing. This could be further developed by combining multiple methods to create a more diverse model, as this review only identified two papers using this approach (Markov and simulation, and queueing theory with simulation). The features of the data which is routinely collected within community or hospital care settings might impact which OR/MS method is chosen. A further research possibility is the use of soft OR methods, which although included within the search terms discussed in Section 2.2.1 did not produce any relevant results.

2. More research would be beneficial in care planning across the care pathway for frail and elderly. Only nine papers were identified where the focus was on the combination of multiple hospitals and community care. Community care had the highest setting focus followed by single hospital settings. Whilst

it is important to consider these separately for improvements in efficiency, they may not be able to be implemented successfully without consideration for one another. There were eight papers identified that concentrated on a specific illness and tended to focus on general wards for care of the elderly. Hospitalised frail and elderly patients can be admitted to specialised wards for a specific medical condition or general geriatric wards. This means it can be difficult to plan for the next step of the pathway when they are medically fit to be discharged. To predict future demands and assist capacity planning in both hospitals and community care, it would be valuable to understand the complete patient journey from the first point of contact to discharge. Many frail and elderly care pathways do not differ from other patient pathways, and as a result, these groups are included in more general studies. However, frail and elderly patients suffer from more age-related issues, often with longer recovery times, so it is important to consider frail and elderly patient care planning separately for a successful healthcare system.

3. There was only one paper with research conducted on how systems would manage if a sudden rise in frail and elderly patients were to occur [98]. Sudden increase in demand is not a novel area to healthcare modelling with a high quantity of papers investigating this issue, e.g., intensive care units [42, 117, 118]. The Covid-19 pandemic has demonstrated why research is important to help healthcare providers meet increasing and sudden changes within demand [119]. Future research could investigate the effect Covid-19 has had within long-term care settings and the effectiveness of different Governmental policies for frail and elderly patients.

2.3 Hierarchical Prediction Models for Elderly and Frail Patients' Lengths of Stay

Hierarchical prediction models expand the flexibility of prediction models by accommodating grouped data. According to Luna [120], the term "hierarchical" is a general term for group structured models and describes a relationship in which the entities are grouped. In the literature, these are often referred to as multilevel, mixed effect or random effects models and are frequently used interchangeably.

The format of this section is as follows: The search criteria used to identify publications are introduced in Section 2.3.1. Section 2.3.2 systematically classifies the identified papers into general, medical and methodological contents. The research gaps discovered from the classification of the papers are described in Section 2.3.3. The findings of the review are then discussed in Section 2.3.4. Similar to the previous literature review, an appendix table has been supplied that lists the references for each figure that discusses a categorisation result. A complete list of the 90 papers that have been identified, together with each classification category, is provided in Table B.8 of the Appendix.

2.3.1 Methods

The Webster and Watson's [43] methodology, which was previously discussed in Section 2.2, was utilised to find relevant literature. To maintain consistency across the two literature reviews, the same methodology was applied. Scopus was also utilised to locate journal and conference proceedings articles of English-language publications.

Further to the previous literature review, an additional two years of papers were included in the criteria, therefore the search was performed from January 2000 to December 2022. Five Clarivate Journal Citation Report (JCR) categories were used: GG, HPS, IE, MI and OR/MS, along with upcoming journals not included in a JCR category, again similar to Section 2.2.

The same procedure as before was used to create a search string to ensure the search was not limited to a single methodology, journal or geographical area [43]. Therefore, at least one OR/MS prediction method, one length of stay (LOS) phrase and one age category, were included in the search string. The classification of people listed in Table 2.4 and those listed in Table 2.1 are identical, however, the other two columns are different.

The search string parameters used to find published studies on LOS modelling for frail and elderly patients are shown in Table 2.4.

2.3.1.1 Study Selection and Data Extraction

The initial search produced 943 papers. In order to eliminate publications that did not focus on frail and elderly patients, the papers underwent abstract analysis. Therefore, if within the abstract the papers did not mention the age of the patients within the study or if the age of the patients were not classified as elderly, the paper was excluded. Additionally, if the papers did not discuss the OR/MS method used, it was excluded. As a result of this, the number of papers were reduced to 66 papers. As discussed within Webster and Watson's research [43], and within Section 2.2.1, a forward and backward search was conducted. This resulted in a total number of 90 publications that were found to be relevant, including one literature review. A PRISMA flow diagram has been created and can be visualised in Figure 2.13. The study protocol remains the same as the first literature review, with a general

Prediction Method	Length of Stay	Classification of People
CART	Length of Stay	Elderly
Classification	LOS	Elderly care
Clustering	Patient Stay	Frail*
Data Mining		Geriatr*
Decision Tree$		Home care
Forecast*		Long term care
Hierarchical		Nurs* care
Linear Regression		Old people
Logistic Regression		Older people
Multilevel		Palliative care
Mixed Effect		>65
Naive Bayesian		
Neural Network$		
Predictive Model*		
Random Effect$		
Random Forest$		
Regression		
Support Vector Machine$		
SVM		
Time Series Analy*		
XGBoost		

Table 2.4: Scopus search terms for hierarchical prediction literature search to include at least one ‘Prediction Method’, one ‘Length of Stay’ term and one ‘Classification of People’ term .

classification being provided, the medical context discussed and then the planning decisions and types of OR/MS methods applied.

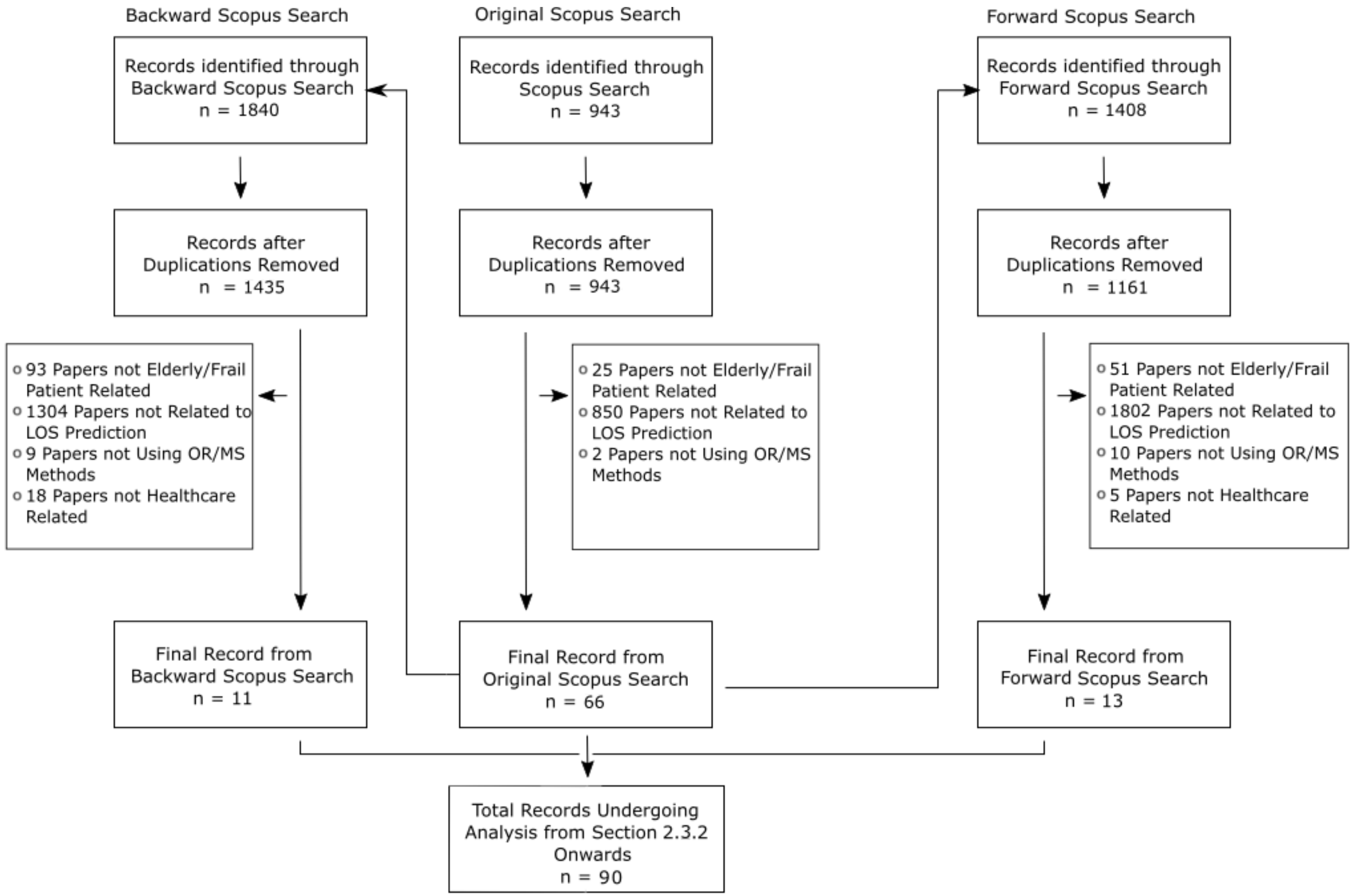


Figure 2.13: The PRISMA flow diagram for the hierarchical prediction literature review detailing the Scopus searches, the number of abstracts screened, the reason for exclusion and the number of full texts retrieved.

Through the search parameters, no literature reviews were identified. Through an additional separate search, Hunt-O'Connors et al. [121] literature review was identified. It should be noted that because the Journal of Nursing Management is not listed in one of the five chosen JCR categories, this was not found in the Scopus search. Additionally, it does not focus specifically on frail and elderly patients. In their assessment of 181 studies, Hunt-O'Connors et al. [121] examined the impact discharge planning on LOS and readmission rates of older adults in acute hospitals. Interestingly, the authors findings suggest that discharge planning did not have a statistically significant difference on LOS. One study, Pellett [122], revealed that medical and nursing teams in the UK had a strong commitment in promoting discharge planning. This ultimately led to a small reduction in LOS. One conclusion Hunt-O'Connor et al. [121] alluded to was that because of heterogeneity, studies could not be easily replicated. As a result, the methodology and findings in this thesis will be straightforward to replicate and modify.

In the following section, the 90 papers identified through the search string will be subjected to the same study protocol as previously outlined.

2.3.2 Results

The 90 papers that were found to be relevant can be divided into general, medical, and methodological themes to help determine what the gaps in the current body of literature are. Summary statistics outlining the results are provided in the sections that follow.

2.3.2.1 General Classification

The country and JCR category where the research was conducted are shown in Table 2.5. Despite the inclusion of five JCR categories, only four categories yielded results. The categories' sizes differed by varying degrees, with the GG category accounting for 72% of publications. This would imply that instead of other healthcare journals, this topic of study typically tends to be published in specialised journals for the frail and elderly journals. The 'Other' category is generated from papers from the forward and backward, which did not have the JCR category as a limitation.

The findings demonstrate the variety of places where this research has been published. However, even in nations where English is not the first language (such as Italy, Japan, and Germany), English-language publications are still produced.

Similar to the previous literature review, IE and MI JCR categories had the fewest articles (with the exception of OR/MS which yielded zero results). These publications should still be considered in the analysis because they offer different aspects

Country	UK	USA	Australia	Italy	France	Canada	China	Japan	Finland	Germany
GG	8	8	6	8	4	4	6	5	5	3
HPS	3	0	1	0	0	0	0	1	0	0
IE	0	0	0	0	1	0	0	0	0	0
MI	0	0	0	0	0	0	1	0	0	0
Other	2	2	1	0	2	2	0	0	0	1
Total	13	10	8	8	7	6	6	6	5	4
Country	**Ireland**	**Brazil**	**Israel**	**Singapore**	**South Korea**	**Sweden**	**Netherlands**	**Jordan**	**Spain**	**Total**
GG	2	1	2	0	1	1	2	0	0	65
HPS	1	0	0	0	1	0	0	0	0	7
IE	0	0	0	0	0	0	0	0	0	1
MI	0	0	0	0	0	0	0	1	0	2
Other	0	1	0	2	0	1	0	0	1	15
Total	3	2	2	2	2	2	2	1	1	90

Table 2.5: Number of papers which fall into each JCR category and the location of the where the research was conducted.

of LOS modelling with different focuses.

Figure 2.14 illustrates the number of papers published broken down into three year intervals. The period 2018 to 2020 yielded the highest number of papers with 32% of research being published within this time frame. The data shows varying numbers of papers published during each period, with a consistent level of 13 papers published between 2009-2011, 2012-2014 and 2015-2017.

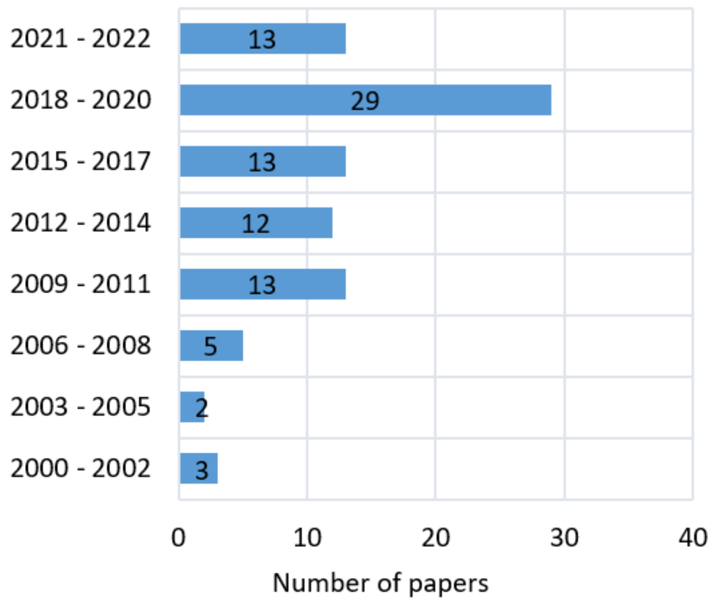


Figure 2.14: Bar chart of the 90 papers identified through the Scopus search, by publication year.

2.3.2.2 Medical Context

This subsection examines the articles medical background, which can be grouped into two categories: the patient's condition area and the treatment's medical environment. Recall the definition of medical setting, which refers to the site where care is being provided, and condition area, which refers to whether a patient's medical condition was short-term or long-term.

To understand where these individuals were receiving care, three medical settings were examined. These are listed below:

- Single Hospital
- Multiple Hospitals
- Community Care

Figure 2.15 displays the results for the number of publications within each care setting. Single hospital settings were the most common with 54 papers (60%). These papers concentrated on hospital wards, such as cardiac wards [123], emergency departments (EDs) [124] and geriatric wards [125]. Only five papers, [87, 126, 127, 128, 129], were based in the community. Cai et al. [126], Hoben et al. [127] and Welberry et al. [87] evaluated the LOS in nursing homes based on several factors, such as prior home care and policy variations. Johnson et al. [128] discussed the differences in location before being admitted to hospices and the effect this has on LOS. Park et al. [129] determined the effects of race and ethnicity in the LOS within hospice care.

Interestingly, there were only three papers which explored hospital and community care settings simultaneously. Fan et al. [130] used frailty as a predictor to determine the usage of healthcare, including LOS inside hospitals. Gordon et al. [65] applied Coxian phase-type distributions to predict patient LOS in hospital and across community care. Walsh et al. [131] analysed whether formal home care decreased LOS in hospitals. There was no overlap between multiple hospitals and community care. It is important for this area to be considered to create a robust and holistic pathway. This therefore suggests a possible direction for further study by using prediction models to calculate LOS across various services.

Another aspect to investigate is the condition area of a patient. There are three types of conditions included in this group: acute, chronic, and surgical. There are studies based on surgery in addition to the literature review covered in Section 2.2. Surgery papers are characterised as, medical disorders for which particular surgical procedures are used.

Figure 2.16 displays the quantity of papers in each medical condition category.

Chronic conditions were the most populous area of research yielded 38 papers of relevance. Since elderly and frail patients suffer with more chronic conditions than the general adult population, the higher number of chronic papers compared to acute is expected. ED's [124], strokes [132], hip fractures [133], and other acute

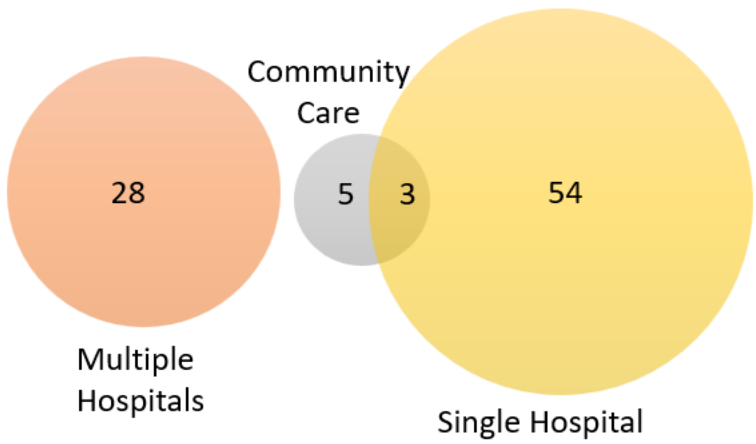


Figure 2.15: Venn diagram of the 90 papers identified through the Scopus search, by medical setting.

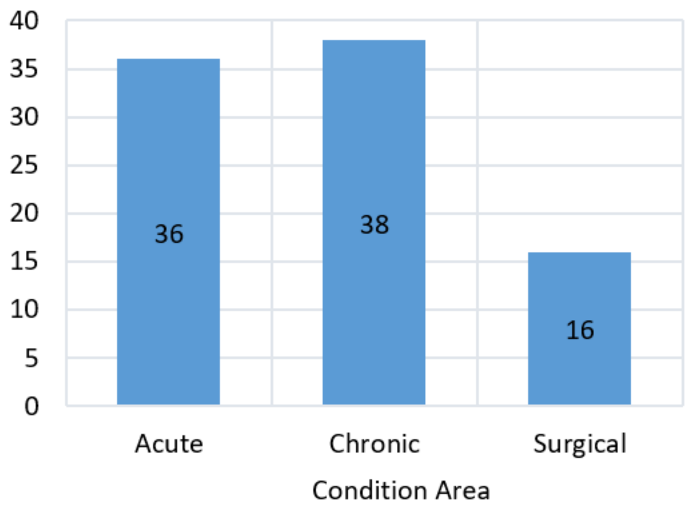


Figure 2.16: Bar chart of the 90 papers identified through the Scopus search, by 'Acute', 'Chronic', and 'Surgical' medical conditions.

medical conditions were the second most common condition area with 36 papers. If these conditions are not treated, they can frequently lead to more serious medical problems. Finally, there were 16 surgical articles focusing on various sub-specialties of surgery were included: Abdominal [134], cancer [135], cardiac [123, 136, 137, 138], colorectal [139], hip [140, 141] and the combination of hip and knee [142]. In six of the publications [143, 144, 145, 146, 147, 148], the surgical specialty was not specifically addressed.

2.3.2.3 Methodological Content

This subsection will analyse the technical aspect of the papers such as the goals of the research, the planning decision and the various LOS prediction techniques.

2.3.2.3.1 Research Aims

The papers can be categorised to define the direction of the research using the three main categories of research aim: examining, forecasting, and improving.

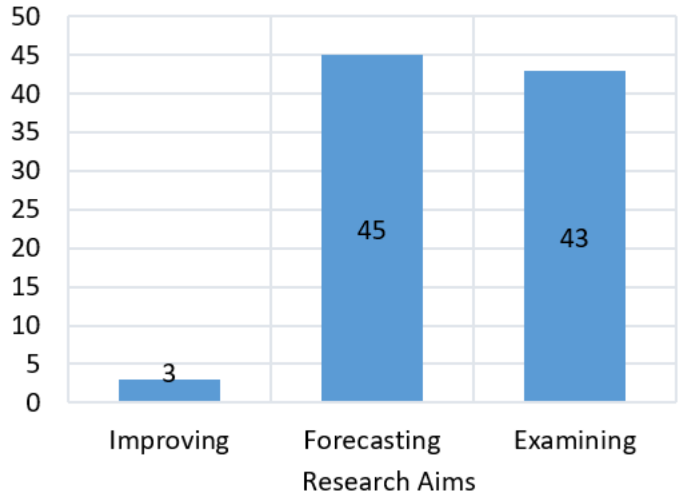


Figure 2.17: Bar chart of the 90 papers identified through the Scopus search, by research focus.

Figure 2.17 displays the quantity of papers in each of the research aim categories. Abe et al. [70] using both examining and forecasting research aims by determining the effects of reducing polypharmacy on LOS on gastrointestinal surgery patients.

The findings indicate that, with 50% of publications, forecasting is the most prevalent study goal. In these 45 studies, 25 were only concerned with LOS prediction [65, 70, 111, 123, 127, 132, 134, 138, 141, 148, 149, 150, 151, 152, 153, 154, 155, 156, 157, 158, 159, 160, 161, 162, 163]. The remaining 20 articles predicted other variables in addition to LOS such as cost [164, 165], healthcare utilisation [130], mortality [124, 164, 166] and hospital readmission [124, 133, 167].

Only three publications examined how to improve LOS [113, 131, 168]. According to Basic and Khoo [168], the discovery of novel medical diagnoses would have an impact on funding models based on diagnosis-related groups (DRGs), but it might also improve patient care and their associated LOS. Hamdani et al. [113] examined the use of Markov chains to model a hospital and the flow of elderly patients. Their model is then used to assess the effectiveness of intra-hospital care and predicting LOS.

2.3.2.3.2 Planning Decisions

Within Hulshof et al's. [44] research on the taxonomic classification in healthcare systems, three different decision levels were discussed: strategic, tactical and operational.

Figure 2.18 displays the number of publications by planning decision level.

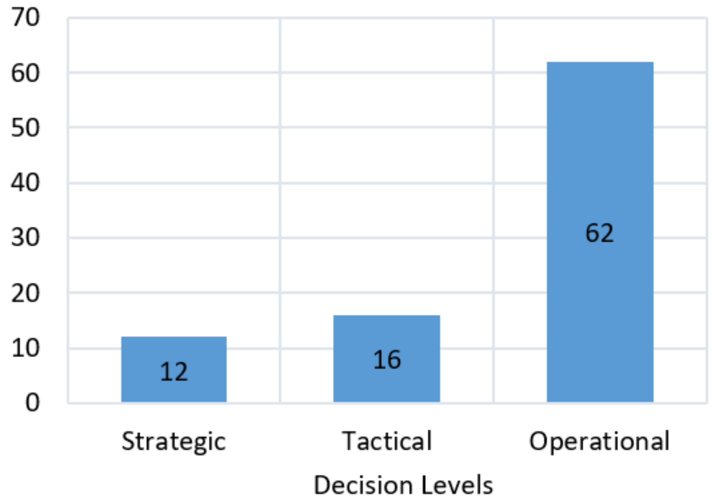


Figure 2.18: Bar chart of the 90 papers identified through the Scopus search, by planning decision level.

The majority of papers focused on operational planning, the day-to-day running of units. This is primarily due to the fact that the publications' main study goals involve predicting a patient's LOS in the hospital, and the authors do not elaborate on this information to aid with long-term planning.

There were only 12 articles that made long-term planning decisions (strategic planning). This suggests that prediction modelling is more suited to short-term, day-to-day decisions rather than long-term, wider policy decisions. The research by Hoben et al. [127] is an illustration of a strategically categorised paper. The authors focused on LOS in nursing homes across three distinct Canadian regions. They investigated how LOS varied based on various regional policies as well as different characteristics of patients.

Figure 2.19 displays the cross analysis between the planning decision and the medical setting. Despite having the fewest number of papers, strategic and tactical planning methods were the only planning decision to be addressed across all medical settings. Operational planning was divided into papers based on single hospitals (47%) and papers based in multiple hospitals (21%). Only one paper [65], had an operational planning level across hospital and community settings. These findings demonstrate the applicability of prediction modelling in a variety of care settings, including short-term and long-term care planning.

2.3.2.3.3 Prediction Methods

The final area to investigate was the prediction methods that have been utilised.

Figure 2.20 displays the quantity of each of the OR/MS methods. The methods

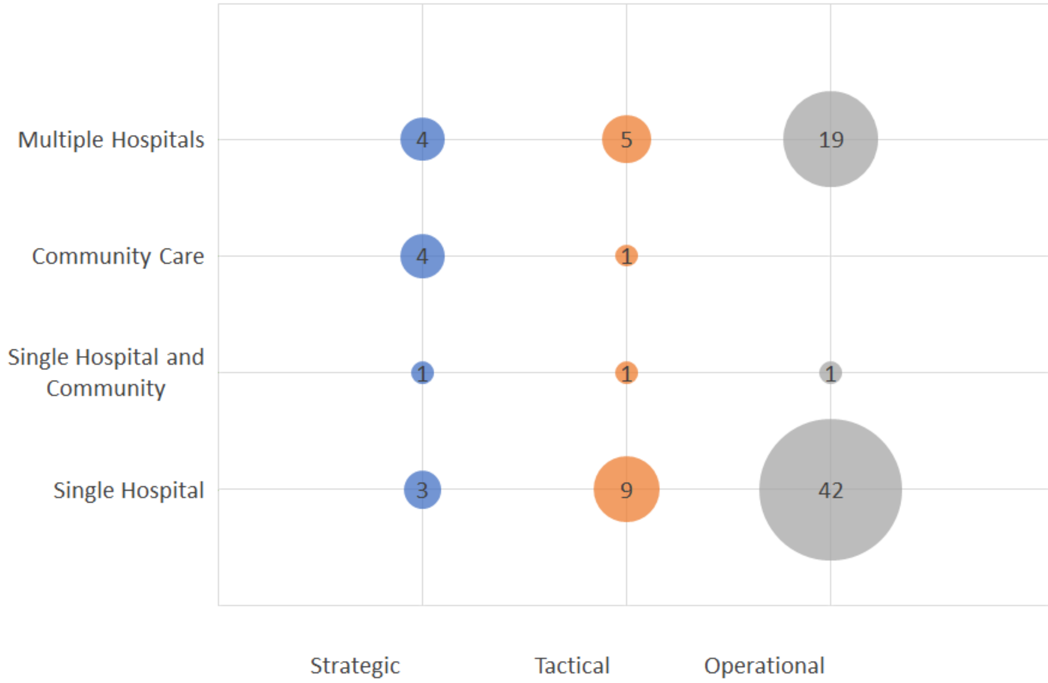


Figure 2.19: Cross analysis of planning decision and medical setting of the 90 papers identified through the Scopus search.

shown vary from those discussed in the first literature review (Section 2.3.2.3.3, as due to the variety of approaches within the first, larger grouping categories were created. For instance, the machine learning category encompassed logistic and linear regressions, however, because these were the most popular techniques in the second literature review these were classified separately. Similarly, Cox regression was grouped within the statistical analysis, however, produced 15 papers of relevance in the second review and therefore was included as its own category.

Linear and Logistic regressions were the most prevalent techniques with 18 and 36 studies, respectively. Four articles, [160, 169, 170, 171], used both linear and logistic regression models in their research. Only two publications employed traditional hierarchical OR methods such as decision trees and CART [149, 161]. In order to ascertain whether there was a relationship between LOS and preventable readmissions, Alyahya et al. [149] developed decision trees. The authors discovered a direct correlation. Within the field of internal medicine, the authors were able to advise clinical decision-makers on the recommended length of hospitalisation for patients. Nishino et al'.s [161] research involved building CART models to forecast long LOS's. Systolic blood pressure (155 mmHg) and serum albumin (3.4 g/dL, a blood protein present in albumin) readings were identified as lower bounds into predicting long LOS's.

There were three papers that used regression techniques that did not involve Cox, linear or logistic regression models. Multivariate multilevel regression was utilised by

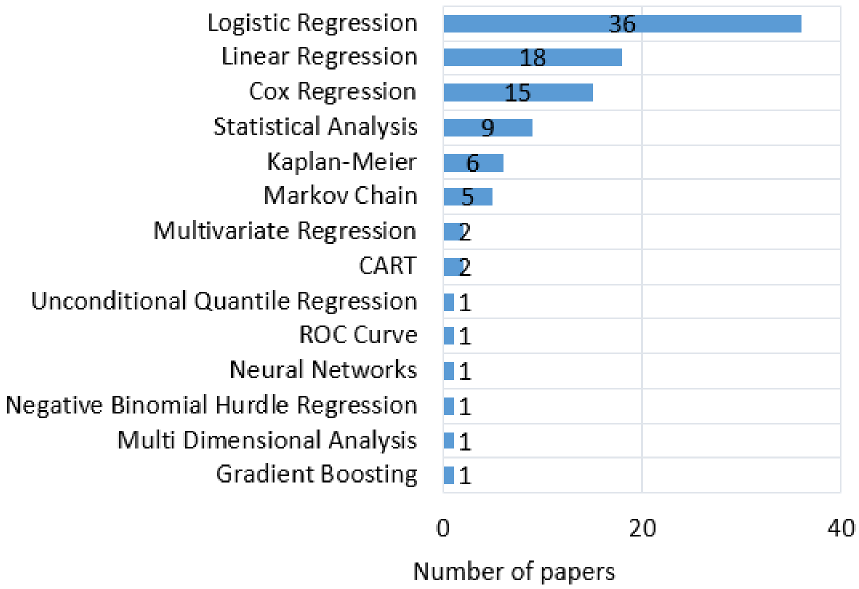


Figure 2.20: Bar chart of the 90 papers identified through the Scopus search, by mathematical method.

Chung et al. [172] to identify the variables that affected patients LOS in psychiatric wards. The type of medical institution, patient diagnosis, type of health insurance provider, and patient makeup of medical institutions were all shown to be significant determinants by the authors. Negative binomial hurdle regression models were used by Motzek et al. [173] to establish that longer hospital stays in Germany were caused by greater hospitalisation rates for dementia patients. Using unconditional quantile regression, Walsh et al. [131] identified a relationship between hospital LOS and formal home care.

Beauchet et al. [152] utilised three methods within their research: cox and logistic regression with Kaplan-Meier models. The impact of several characteristics, such as a history of falls and temporal disorientation, on long hospital LOS's was examined by the authors.

Table 2.6 refers to the prediction method used by each paper against its corresponding care setting. Since seven papers make use of various approaches, there are a total of 67 papers. Amongst the methods where there are five or greater papers published, a range of different care settings are used for the research. This demonstrates that, regardless of the situation, numerous strategies can be used in LOS modelling for frail and elderly patients. CART models are only used within single hospitals, and provides potential for research to focus on other care settings when using CART. Seven different methodologies were used in both single and multiple hospital settings, therefore having the most variety of approaches.

	Single Hospital	**Single Hospital and Community**	**Community Care**	**Multiple Hospitals**	**Total**
Logistic Regression	[70, 140, 144, 146, 147, 148, 151, 152, 155, 156, 159, 165, 169, 174, 175, 176, 177, 178, 179]	[130]	[128]	[133, 137, 139, 142, 143, 150, 160, 163, 170, 171, 180, 181, 182, 183, 184]	36
Linear Regression	[123, 135, 136, 138, 140, 141, 169, 185, 186, 187, 188]			[160, 170, 171, 189, 190, 191, 192]	18
Cox Regression	[132, 145, 152, 168, 174, 193, 194, 195]		[87, 126, 127, 129]	[166, 196, 197]	15
Statistical Analysis	[125, 134, 154, 198, 199, 200, 201, 202]			[203]	10
Kaplan-Meier	[124, 152, 162, 167]		[128]	[164]	6
Markov Chain	[111, 113, 204]	[65]		[205]	5
Multivariate Regression	[206]			[172]	2
CART	[149, 161]				2
Unconditional Quantile Regression		[131]			1
ROC Curve	[207]				1
Neural Networks	[158]				1
Negative Binomial Hurdle Regression				[173]	1
Multi Dimensional Analysis	[153]				1
Gradient Boosting				[157]	1
Total	59	3	6	31	99

Table 2.6: Number of papers which fall into each medical setting and OR/MS method within the published research.

Note: Adamis et al. [169], Basic et al. [174], Beauchet et al. [152], Johnson et al. [128], Lisk et al. [160], Motohashi et al. [170], Naouri et al. [171], all utilise multiple methods and therefore appear multiple times within the table. This resulted in a total of 99 publications.

2.3.2.4 Common Themes

Section 2.3.2 provided an overview of prediction modelling for frail and elderly patients LOS in both hospitals and community care. Logistic regression was the most widely utilised method, with 36 papers developing these types of models. The last five years have seen the publication of more sophisticated techniques like CART, which are frequently derived from logistic and linear regression models. Operational planning was the most popular form of planning, indicating that prediction modelling is simpler to implement for daily operations of units and patient care. The papers include a variety of research objectives, with forecasting approaches being the most prevalent and frequently used to forecast patients LOS. Finally, hospitals have served as the primary environment for these publications, either based in a single hospital ward or analysed across multiple hospitals. The research identified through the Scopus search has demonstrated that a wide variety of OR/MS techniques can be applied to the care of frail and elderly LOS prediction modelling. Section 2.3.3 will discuss the areas for future research.

2.3.3 Research Gaps

Within ageing prediction modelling, the 90 studies found in this study addressed a wide range of facilities, regions, and patient types. Three significant gaps still exist in the literature, which could be the subject of future study.

2.3.3.1 Gaps in terms of Methodology

The 90 publications used 14 distinct techniques, with a significant emphasis on linear and logistic regression. This is perhaps because the data lends itself nicely to these prediction techniques. There were in total eight approaches that accounted for 11% of the paper methods, which used alternative prediction methods, indicating the potential for further research into these less commonly used techniques. The topic of frail and elderly prediction modelling underutilised more sophisticated and complex techniques like CART. In comparison to linear and logistic regression, CART models provide a deeper understanding into the data. Neural networks only produced one relevant study, [158], and clustering produced no relevant results in spite of being included in the Scopus search string in Table 2.4. Future studies may choose to concentrate on these hierarchical techniques, such as CART, clustering, and neural networks.

Patients with long LOS's typically account for a high proportion of overall bed days [208]. There are no laws or regulations regarding patient LOS, and it should be determined by when the patient is clinically fit for discharge. Bed occupancy has consistently exceeded 91% in Canada [209], 85% in England [210] and 66% in the

USA [211]. There were only 17 papers which specifically addressed predicting the longer LOS in hospitals [70, 123, 126, 136, 142, 152, 153, 155, 158, 159, 170, 178, 182, 185, 194, 199, 201, 203]. Since trends in bed occupancy rates are still rising, particularly following the Covid-19 outbreak, projecting lengthy hospital LOS's and enacting policies to shorten them may lead to a decrease in occupancy rates.

2.3.3.2 Gaps on the Intersection between Research Aims and Decision Levels

Both the long-term and day-to-day planning scenarios should be taken into account for effective care planning. Only 28 publications examined the strategic and tactical decision levels, compared to the 62 papers that concentrated on the operational decision level. Decision-makers do not take into account how to best plan for the future because they are primarily concerned with day-to-day planning. Future, increased needs would not be satisfied as a result of this. All three locations for care, as well as the overlap between hospital wards and community care, were taken into account when determining strategic planning levels. On some level, this does indicate that all care facilities are being taken into account and analysed for the long-term horizon of LOS prediction modelling for frail and elderly patients. Policy makers will be able to make educated decisions based on these research papers regarding how to adapt units to future clinical and demographic changes.

The research's objectives were also evaluated and divided into three groups: examining (43 papers), forecasting (45 papers) and improving (3 papers). In total 13 of the 14 different methods were used to achieve the most popular study goal, forecasting. With 25 publications, forecasting was primarily employed to predict only LOS, with 20 papers demonstrating how LOS prediction can support care planning by working in conjunction with other predictors. Only three studies, all of which took place in a single hospital, were related to the improving research aim [113, 131, 168]. Although it does open up a new line of inquiry for researchers, it may also imply that the study methods these publications examined were simpler to apply to single settings.

2.3.3.3 Shortcomings on the Intersections between Medical Settings

There were only eight studies that addressed LOS prediction modelling in the context of community care [65, 87, 126, 127, 128, 129, 130, 131]. This shows that LOS prediction modelling may not be well adapted to community care settings, given that patients are often unlikely to return to their own homes after being admitted to a nursing or care facility. Fan et al's. [130] and Walsh et al.'s [131] research using a single hospital and community care setting were the sole crossovers between community and hospital settings. This presents an opportunity to examine if patients who receive some form of community care have different LOS's when admitted. Two potential hypotheses could be tested: whether patients are discharged sooner because they have a home to return to in the community and do not bed block, or because these patients are typically more unwell and take longer to recover as they are already receiving care and hence stay in hospital for longer.

In total, 28 studies addressed multiple hospital settings. Although being able to forecast LOS at the time of admission can help with resource allocation for a patient's hospital stay, none of the publications were found to do more than merely investigate the variables that influence LOS. The development of these LOS prediction models and their use in resource planning would fill a significant research gap. There were 11 of the 28 multiple hospital papers, [160, 166, 171, 180, 181, 184, 190, 191, 192, 197, 205], did not concentrate on a particular disease but instead analysed the elderly admission and were able to compare LOS based on various illnesses. These publications do not, however, examine or make forecasts for the future on a bigger scale. This leaves open the possibility for future studies to examine the effects of various medical conditions on patient demands and LOS, as well as how these factors may alter over time based on clinical and demographic changes.

2.3.4 Literature Review Findings

This section has established a framework for categorising general, medical, and methodological components of prediction modelling for frail and elderly patients. The literature on healthcare has been categorised for a total of 23 years. The significance of bridging the gap between GG journals and more conventional health mathematics journals (HPS, IE, and MI) has been emphasised, similar to the literature review in Section 2.2. As a consequence, the same three overarching research opportunities, as Section 2.2, have been identified.

1. The use of linear and logistic regression was the focus of 55% of the publications. Despite the fact that they were employed in conjunction with other techniques, this leaves potential for further study to build on other methods techniques such as CART. Given that our study only found six publications utilising this strategy, it might be further developed by integrating other strategies to produce a more diverse model. The type of prediction method employed may vary depending on the environment and types of data that are regularly collected. Another avenue for investigation is the application of other hierarchical techniques, such as neural networks or clustering, which, despite being part of the Scopus search term, only yielded one outcome.

2. Predicting LOS for frail and elderly patients across the care pathway would benefit from more research. Only three studies with a single hospital and

community care as their primary settings were identified. Due to the bed blockage issue that many healthcare facilities are experiencing, it is crucial that hospitals and community care are integrated. When a patient is medically fit for release but cannot be transferred because there is nowhere for them to go, this is known as bed blocking. This frequently happens as a result of the time it takes to arrange for people who need home care or who need to be admitted to a nursing or care facility. As a result, this leads to longer LOS in hospitals and this cannot be resolved without adequate care in the community. This is supported by the discovery of just nine studies with a specific focus on predicting longer LOS's within hospitals.

3. The final research possibility focuses on the increased demands and pressures that healthcare facilities are currently experiencing. None of the publications addressed the possibility that hospital LOS would rise in response to rising demand for admissions of frail and elderly patients. First of all, if there is a greater influx of patients, staff would have to care for more patients and as a result become overworked, which might significantly affect patients recovery times. Additionally, because there would be more demand for inpatient treatments like radiology or pathology, patients would have to wait longer for these examinations, extending their LOS.

2.4 Overlap Between the Two Literature Reviews

Two literature reviews were studied for this research. Firstly, by conducting a broad literature review in Section 2.2, we gained a comprehensive understanding of the OR/MS methods applied to frail and elderly patient care planning. This foundational knowledge allowed us to better contextualise and appreciate the specific hierarchical approaches examined in Section 2.3 for predicting LOS for these patients. The second review, with its more in-depth analysis, allowed us to delve deeper into the predictive methodologies, enabling us to identify key insights and nuances that may have been overlooked in a single, all-encompassing review. Due to the nature of the studies, there was an overlap of the OR/MS methods and therefore there was cross over between the two reviews.

There were four papers which appeared in both literature reviews: Abe et al. [70], Hamdami et al. [113], Marshall and McClean [111] and Welberry et al. [87]. Abe et al. [70] focused on investigating the influence of polypharmacy on LOS for gastrointestinal surgery patients. Their study employed logistic regression to analyse the relationship. Hamdani et al. [113] utilised Markov chains to track and predict the movement through the Hospital Center of Roanne in France, offering valuable insights into patient flow and LOS dynamics. Marshall and McClean [111] also employed Markov chains to explore the characteristics of patients that affect their LOS within hospitals, providing a deeper understanding of LOS variability. In a different context, Welberry et al. [87] discussed the impact on prior home care on the LOS for dementia patients in residential care. To achieve this, they utilised Cox's regression analysis, yielding valuable findings on the subject. This cross-referencing provided an additional layer of validation for their significance and reinforced the relevance of their contributions to the field.

Furthermore, exploring the OR/MS methods from different angles and within various contexts allowed connections to be made. By doing so, a more comprehensive understanding of the literature could be built, ultimately contributing to a more well-rounded and informed research study that can offer practical implications for improving the planning and care of frail and elderly patients in healthcare settings.

2.5 Summary

The practice of OR/MS approaches in the planning of care for the frail and elderly was the main topic of this chapter's literature reviews. The underutilisation of OR/MS techniques, the absence of comprehensive holistic care planning, and the implications of increases in demand on healthcare systems have all been noted as gaps in the literature. Within this thesis, we seek to address all three aspects by using underutilised OR/MS methods and applying them to multiple hospitals in Chapter 5 and discussing growing demand in Chapter 6 using scenario analysis.

The application of OR/MS methods to frail and elderly patients literature review (Section 2.2) included a higher quantity of strategic level publications, hence the two literature reviews did dispute in terms of planning decisions. There was a stronger emphasis on operational planning choices in the literature review for the hierarchical prediction models for patients' LOS literature review (Section 2.3). Therefore, all three planning decisions will be examined in this research, including how to plan the day-to-day running of units while also taking into consideration long-term demands and predictions for frail and elderly patients.

In the following chapter, Chapter 3, we discuss a number of the predictive analytical methods identified within the literature reviews. Through a realistic, simplified example illustrating the procedures and outcomes, the methods will be applied to the case of frail and elderly patients.

Chapter 3

Predictive Analytics to Support Capacity Planning for Frail and Elderly Patients

3.1 Introduction

Within industry today, OR and predictive analytics are allowing companies to make more informed decisions about their businesses, with the effects of new policies to be understood prior to their implementation. The use of analytics has also developed rapidly within healthcare in the 21^{st} century, and with the NHS being one of the UK's and Europe's largest employers, many individuals are now active as healthcare consumers or healthcare providers [212]. The need for new, more sophisticated technologies and recent demographical changes, such as an ageing population, are the main causes of the rising expenses associated with healthcare treatment [213]. Patients, as consumers, are no longer prepared to accept subpar services and expect: shorter waiting times, increased appointment availability and faster treatment times [214]. By using predictive analytics within healthcare, it allows costs to be minimised and performance to be optimised within the NHS. Predictive analytics uses statistics and modelling techniques to make predictions about future events [215]. The analysis of both recent and historical data can help healthcare systems become more dynamic and proactive by looking forward to detect patterns or behaviour.

<u>Research Aim</u> - This chapter will identify predictive techniques which will be applied to the frail and elderly case study within ABUHB, to determine LOS within hospitals. The theory discussed in this chapter will be applied to address the following research aim, 'How do the clinical and demographical attributes of frail and elderly patients affect their length of stay within hospital?' to be addressed later in this thesis.

Predictive and forecasting techniques were the most frequent study objectives within the healthcare literature, as discussed in Chapter 2. The studies, however, tended to accomplish this via more traditional OR/MS techniques, such as simulation and Markov chains. The goal of this thesis is to use predictive approaches that are not typically used in healthcare to see if similar patient groups can be found for the frail and elderly.

The chapter is structured as follows: Section 3.2 will go through the prediction techniques utilised in this thesis, while Section 3.3 will discuss classification and regression trees (CART), as well as the extension of random forest metrics. Then, in Section 3.4 we will determine how these analytical methods can be implemented within Python. A practical example will be demonstrated throughout the chapter to provide a greater understanding of the theory.

3.2 Predictive Techniques

Predictive analytics extracts information from existing data using algorithms and machine learning to forecast future events. Predictive analytics can be divided into two subcategories: supervised learning and unsupervised learning.

3.2.1 Differences Between Supervised and Unsupervised Learning

To train algorithms for data classification or outcome prediction, supervised learning makes use of labelled data sets. During a 'training phase', when data is inputted into the model, patterns within the data are identified and connections between dependent and independent variables are found. In order to demonstrate this relationship, a mathematical equation is created. The training process can produce more accurate outcomes if more data and information are fed into it. Cross-validation then occurs on a 'testing' data set which the model has not been trained upon, to assess the accuracy and reliability of the model.

Table 3.1, contains a collection of the most often used supervised learning methods applied within healthcare, along with relevant healthcare references.

Using supervised learning has various advantages, including improved accuracy levels because the desired result is already known. Supervised approaches are very simple for practitioners without a mathematical background to comprehend and apply in the context of healthcare. However, supervised learning models always re-

Method	**Healthcare Examples**
Regression	Heart Failure Outcomes [72], LOS Prediction [70, 87]
Classification	Ambulance Delay [216], Blood Testing & Donation [217, 218]
Naive Bayesian Model	Diabetes Prediction [219], Heart Disease Detection [220, 221]
Random Forest Model	Cancer Prediction [222], Diabetes Detection [223, 224]
Neural Networks	Biomedicine [225, 226], Healthcare Access [227]
Support Vector Machines	Dementia Prediction [228], Stroke Patients [229, 230]

Table 3.1: Supervised learning methods and their applications across healthcare.

quire updating in order to keep up with evolving data. The model must repeat the training and validation phases whenever new data variables are added in order to retrain with the new variables. These models can take a long time to compute, especially when employing large amounts of data. If there is unnecessary data included, efficiency may also suffer.

Unsupervised learning, in contrast to supervised learning, does not use labelled data to learn from, instead, the algorithm identifies clusters of similar patterns or groupings. The algorithm will choose the optimal number of classes the data should be divided into because it is unlabelled, therefore it does not need to go through the training and testing stages that supervised learning requires.

The most popular unsupervised learning techniques are shown in Table 3.2, along with real-world healthcare applications. Each technique may incorporate a variety of techniques, for instance, clustering may include, ‘K-Means’, ‘hierarchical’ and ‘probabilistic’ techniques.

Method	**Healthcare Examples**
Anomaly Detection	Heart Rate Anomalies [231, 232], Lung Cancer [233]
Clustering	Breast Cancer Recurrence [234], Heart Disease [235], Thyroid Cancer [236]
Dimensionality Reduction	Alzheimer’s Diagnosis [94, 237], Diabetic Retinopathy Detection [238]

Table 3.2: Unsupervised learning methods and their applications across healthcare.

When used in the healthcare industry, unsupervised learning offers a host of advantages, particularly for data that naturally lends itself to labelling. By deciding on the groupings, the unsupervised learning algorithm can save clinicians critical time. Due to the dynamic nature of healthcare data, it is possible to use the algorithms to track how clusters change over time. However, because there is no expectation of labelled results, it is uncertain whether the yielded results will be useful. As a result, accuracy levels frequently fall short of those of supervised methods.

Five predictive approaches will be used within this research. The first two, linear and logistic regressions, were the most common methods identified through the literature search in Section 2.3.2.3.3. However, since these methods are prone to over fitting or over simplifying problems [239], more advanced techniques should be investigated. When investigating other more complex predictive algorithms, literature suggests that with support vector machines performance goes down with large data sets due to the increase in training time [239]. Naïve Bayesian models have the assumption that all features are independent [240] and neural networks have a limited ability to explicitly identify causal relationships [241]. Therefore these three methods would not be suitable for this research. Classification and regression tree (CART) models offer a more notable advantage in healthcare due to their interpretability [239]. By generating a visual decision tree, they simplify the understanding of model outcomes, making it accessible even to healthcare practitioners with limited mathematical knowledge. The final method investigated will be random forests, which are developed from classification and regression trees.

Within this research, supervised learning approaches will be examined and applied to the case study of frail and elderly patients. Prescriptive approaches will be used in conjunction with models for regression, classification, and random forests. Discussion into the application of unsupervised methods will take place within Chapter 8.

3.2.2 General Model Formulation

The following section will analyse the general formulation of each of the predictive methods. Each method will be accompanied by a worked example to demonstrate it can be effectively applied to a simplified patient data set. This approach will enable healthcare professionals to quickly implement these techniques in their own data analysis.

3.2.2.1 Linear Regression

Linear regression is used to establish the relationship between a continuous dependent variable and either continuous or discrete independent variables. There are two types of linear regression; simple linear regression in which there is only one independent variable, and multiple linear regression in which there are two or more independent variables.

In a regression problem, the key quantity is the mean value of the outcome, given the value of the independent variable or variables. This quantity is known as the conditional mean, expressed as '$E(Y|x)$' where Y denotes the outcome variable and x denotes a value for the independent variable. In simple linear regression, it is assumed that the conditional mean may be expressed as a linear equation in x as follows [242, 243]:

$$E(Y|x) = \beta_0 + \beta_1 x \tag{3.1}$$

where β_0 is the y-intercept and β_1 is the gradient of the line. It is also required for an error term to be included within the equation, denoted ϵ. This error term expresses an observation's deviation from the conditional mean. It is therefore assumed that an observation outcome of the variable may be expressed as [242, 243]:

$$y = E(Y|x) + \epsilon \tag{3.2}$$

where ϵ is the error term.

For multiple linear regression, where multiple independent variables are used to predict the dependent variable, Equation (3.1) can be modified to include these additional variables [242].

$$E(Y|x) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_n x_n \tag{3.3}$$

Therefore:

$$y = E(Y|x) + \epsilon \tag{3.4}$$

The findings demonstrate that the dependent variable will increase or decrease by a specific amount for every unit increase in the independent variables. There are four primary assumptions regarding the error value that must be met when using linear regression [244]:

- The error term values have a mean equal to zero
- The error terms have constant variance
- The error terms are normally distributed
- Each error term is independent from other error terms

According to the first three assumptions, the population of potential error term values for each given value of 4x has a normal distribution with a mean of zero and a variance of σ^2. The final assumption implies that the error terms do not follow a pattern and hence most likely to be independent of one another.

To establish the eligibility for linear regression, the null hypothesis must first be rejected, indicating that there is a relationship to be modelled. For a simple linear regression, the hypotheses are as follows:

$$\text{H}_0 = \text{There is no linear relationship (i.e., } \beta_1 = 0) \tag{3.5}$$

$$\text{H}_1 = \text{There is a linear relationship (i.e., } \beta_1 \neq 0) \tag{3.6}$$

These hypotheses can be extended for multiple linear regression:

$$\text{H}_0 = \text{There is no linear relationship (i.e., all } \beta_{1\ldots n} = 0) \quad (3.7)$$

$$\text{H}_1 = \text{There is a linear relationship (i.e., at least one } \beta_{1\ldots n} \neq 0) \quad (3.8)$$

where n is the sample size.

In order to estimate the parameters of the β terms, the method of ordinary least squares (OLS) is used. When calculating these terms, OLS first calculates the sum of the squared standard deviations of the observed values of Y before minimising the outcome.

3.2.2.1.1 Worked Example Linear regression can be used to find the relationship connection between a patient's LOS and the dependent variables. This worked example will utilise a data set of 10 patients to illustrate the practical application of linear regression. Two case studies will be demonstrated, highlighting the difference between categorical and continuous independent variables. Table 3.3 displays the 10 patients used within the analysis.

Patient Number	Age	Hospital	LOS	Specialty	Admission Method	Admission Source	Frailty Level
Patient 1	95	RGH	5	COTE	Emergency	Own Home	3
Patient 2	82	RGH	3	COTE	Emergency	Own Home	2
Patient 3	89	RGH	4	T&O	Emergency	Own Home	2
Patient 4	87	RGH	4	T&O	Elective	Own Home	2
Patient 5	85	NHH	3	COTE	Elective	Transferred	1
Patient 6	76	NHH	1	COTE	Elective	Transferred	1
Patient 7	71	NHH	1	T&O	Emergency	Transferred	1
Patient 8	96	RGH	5	T&O	Emergency	Own Home	3
Patient 9	70	NHH	1	COTE	Emergency	Transferred	1
Patient 10	67	NHH	1	T&O	Elective	Own Home	1

Table 3.3: Data set of 10 patients that will be used for the first worked example.

Age, with a range of 29 years, is one of the continuous independent variables. We can ascertain the relationship between LOS and age using linear regression. The outcomes of the linear regression are shown in Equation (3.9). It should be observed that the equation lacks a constant term. The LOS rises by 0.1494 days for every unit increase in age (Figure 3.1).

$$y = 0.1578x - 10.1052 \quad (3.9)$$

where x is the age of the patient. Therefore, if the age of a person was 70, their predicted LOS would be 0.9408 days.

For categorical variables, one-hot encoding is used which results in Equation (3.10),

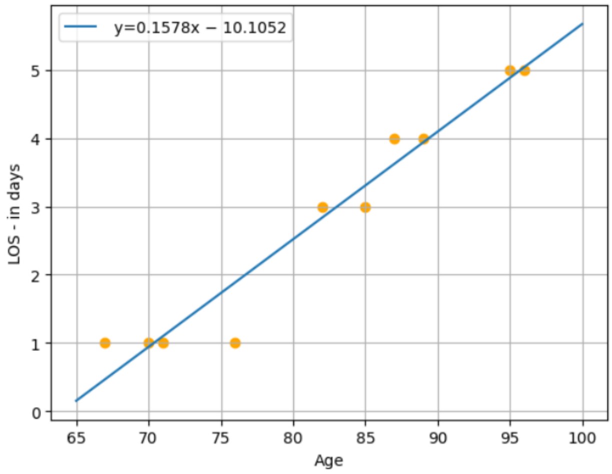


Figure 3.1: Worked example of linear regression model where the dependent variable y represents the Length of Stay (LOS) and the independent variable x is the age of the patient.

if using the variable hospital for prediction.

$$y = 1.6667 - 0.6667x_1 + 2.3333x_2 \tag{3.10}$$

where x_1 is equal to hospital NHH, and x_2 is equal to hospital RGH. Therefore, if NHH is attended, then the predicted LOS is 1 day, otherwise RGH is attended and the LOS is 4 days.

3.2.2.2 Logistic Regression

Dichotomous data is data that has two possible outcomes, data with a binary response in which each participant can only belong to one of the two categories. Applying the conditional mean to dichotomous data requires that it be higher than or equal to zero and less than or equal to one, i.e., $0 \leq E(Y|x) \leq 1$. In order to forecast this binary result, logistic regression is applied. The quantity $\pi(x) = E(Y|x)$ is used to denote the conditional mean, simplifying the notation. This form of the logistic regression model can be seen within Equation (3.11) [242, 243].

$$\pi(x) = \frac{e^{\beta_0+\beta_1 x}}{1+e^{\beta_0+\beta_1 x}} \tag{3.11}$$

In order for a linear regression to be fit, a logit transformation of the dependent variable is required to be performed, defined in terms of $\pi(x)$ as follows [242]:

$$g(x) = \ln\left[\frac{\pi(x)}{1-\pi(x)}\right] \tag{3.12}$$

$$g(x) = \beta_0 + \beta_1 x \tag{3.13}$$

This transformation is important since $g(x)$ is linear in its parameters which may be continuous and range from $-\infty$ to $+\infty$, depending on the range of x. Similar to linear regression, logistic regression also requires an error term for the observation value outcome of the variable, expressed as:

$$y = \pi(x) + \epsilon \tag{3.14}$$

In order to determine the suitability for logistic regression, the null hypothesis is first required to be rejected, suggesting that there is a relationship. The first of these hypotheses is that if $y = 1$ then $\epsilon = 1 - \pi(x)$ with probability $\pi(x)$. Otherwise, $y = 0$ and therefore $\epsilon = -\pi(x)$ with probability $1 - \pi(x)$. Thus, ϵ has a distribution with a mean of zero and variance of equal to $\pi(x)[1 - \pi(x)]$. The conditional distribution of the outcome variable follows a binomial distribution with probability given by the conditional mean, $\pi(x)$.

The hypotheses for logistic regression are as follows:

$$\mathrm{H}_0 = \text{All coefficients are zero (i.e., all } \beta_{1\ldots n} = 0) \tag{3.15}$$

$$\mathrm{H}_1 = \text{All coefficients are zero (i.e., at least one } \beta_{1\ldots n} \neq 0) \tag{3.16}$$

where n is the sample size.

3.2.2.2.1 Worked Example A case study of 15 patients was used (an additional five rows compared to the previous worked example), to avoid the case of perfect separation. For the remainder of this section, the linear regression will be performed using the initial 10 patient table (Table 3.3), while the logistic regression will be performed with the following additional rows added (Table 3.4).

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

Table 3.4: Data set of 15 patients that will be used for the second worked example.

Given that logistic regression requires a grouped variable to predict outcomes, LOS can be divided into groups of '$<$ 4 days' and '$\geq$ 4 days, to identify the groupings of patients who have longer LOS's in hospital.

Performing the logistic regression with age as a continuous variable, Equation (3.17) is generated.

$$y = 0.2661x - 21.7946 \tag{3.17}$$

Therefore, for a patient who was aged younger than 82, they would fall into category '0' and their LOS would be predicted to be less than 4 days. Otherwise, for those aged 82 and over, they would be predicted to fall into the category of '1' and have a LOS equal to or greater than 4 days.

Performing with hospital as a categorical variable, results in the following:

$$y = 2.8904x_{RGH} - 1.7918 \tag{3.18}$$

Within this calculation, the intercept did not have a significant p-value (0.097) and therefore can be considered as zero.

This means, for patients who attend NHH hospital, the predicted grouping would be '0' and therefore have a LOS of less than 4 days. RGH patients would result in $y \geq 1$ and would therefore fall into group '1'.

3.2.3 Evaluation Metrics

In order to determine the success of using models against the data, traditional scoring techniques will be used for both the linear and logistic regressions.

Whilst there is no universal standard for scoring metrics, the rule of thumb is that the higher values indicate better performance. Typically, what defines a good accuracy level in machine learning is subjective to the type of data, field of study and the intended application of the model.

According to Ozili [245], a regression score ranging from 0.1 to 0.5 is considered acceptable when a considerable number of explanatory variables exhibit statistical significance. Conversely, the model must be rejected if all the explanatory variables lack statistical significance. A score value between 0.5 and 0.99 is acceptable when most of the explanatory variables demonstrate statistical significance. However, it is crucial to avoid the possibility of multicollinearity among the explanatory variables, as these factors could potentially inflate the value and lead to misleading conclusions. All of the evaluation metrics covered in this chapter will utilise the same score ranges.

For linear regression, two scoring measures will be used for the evaluation of the models:

1. R^2 Value

2. Adjusted R^2 Value

The theory behind these two scoring methods are discussed below.

3.2.3.1 R^2 Value

The R^2 is the coefficient of determination and calculates how good a model's fit is compared to a given data set. It indicates how close the predicted values are to the actual values.

$$R^2 = 1 - \frac{SS_{RES}}{SS_{TOT}} = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{y})^2} \tag{3.19}$$

where SS_{RES} is the sum of square of the residuals and SS_{TOT} is the total sum of squares. Further expanding into the formula for SS_{RES}, y_i is the observed variable value and $\hat{y}_i$ is the value estimated by the regression line. Similarly for SS_{TOT}, y_i is the observed variable value and $\bar{y}$ is the mean value. The range for the R^2 value is between -∞ to 1, where a negative value indicated the best fit line is performing worse than the average fit line.

3.2.3.1.1 Worked Example When calculating the R^2 value for the linear regressions performed in the worked example, we can use Equation (3.9). Starting with patient 1 who is aged 95, their recorded LOS was 5 with a predicted LOS of 4.8858 days. SS_{RES} can be calculated as follows:

$$\hat{y} = -10.1052 + 0.1578x \tag{3.9 revisited}$$

$$\hat{y} = -10.1052 + (0.1578 * 95) \approx 4.8858 \tag{3.20}$$

$$(y_i - \hat{y}_i)^2 = (5 - 4.8858)^2 = 0.0013 \tag{3.21}$$

This process is performed for all 10 patients, resulting in an SS_{RES} value of 1.3674.

$$\sum (y_i - \hat{y}_i)^2 = 1.3674 \tag{3.22}$$

Then we can calculate the SS_{TOT}, by using the equation $\bar{y} = \frac{\sum y}{n}$. Since there are 10 samples within the data, and the total LOS sums to 29, our $\bar{y}$ value is 2.9.

$$(y_i - \bar{y})^2 = (5 - 2.9)^2 = 4.41 \tag{3.23}$$

Similarly, this process continues for all 10 patients:

$$\sum (y_i - \bar{y})^2 = 25.7 \tag{3.24}$$

Therefore, using Equation (3.19), we can calculate the R^2 value to be:

$$R^2 = 1 - \frac{1.3674}{25.7} = 0.9468 \tag{3.25}$$

Therefore, age accounts for 94.68% of the variation in the LOS.

Categorical variables follow a similar process using Equation (3.10)

$$\hat{y} = 1.6667 - 0.6667x_1 + 2.3333x_2 \tag{3.10 revisited}$$

For each patient, either x_1 or x_2 will be given a value of 1, with the other being zero to calculate $\hat{y}$.

$$\hat{y} = 1.6667 - 0.6667(0) + 2.3333(1) = 4 \tag{3.26}$$

$$(y_i - \hat{y}_i)^2 = (5 - 4)^2 \tag{3.27}$$

To compute SS_{RES}, this procedure can be repeated for all 10 cases.

$$\sum (y_i - \hat{y}_i)^2 = 4 \tag{3.28}$$

Since the value of the SS_{TOT} remains unchanged from the previous calculation, R^2 is determined as follows:

$$R^2 = 1 - \frac{4}{25.7} = 0.8444 \tag{3.29}$$

3.2.3.2 Adjusted R^2 Value

The adjusted R^2 value is a modification of R^2 value that accounts for variables that are not significant in the model. The adjusted R^2 determines the extent of the variance of the dependent variable which is explained by the independent variable.

$$\mathrm{R}^2_{adj} = 1 - \left[\frac{(1 - \mathrm{R}^2)(n-1)}{n-k-1}\right] \tag{3.30}$$

where n is the number of points in the sample and k is the number of independent variables in the model.

3.2.3.2.1 Worked Example The calculated R^2 for the worked example was 0.9468. This value can be entered into the Equation (3.30), to calculate the adjusted R^2 value.

$$\mathrm{R}^2_{adj} = 1 - \left[\frac{(1-0.9468)(10-1)}{10-1-1}\right] = 0.9401 \tag{3.31}$$

Therefore, the adjusted R^2 value is 94.01%.

Similarly, the adjusted R^2 value for the hospital attended can be determined using the following formula:

$$R^2_{adj} = 1 - \left[\frac{(1-0.8444)(10-1)}{10-1-1}\right] = 0.8249 \tag{3.32}$$

To evaluate how well the model fits the data in logistic regression, error measures are also required. The following are the four primary evaluation metrics in logistic regression that are used to determine accuracy and error rates and are as follows:

1. Confusion Matrix
2. Classification Report
3. Accuracy Score
4. Receiver Operating Characteristic Curve (ROC curve)

The data utilised in this thesis will be subjected to these four measurement methodologies.

3.2.3.3 Confusion Matrix

The number of values that are successfully or incorrectly predicted is determined from the confusion matrix. True positives (TP), occur when the model correctly predicts the outcome as positive. True negatives (TN), are when the model correctly predicts the outcome as negative. There are two types of errors where the model incorrectly predicts the outcome, type I and type II errors (Table 3.5).

		Predicted Value	
		Positive	**Negative**
Actual Value	**Positive**	True Positive (TP)	False Negative (FN)
	Negative	False Positive (FP)	True Negative (TN)

Table 3.5: General form of a confusion matrix consisting of actual values against predicted values.

False positives (FP), also known as type I errors, occur when a result that should be negative turns out to be positive. As patients who are not diagnosed may be discharged from the care pathway, these errors are frequently considered worse in the application of healthcare. As a result, individuals can be denied access to treatment. False negatives (FN), or type II errors, occur when a result is projected to be positive however, really turns out to be negative. Patients may experience anxiety due to misdiagnosis, but this would be detected and corrected further along the care pathway.

3.2.3.3.1 Worked Example With 15 patients, we may analyse the aggregated LOS against age using logistic regression. The findings are displayed in Table 3.6 and demonstrate that the model correctly predicts the majority of situations since there is only one type I error and two type II errors.

An identical confusion matrix is produced when the logistic regression is performed for the categorical variable hospital, (Table 3.6), allowing the same conclusions to be drawn.

		Predicted Value	
		Positive	**Negative**
Actual Value	**Positive**	6	2
	Negative	1	6

Table 3.6: Confusion matrix for the logistic regression model used in the second worked example with the data set of 15 patients.

3.2.3.4 Classification Report

The second set of metrics focuses on three parameters, which are derived using Table 3.5. These are the precision, recall and F1 scores [246].

$$\text{Precision} = \frac{TP}{TP + FP} \tag{3.33}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{3.34}$$

$$\text{F1 Score} = 2 * \frac{(\frac{TP}{TP+FP}) * (\frac{TP}{TP+FN})}{(\frac{TP}{TP+FP}) + (\frac{TP}{TP+FN})} \tag{3.35}$$

Or equivalently:

$$\text{F1 Score} = 2 * \frac{Precision * Recall}{Precision + Recall} \tag{3.36}$$

The precision is defined by the number of positive results correctly predicted by the total number within the predicted positive class. The recall score is calculated by dividing the total number of correctly predicted positive results by the number of real positive results. The F1 Score is defined as the harmonic mean between the precision and recall values.

3.2.3.4.1 Worked Example The following precision, recall and F1 score will be the same because the confusion matrix produced by the linear and logistic regressions (Table 3.6) was the same.

We may infer the following values from the confusion matrix: TP = 6, FP = 2, FN = 1 and TN = 6. Consequently, the following are the precision, recall and F1 scores:

$$\text{Precision} = \frac{6}{6+2} = 0.75 \tag{3.37}$$

$$\text{Recall} = \frac{6}{6+1} = 0.8571 \tag{3.38}$$

$$\text{F1 Score} = 2 * \frac{\frac{6}{8} * \frac{6}{7}}{\frac{6}{8} + \frac{6}{7}} = 0.8 \tag{3.39}$$

All three results are greater than 70%, demonstrating the model's strong ability to forecast the LOS categories.

3.2.3.5 Accuracy Score

The proportion of correctly classified predictions over all of the predictions is known as accuracy. Equation (3.40) can be used to express this [246].

$$\text{Accuracy} = \frac{TN + TP}{TN + FP + TP + FN} \tag{3.40}$$

A balanced distribution of the data is important when considering accuracy as a scoring criterion. When there is an imbalance in the data set, the prediction frequently skews the findings significantly so that they fall into one of the predicted categories. A high accuracy result would be the outcome of this.

3.2.3.5.1 Worked Example The accuracy can be determined once more by using the confusion matrix from the previous calculation.

$$\text{Accuracy} = \frac{6+6}{6+2+1+6} = 0.8 \tag{3.41}$$

An accuracy score of 80% shows that patients are appropriately assigned to the LOS group 80% of the time.

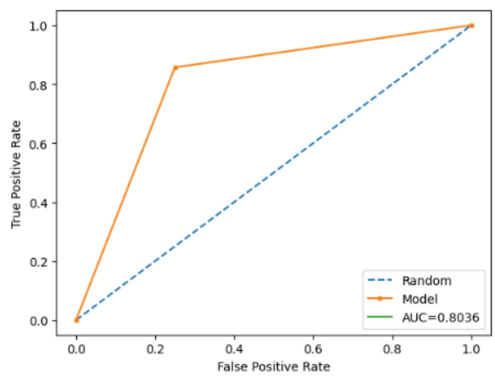


Figure 3.2: ROC curve for continuous age and LOS prediction.

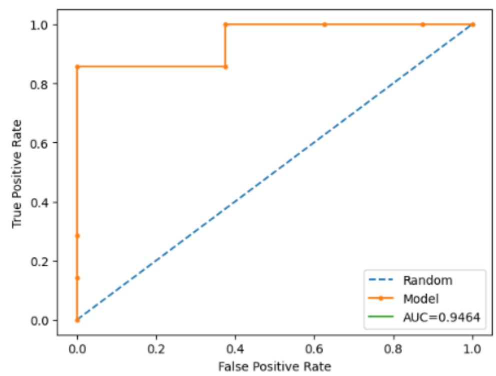


Figure 3.3: ROC curve for categorical hospital and LOS prediction.

3.2.3.6 ROC Curve

The ROC curve, which displays the TP rate against the FP rate at various classification thresholds, is the final metric. The area under the curve (AUC) represents the probability that a randomly chosen positive example will be ranked higher by the model than a randomly chosen negative example. The model receives a score between 0 and 1, with the higher the AUC value, the greater the percentage of properly predicted values.

3.2.3.6.1 Worked Example In Figures 3.2 and 3.3, the ROC curves are displayed. The AUC for the age example was 80.36%, which indicates that a large percentage of projected values are successfully predicted. The higher AUC value of 94.64% implies that the hospital variable is more effective at predicting the LOS group when directly compared to the age example.

This subsection has described the scoring criteria for both linear and logistic regressions and shown how these models function with worked examples of 10 and 15 patients, respectively. Chapter 5 of this thesis will further apply these predictive analytics to a data set of frail and elderly patients.

3.3 Classification and Regression Trees

Classification and regression trees (CART) are a data mining technique in which variables predict an outcome. The parameters that make up the final groups are visually represented by a decision tree. The decision tree asks a series of questions that decide the groups into which the data is sorted. CART models are a form of binary recursive partitioning, where each node is split into two groups.

In data mining, a decision tree is a prediction model that may be applied to both classifiers and regression models in data mining. In contrast, the term 'decision tree' in operations research refers to the hierarchical model of decisions and their consequences.

CART is comprised of four main components. The dependent variable that the algorithm seeks to predict is the first element. The second component is the independent variables that are related to and used to predict the dependent variable. The third component is the training data set, a subset of the main data set that includes both the dependent and independent variables. This is used to train and allow the algorithm to learn. The testing data set, which is the final component, will assess the precision and reliability of the algorithm's predictions.

The decision tree illustrates the clinical judgements required to reach the final classification grouping. It is composed of root, decision and terminal nodes. The tree's root node, which symbolises the whole population, is at the top of the structure. The population is then divided up into decision or terminal nodes based on certain criteria. The tree's terminal nodes, which hold the information indicating to which grouping the data belongs, are located at its very end.

Depending on the maximum tree depth or the maximum number of leaves (terminal nodes) in the model, the tree's size will vary. Figure 3.4 displays a decision tree with a max depth of one or two maximum leaves.

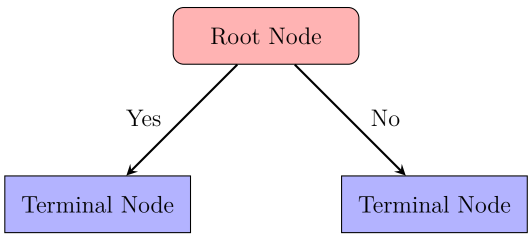


Figure 3.4: Example of a decision tree with one root node and two terminal nodes.

The depth or maximum number of leaves can be increased to allow the potential for more splits (see Figure 3.5). From this, the following rules can be deduced:

1. If (Root-Node = True) AND (Decision-Node-1 = True) THEN Terminal-Node-1 = True
2. If (Root-Node = True) AND (Decision-Node-1 = False) THEN Terminal-Node-2 = True
3. If (Root-Node = False) AND (Decision-Node-2 = True) THEN Terminal-Node-3 = True
4. If (Root-Node = False) AND (Decision-Node-2 = False) AND (Decision-Node-3 = True) Then Terminal-Node-4 = True

5. If (Root-Node = False) AND (Decision-Node-2 = False) AND (Decision-Node-3 = False) Then Terminal-Node-5 = True

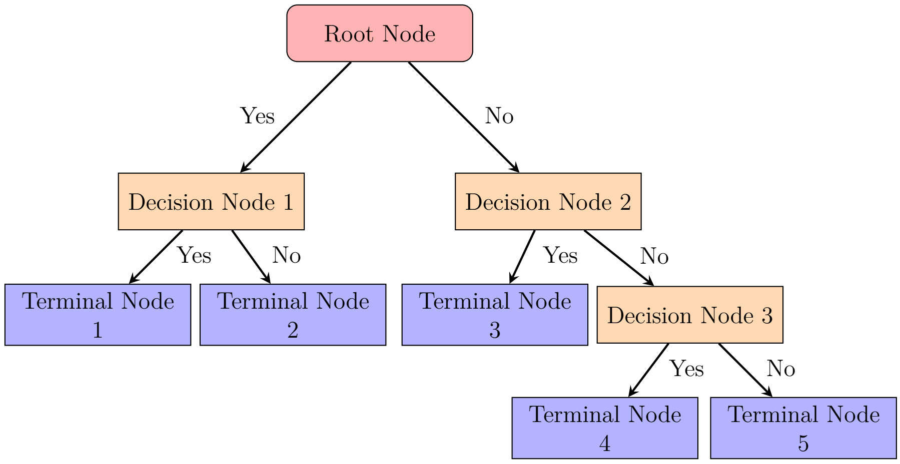


Figure 3.5: Example of a decision tree with one root node, three decision nodes and five terminal nodes.

The algorithm determines the most important splitting criteria in order to gain the most information.

3.3.1 Generalised Formulation

Within both regression and classification trees, categorical variables are present within the data. Because of the nature of the algorithm, the variables have to undergo preprocessing to change these to numerical data. Since there is no ordinal relationship in the categorical data, one-hot encoding must be used instead of integer encoding. Each distinct integer value is represented by a brand-new binary variable, which replaces the categorical variable.

The newly one-hot encoded variables can then be run with the numerical variables into the CART algorithm.

3.3.1.1 Regression Trees

A continuous outcome variable is predicted via regression trees. The regression algorithm known as *'DecisionTreeRegressor'* is part of the machine learning toolkit called 'Scikit-Learn' [247] in Python. Regression trees can be created and developed as a result.

The decision-making process of the algorithm is based on the mean square error (MSE), which also helps to establish the final groupings of data. The MSE informs the user as to how much their prediction deviates from the original target (Equation (3.42)) [248]. Since regression trees are aiming to predict a continuous variable, once the final groupings are determined, the average of the dependent variable is calculated.

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2 \tag{3.42}$$

where Y is the actual value and $\hat{Y}$ is the prediction. The R^2 value determines the coefficient of determination of the prediction, given in Equation (3.45).

The process of creating a regression tree is described by Algorithm 1.

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^n_a
    $x_a = MSE^n$ - MSE^n_a
  end
  Set $MSE^n = Max(\ x_a)$
  Create two new nodes, n' and n" and calculate new MSE^n *for each*
end

A decision tree of the regression algorithm can then be constructed to provide the user with a visual representation of the clinical decisions.

3.3.1.1.1 Worked Example Reverting to the prior example, we apply Algorithm 1 to forecast the continuous LOS for the 15 hospitalised patients.

Figure 3.6 shows a regression tree with a test set size of 20% and a maximum of four leaf nodes. For the leaf nodes where the MSE is equal to zero, $Y_i - \hat{Y}_i$ is also equal to zero and therefore there is a perfect prediction. The figure also shows that the most crucial element in deciding a patient's LOS is whether they are 86 years old or younger. The LOS is represented by colours, and the shorter the LOS, the lighter the colour. The node with the value 4.5, is the one with the darkest colour, indicating that this has the longest LOS prediction in the model.

3.3.1.2 Classification Trees

When the dependent variable is categorical and the tree algorithm seeks to predict the class, classification trees are utilised. The classification technique, *'DecisionTreeClassifier'*, can be used with the Scikit-Learn [247] Python library. Although

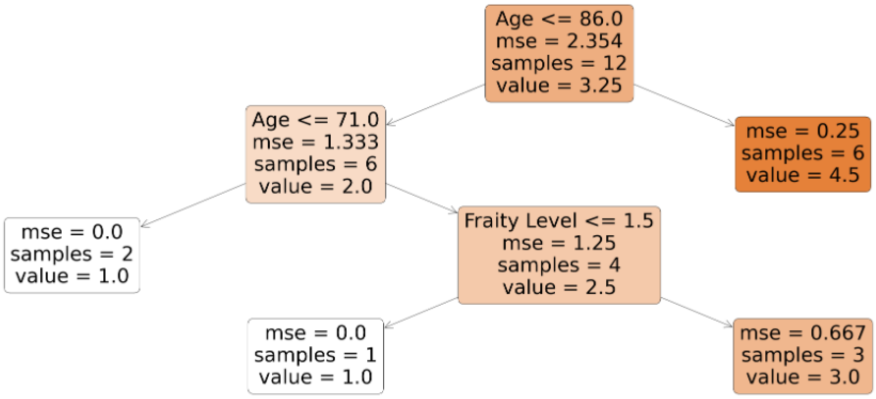


Figure 3.6: Regression tree of the first worked example consisting of four terminal nodes. Note that a darker colour of the node indicates a longer LOS.

classification trees adopt a methodology very similar to that of regression trees, MSE should not be applied because they forecast a categorical outcome variable. Instead, the optimum splitting choice is determined using the Gini Index or Gini impurity. The Gini Index yields a number between 0 and 1, with a smaller value indicating greater sample homogeneity. The Gini Index is calculated using Equation (3.43), which involved subtracting the sum of the squared probabilities of each class from one [249].

$$\text{Gini Index} = 1 - \sum_{i=1}^{n} p_i^2 \tag{3.43}$$

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

Algorithm 2 illustrates the steps involved in creating a classification tree and has been extended from Algorithm 1.

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^n_a
        x$_a$ = Gini-Index^n - Gini-Index^n_a
    end
    Set Gini-Index^n = *Max* (x_a)
    Create two new nodes, n' and n" and calculate new Gini-Index for each
end

3.3.1.2.1 Worked Example For the classification tree, patients were split into two groups based on their LOS: those admitted for four days or less and those admitted for four days or more. Once more, a test set of 20% was used.

The classification tree is depicted in Figure 3.7, where the maximum number of leaf nodes is set to four. Using the training data, perfect prediction occurs since all four end nodes have a Gini Index of zero. The regression tree's findings (Figure 3.6) showed that a patient's age was the primary determinant in identifying the LOS group. The rest of the tree breaks differently, though. The number of patients that fit into a certain node is indicated by the colour depth in the diagram. White nodes show an equal distribution of patients inside that node.

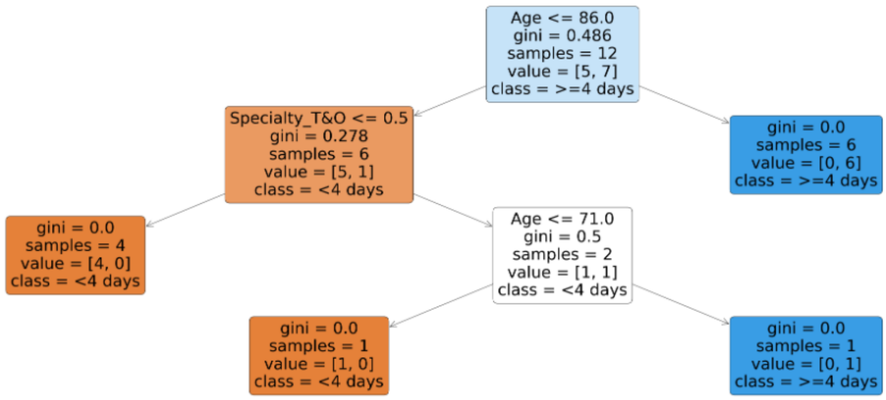


Figure 3.7: Classification tree for second worked example consisting of four terminal nodes. Note that blue indicates the class of ≥ 4 days, and orange indicates the class of < 4 days.

3.3.2 Extension of Random Forests

Random forests can be derived from CART models. Random forests are a bagging method which constructs decision trees on several samples, using the majority vote for classification and the mean results for regression models. The more trees there are in the forest, the more accurate the model is likely to be and the likelihood of overfitting is reduced. The algorithm's operation is shown in Figure 3.8.

The individual trees are constructed using bootstrap samples as opposed to the original sample (Algorithm 3). When it comes to classification and regression trees, the splits of the tree depend on the MSE or Gini Index, respectively. Similarly, random forests employ the Scikit-Learn library [247] and uses the classes 'RandomForestRegressor' and 'RandomForestClassifier' for regression and classification trees, respectively.

3.3.2.0.1 Worked Example The working example may be used with both the random forest's regression and classification versions. Due to their nature, random

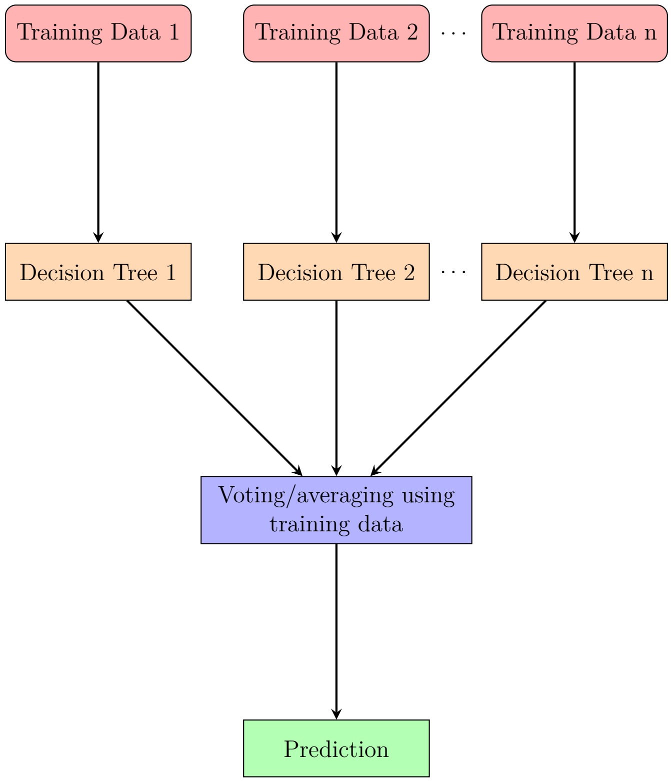


Figure 3.8: Generic example of a random forest consisting of n decision trees and n training data sets.

Algorithm 3: Random Forests - Adapted from [250]

```
Determine stopping criteria:
B = Number of subtrees
while MSE or Gini-Index^n >0 or stopping criterion not met do
    for i = 1 to B do
        Draw a bootstrap sample of size N from the training data
        while node size != minimum node size do
            randomly select a subset of m predictor variables from total p
            for j = 1 to m do
                if jth predictor optimises splitting criterion then
                    split internal node into two child nodes
                    break;
                end
            end
        end
    end
end
return the ensemble tree of all B subtrees generated in the outer for loop;
```

forests often perform better when there is more data available, as it allows them to sample more decision trees. However, random forests can become difficult to visualise. The number of decision trees will be displayed in accordance with the number of decision tree iterations selected, although with large numbers of iterations, this can become too computationally expensive.

3.3.3 Feature Parameters

There are parameters and attributes within each of the four cases; *'DecisionTreeRegressor'*, *'DecisionTreeClassifier'*, *'RandomForestRegressor'* and *'RandomForestClassifier'*, that can be changed to increase the accuracy of the predictions. Tables 3.7 and 3.8 display the parameter and attribute options for each of the methods, with the default entry for each shown in blue. The best parameter or value to select in order to obtain the highest accuracy score can be determined by subjecting these parameters to parameter optimisation.

All four methods, share 11 of the same parameters used within the model. The criterion is a function that assesses the split's quality and ultimately determines where the split occurs. The splitter is the process used to decide whether the best split will be employed at each node, or if a random split will be selected. The maximum depth of the tree is determined by its max_depth; if no parameter is chosen, the node expands until all of its leaves are pure or until all of its leaves include fewer than the min_samples_split. The min_samples_split is the minimum number of samples required to split an internal node. Similarly, the min_samples_leaf is the minimum number of samples required to be at a leaf node. Only split points that will leave two nodes with at least min_samples_leaf training samples will be taken into consideration. The min_weight_fraction_leaf is the fraction of the input samples required to be at a leaf node. The max_features determines the number of features to consider when looking for the best split. If *auto, sqrt, log2* or *None* are selected, then max_features will be equal to the attribute *n_features*. The parameter random_state controls the randomness of the estimator and the features are always randomly permuted at each split, even if 'splitter' is set to 'best'. To obtain deterministic behaviour during fitting, random_state has to be fixed to an integer. Max_leaf_nodes set the maximum number of end nodes that will be constructed in the model, if no value is selected then there will be an unlimited number of leaf nodes. Min_impurity_decrease causes a node to be split if the split induces a decrease of the impurity greater than or equal to its value. The following equation calculates the min_impurity_decrease:

$$\frac{N_t}{N} * (impurity - \frac{N_{t_R}}{N_t} * right\_impurity - \frac{N_{t_L}}{N_t} * left\_impurity) \tag{3.44}$$

where N is the total number of samples, N_t is the number of samples at the current node, N_{t_L} is the number of samples in the left child node and N_{t_R} are the number of samples in the right child node. In essence, the formula accounts for the parent nodes contribution to the whole tree $\frac{N_t}{N}$. The ccp_alpha complexity parameter, which selects the subtree with the largest cost complexity that is smaller than the ccp_alpha value, is the final parameter that is shared by all four models.

Additionally, the '$DecisionTreeClassifier$' and '$RandomForestClassifier$ require the parameter class_weight. This parameter can be used to account for unbalanced classes and gives a class with a high population greater weight.

The random forest algorithms require further parameters. Firstly, n_estimators tells the user the number of trees within the model. Bootstrap determines whether samples of the data are used within the model or if the whole data set is used. Following on, if bootstrap = True, then there is the option for oob_score which uses out-of-bag samples to evaluate its performance. The n_jobs parameter controls how many jobs will run concurrently, and how many processors will be available. Warm_start allows parts of the model that were learned from previous parameter values to be reused, which ultimately saves time. Finally, max_samples requires the user to select the number of samples to draw from to train each base estimator, only if bootstrap = True.

Parameter	**DecisionTreeRegressor**	**DecisionTreeClassifier**
criterion	"squared_error", "friedman_mse", "absolute_error", "poisson"	"gini","entropy","log_less"
splitter	"best", "random"	"best","random"
max_depth	integer, (None)	integer, (None)
min_samples_split	integer, float, (2)	integer, float, (2)
min_samples_leaf	integer, float, (1)	integer, float, (1)
min_weight_fraction_leaf	Float (0)	Float (0)
max_features	integer, float, "auto", "sqrt", "log2", (None)	integer, float, "auto", "sqrt", "log2", (None)
random_state	integer, "random_state", (None)	integer, "random_state", (None)
max_leaf_nodes	integer, (None)	integer, (None)
min_impurity_decrease	float, (0)	float, (0)
class_weight	(N/A)	dict, list of dicts, "balanced", (None)
ccp_alpha	non-negative float, (0)	non-negative float, (0)
Attributes	**DecisionTreeRegressor**	**DecisionTreeClassifier**
classes	(N/A)	ndarray of shape or list of ndarray
feature_importances_	ndarray of shape	ndarray of shape
max_features	integer	integer
n_classes_	(N/A)	integer, list of integers
n_features_	integer	integer
n_features_in	integer	integer
feature_names_in_	ndarray of shape	ndarray of shape
n_outputs_	integer	integer
tree_	Tree instance	Tree instance

Table 3.7: 'DecisionTreeRegressor' and 'DecisionTreeClassifier' associated parameters and attributes the user can select. The default variable is highlighted in blue. For variables which are an integer, float or there is no default, this is listed in brackets.

Parameter	**RandomForestRegressor**	**RandomForestClassifier**
n_estimators	integer, (100)	integer, (100)
criterion	"squared_error", "absolute_error", "poisson"	"gini","entropy","log_less"
max_depth	integer, (None)	integer, (None)
min_samples_split	integer, float, (2)	integer, float, (2)
min_samples_leaf	integer, float, (1)	integer, float, (1)
min_weight_fraction_leaf	Float (0)	Float (0)
max_features	integer, float, "sqrt", "log2", (None)	integer, float, "sqrt", "log2", (None)
random_state	integer, "RandomState instance", (None)	integer, "RandomState instance", (None)
max_leaf_nodes	integer, (None)	integer, (None)
min_impurity_decrease	float, (0)	float, (0)
bootstrap	bool, (True)	bool, (True)
oob_score	bool, (False)	bool, (False)
n_jobs	integer, (None)	integer, (None)
warm_start	bool, (False)	bool, (False)
ccp_alpha	non-negative float, (0)	non-negative float, (0)
max_samples	integer, float, (None)	integer, float, (None)
class_weight	(N/A)	dict, list of dicts, "balanced","balanced_subsample", (None)
Attributes	**RandomForestRegressor**	**RandomForestClassifier**
base_estimator_	DecisionTreeRegressor	DecisionTreeClassifier
estimators_	list of DecisionTreeRegressor	DecisionTreeClassifier
classes	(N/A)	ndarray of shape, list of such arrays
feature_importances_	ndarray of shape	ndarray of shape
n_features_	integer	integer
n_features_in	integer	integer
feature_names_in_	ndarray of shape	ndarray of shape
n_outputs_	integer	integer
oob_score	float	float
oob_prediction_	ndarray of shape	ndarray of shape

Table 3.8: 'RandomForestRegressor' and 'RandomForestClassifier' associated parameters and attributes the user can select. The default variable is highlighted in blue. For variables which are an integer, float or there is no default, this is listed in brackets.

3.3.4 Evaluation Metrics

To ascertain the success and how well the model predicts LOS, a series of evaluation measures can be applied to each of the different CART and random forest models.

3.3.4.1 Regression Models

The same coefficient of determination (R^2) metric can be used to compare the regression tree and random forest regression models.

Recall the linear regression equation, Equation (3.19). This equation can be modified to create an equation for calculating success when using regression trees (Equation (3.45)) [251].

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)}{\sum_{i=1}^{n}(Y_i - \bar{Y}_i)} \tag{3.45}$$

where Y_i represents the true y value, $\hat{Y}_i$ is the value of the predicted y value, $\bar{Y}_i$ represents the mean of all values and n is the total number of observations. The higher the R^2 value, the more reliable the model is.

To calculate the R^2 for the random forest, the average R^2 is calculated (Equation (3.46))

$$R^2_{average} = \frac{1}{k}\sum_{i=1}^{k} R^2_k \tag{3.46}$$

where k is the number of decision trees selected to combine to the random forest.

3.3.4.1.1 Worked Example A test set of 20%, or three patients, was included in the regression example so that the accuracy could be assessed. Three patients were chosen at random, and their LOS's were three days, one day and one day. The predicted values for each of these variables were three days, one day and one day, respectively. Therefore, the u term is equal zero when computing R^2 from Equation (3.45) and as a result, R^2 is equal to one.

The random forest regression model generated a negative R^2 score. This would imply that the model is not a very good predictor for LOS. The negative score is caused by $\sum_{i=1}^{n}(Y_i - \hat{Y}_i) > \sum_{i=1}^{n}(Y_i - \bar{Y}_i)$.

3.3.4.2 Classification Models

Some of the same metrics used for logistic regression analysis can also be used to analyse the classification tree and random forest (classification) models.

The TP, FP, TN and FN rates will be obtained from Table 3.5, which will then be used to calculate the precision, recall and accuracy as follows [246]:

$$\text{Precision} = \frac{TP}{TP + FP} \tag{3.33 revisited}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{3.34 revisited}$$

$$\text{Accuracy} = \frac{TN + TP}{TN + FP + TP + FN} \tag{3.40 revisited}$$

The accuracy scoring method will be the primary one employed because it takes into account both the TP and TN values to determine how accurate the prediction is. When all decision trees are performed on the testing set, the answer that appears the most frequently is chosen as the final result.

Using CART models for prediction has several benefits because they can automatically identify key variables and the order in which they should be prioritised. The algorithm also generates a visual representation of the decision tree which simplifies and clarifies the final result. The tree structure, however, might become unstable and introduce variance if the data set experiences even a slight shift. If some classes are unbalanced, it is also possible to produce under or over fitted trees.

Due to the averaging of the outcome, using random forests might increase accuracy by eliminating under or over fitting. However, the larger size challenge necessitates more computational time and resources.

3.3.4.2.1 Worked Example A test set consisting of 20% of the original data, or three patients, was created for the classification tree. Due to the seed within the code being set to 0 to guarantee findings are reproducible, the three selected patients were the same as those in the regression tree example. The patients LOS was three, one and one days. All three patients were predicted to be in the ‘<4 days’ category. As a result, TN, FP and FN rates are all equal to zero, while the TP rate is equal to three. Consequently, the values for precision, recall and accuracy are all equal to one.

A maximum of four leaf nodes were used in the random forest classification algorithm’s analysis of the data. It is interesting to note that the TP rate and FN rates were both calculated to be two, while the TN and FP rates were both zero. The model’s recall therefore stands at 0.67. This would imply that utilising the straightforward classification tree yields superior outcomes in this instance.

3.4 Python Development

The development of the linear regression, logistic regression, and random forest models was executed within Python (Version 3.8.9). It is advantageous to utilise Python, an open source programming software, for data analysis and data visualisation when there are large amounts of data.

Using the Pandas library [252] (Version 1.2.0), the hospital admission data may be imported into Python as a dataframe. Additionally the library, statsmodels.api [253] (Version 0.12.2), was utilised to create statistical models. There is a list of additional libraries that are needed in the corresponding code. Data can then be subjected to analysis to identify patterns within the data.

3.4.1 Linear Regression

Depending on whether the independent variables are continuous or categorical, there are two coding strategies that can be used for linear regression models.

3.4.1.1 Linear Regression - Continuous Independent Variable

The following gives an illustration of how the linear regression model can be built with a continuous independent variable:

```
import pandas as pd
import statsmodels.api as sm
x = df['Age'] #Independent variable
y = df['LOS'] #Dependent variable
x=sm.add_constant(x) #Ensures there is a constant term
model = sm.OLS(y, x).fit() #Using Ordinary Least Squares method
print(model.summary())
```

Figure 3.9 presents the relevant outputted results. According to Equation (3.1), the 'coef' term specifies how much an increase in one unit will increase the total dependent variable. To confirm that the results are statistically significant, it is also crucial to look at the p value, which is shown in Figure 3.9 as 'P> |t|'. The R^2 and the adjusted R^2 values can also be extracted directly from this table.

3.4.1.2 Linear Regression - Categorical Independent Variable

Similar coding methods can be applied for those variables which are categorical, with the addition of one-hot encoding.

```
import pandas as pd
import statsmodels.api as sm
x = df['Hospital'] #Independent variable
x = pd.get_dummies(data=x) #Applying one-hot encoding
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                    LOS   R-squared:                       0.947
Model:                            OLS   Adj. R-squared:                  0.940
Method:                 Least Squares   F-statistic:                     141.8
Date:                Wed, 26 Oct 2022   Prob (F-statistic):           2.27e-06
Time:                        17:06:47   Log-Likelihood:                -4.2406
No. Observations:                  10   AIC:                             12.48
Df Residuals:                       8   BIC:                             13.09
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const        -10.1052      1.092     -9.257      0.000     -12.623      -7.588
Age            0.1578      0.013     11.907      0.000       0.127       0.188
==============================================================================
```

Figure 3.9: Excerpt of the OLS results for continuous linear regression showing the R^2 value, the coefficient values, and their corresponding p-values.

```python
y = df['LOS'] #Dependent variable
model = sm.OLS(y, x).fit() #Using Ordinary Least Squares method
print(model.summary())
```

Figure 3.10 contains the outputted findings. The fundamental distinction between categorical and continuous coefficients is that, if one exists, the constant term is increased by a value of either "-0.6667" or "2.3333" depending on which of the categorical coefficients is present.

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                    LOS   R-squared:                       0.844
Model:                            OLS   Adj. R-squared:                  0.824
Method:                 Least Squares   F-statistic:                     43.20
Date:                Thu, 27 Oct 2022   Prob (F-statistic):           0.000174
Time:                        14:14:16   Log-Likelihood:                -9.6079
No. Observations:                  10   AIC:                             23.22
Df Residuals:                       8   BIC:                             23.82
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          1.6667      0.152     10.954      0.000       1.316       2.018
NHH           -0.6667      0.255     -2.619      0.031      -1.254      -0.080
RGH            2.3333      0.226     10.340      0.000       1.813       2.854
==============================================================================
```

Figure 3.10: Excerpt of the OLS results for categorical linear regression showing the R^2 value, the coefficient values, and their corresponding p-values.

3.4.2 Logistic Regression

Depending on whether the independent variables are continuous or categorical, two coding strategies for the logistic regression models can be used, adapting from the prior linear regression models.

3.4.2.1 Logistic Regression - Continuous Independent Variable

The following gives an illustration of how the logistic regression model might be used with a continuous independent variable:

```python
import statsmodels.formula.api as smf
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
from statsmodels.formula.api import logit
from sklearn.metrics import roc_auc_score, roc_curve

conditions = [(df['LOS'] <4),(df['LOS'] >=4)] #Creating grouped LOS
values = [0,1] #Assigning values to the new groups
df['LOS_group'] = np.select(conditions,values) #Creating LOS_group column
model = smf.logit("LOS_group ~ Age", data = df)
results = model.fit_regularized()
results.summary()

results.pred_table() #Prints confusion matrix

def plot_roc_curve(Y_test, model_probs): #Create ROC curve function
    random_probs = [0 for _ in range(len(Y_test))]
    model_auc = roc_auc_score(Y_test, model_probs) #Calculate AUC
    random_fpr, random_tpr, _ = roc_curve(Y_test, random_probs) #Calculate ROC Curve for Random Model
    model_fpr, model_tpr, _ = roc_curve(Y_test, model_probs) #Plot ROC curves
    plt.plot(random_fpr, random_tpr, linestyle='--', label='Random')
    plt.plot(model_fpr, model_tpr, marker='.', label='Model')
    plt.plot(model_auc, label= 'AUC=%.4f' % model_auc)
    plt.xlabel('False Positive Rate') #Plot x axis label
    plt.ylabel('True Positive Rate') # Plot y axis label
    plt.legend()
    plt.show()
y_pred = results.fittedvalues #Determine predicted values
plot_roc_curve(df['LOS_group'], y_pred) #Plot the ROC curve
```

The associated outputted results are given in Figure 3.11. The 'coef' term, which is also used in linear regression, estimates how much an increase in one unit will result in an overall rise in the dependent variable (from Equation (3.17)). As previously stated, it is crucial to verify the p value, represented by the notation 'P> |t|', to make sure the results are statistically significant. The confusion matrix and ROC curve generation instructions are also included in the code above.

```
Optimization terminated successfully    (Exit mode 0)
            Current function value: 0.3390099350344781
            Iterations: 15
            Function evaluations: 18
            Gradient evaluations: 15
```

Logit Regression Results

Dep. Variable:	LOS_group	**No. Observations:**	15
Model:	Logit	**Df Residuals:**	13
Method:	MLE	**Df Model:**	1
Date:	Thu, 27 Oct 2022	**Pseudo R-squ.:**	0.5093
Time:	14:51:52	**Log-Likelihood:**	-5.0851
converged:	True	**LL-Null:**	-10.364
Covariance Type:	nonrobust	**LLR p-value:**	0.001157

	coef	std err	z	P>|z|	[0.025	0.975]
Intercept	-21.7946	9.962	-2.188	0.029	-41.320	-2.270
Age	0.2661	0.121	2.206	0.027	0.030	0.502

Figure 3.11: Excerpt of the Logit regression results for continuous logistic regression showing the pseudo R^2 value, the coefficient values, and their corresponding p-values.

3.4.2.2 Logistic Regression - Categorical Independent Variable

The logistic regression Python code for categorical variables is identical to the coding for the continuous variables, since the ‘logit’ function can determine the type of variable inputted and act accordingly.

Furthermore, the findings of the output are very similar to those of the continuous logistic regression. The outcomes for the worked example are shown in Figure 3.12. The number of variables contained within the independent variable, minus one, is displayed in the results. Since there was only one alternative hospital choice in this particular example, the value would simply be the intercept (i.e., -1.79 or category ‘0’) if the hospital attended was not RGH.

3.4.3 Regression Trees

For regression trees, both the R^2 score and visualisation can be demonstrated within Python. From the original data set entered into the notebook, the following code generates testing and training data sets. Different parameters can be entered into the ’DecisionTreeRegressor’ function within this model, as detailed in Table 3.7.

```python
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from sklearn.tree import DecisionTreeRegressor
from sklearn import tree
```

```
Optimization terminated successfully    (Exit mode 0)
            Current function value: 0.4912996923313728
            Iterations: 13
            Function evaluations: 13
            Gradient evaluations: 13
```

Logit Regression Results

Dep. Variable:	LOS_group	**No. Observations:**	15
Model:	Logit	**Df Residuals:**	13
Method:	MLE	**Df Model:**	1
Date:	Thu, 27 Oct 2022	**Pseudo R-squ.:**	0.2889
Time:	15:30:43	**Log-Likelihood:**	-7.3695
converged:	True	**LL-Null:**	-10.364
Covariance Type:	nonrobust	**LLR p-value:**	0.01440

	coef	std err	z	P>|z|	[0.025	0.975]
Intercept	-1.7918	1.080	-1.659	0.097	-3.909	0.325
Hospital[T.RGH]	2.8904	1.354	2.135	0.033	0.237	5.544

Figure 3.12: Excerpt of the Logit regression results for categorical logistic regression showing the pseudo R^2 value, the coefficient values, and their corresponding p-values.

```python
from sklearn.model_selection import train_test_split
from sklearn.tree import plot_tree
from sklearn.metrics import confusion_matrix

x = df.drop('LOS',axis = 1).copy() #Removing dependent variable from analysis
x = pd.get_dummies(x, columns = ['Hospital',
                                 'Specialty',
                                 'Admission Method',
                                 'Admission Source'
                                ]) #Applying one-hot encoding
df['LOS'] =df['LOS'].astype(float)
y = df['LOS'].copy() #Ensuring dependent variable is in own dataframe
X_train, X_test, Y_train, Y_test = train_test_split(x,y, test_size=0.2, random_state=0) #Creating training and testing data
clf_dt = DecisionTreeRegressor(random_state=0,max_leaf_nodes=4)
clf_dt.fit(X_train, Y_train) #Determines fit of the algorithm
Y_predict=clf_dt.predict(X_test)
print(clf_dt.score(X_test,Y_test)) #Prints R-squared score

plt.figure(figsize=(60,20))
plot_tree(clf_dt,
         filled=True,
         rounded=True,
```

```
        feature_names = x.columns); #Code to plot regression tree
    visual
plt.show()
```

3.4.4 Classification Trees

To choose the optimum splitting node, classification trees generate the Gini Index at each level using the Python class 'DecisionTreeClassifier'. The following code produces testing and training data sets similar to regression trees and uses parameters from Table 3.8 to establish the stopping criterion.

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from sklearn import tree
from sklearn.tree import DecisionTreeClassifier
from sklearn import metrics
from sklearn.model_selection import train_test_split
from sklearn.metrics import classification_report
from sklearn.tree import plot_tree
from sklearn.metrics import confusion_matrix

conditions = [(df['LOS'] <4),(df['LOS'] >=4)] #Creating categories
    to predict
values = [0,1]
df['LOS_group'] = np.select(conditions,values)
x = df.drop('LOS_group', axis =1).copy() #Removing dependent
    variable from analysis
x =pd.get_dummies(x,columns=['Hospital',
                             'Specialty',
                             'Admission Source',
                             'Admission Method'
                             ]) #Applying one-hot encoding
df['LOS_group']=df['LOS_group'].astype(float)
y = df['LOS_group'].copy() #Ensuring dependent variable is in own
    dataframe

X_train, X_test, Y_train, Y_test = train_test_split(x,y, test_size
    =0.2, random_state=0) #Creating training and testing data

clf_dt = DecisionTreeClassifier(random_state=0,max_leaf_nodes=4) #
    Applying the algorithm
clf_dt.fit(X_train, Y_train) #Determines fit of algorithm
Y_predict=clf_dt.predict(X_test)
print("Accuracy:", metrics.accuracy_score(Y_test, Y_predict)) #
    Prints accuracy score
```

```
print(classification_report(Y_test, Y_predict)) #Prints Precision, Recall and F1 Scores
print(confusion_matrix(Y_test, Y_predict)) #Prints confusion matrix

plt.figure(figsize=(60,20))
plot_tree(clf_dt,
          filled=True,
          rounded=True,
          class_names = ['<4 days', '>=4 days'],
          feature_names = x.columns); #Code to plot classification tree visual
plt.show()
```

3.4.5 Random Forests - Regression

The following code is adapted from the regression tree with the class changed to 'RandomForestRegressor'. Due to the difficulty in visualising these trees as was previously mentioned in Section 3.3.2, there is no code to plot the random forest. To assess the model's fit, the R^2 is also outputted.

```
import pandas as pd
import numpy as np
from sklearn.metrics import accuracy_score
from sklearn.ensemble import RandomForestRegressor
from sklearn.model_selection import train_test_split
from sklearn.metrics import confusion_matrix

df['LOS'] = np.select(conditions,values)
x = df.drop('LOS', axis =1).copy() #Removing dependent variable from analysis
x =pd.get_dummies(x,columns=['Hospital',
                             'Specialty',
                             'Admission Source',
                             'Admission Method'                ]) #Applying one-hot encoding

df['LOS']=df['LOS'].astype(float)
y = df['LOS'].copy() #Storing dependent variable in own dataframe

X_train, X_test, y_train, y_test = train_test_split(x, y, random_state=0,test_size=0.2) #Creating training and testing data

forest = RandomForestRegressor(n_estimators= 10, max_leaf_nodes =4) #Applying the algorithm
forest.fit(X_train, y_train) #Determine fit of algorithm

```

```
print(forest.score(X_test, y_test))
```

3.4.6 Random Forests - Classification

Adapting the code from the regression to the classification random forest is shown below. To assess the random forest's fit, the score metrics are also printed.

```
import pandas as pd
import numpy as np
from sklearn.metrics import accuracy_score
from sklearn.ensemble import RandomForestClassifier
from sklearn.model_selection import train_test_split
from sklearn.metrics import confusion_matrix

conditions = [(df['LOS'] <4),(df['LOS'] >=4)] #Creating categories
    for prediction
values = [0,1]
df['LOS_group'] = np.select(conditions,values)
x = df.drop('LOS_group', axis =1).copy() #Removing dependent
    variable from analysis
x =pd.get_dummies(x,columns=['Hospital',
                             'Specialty',
                             'Admission Source',
                             'Admission Method'               ]) #
    Applying one-hot encoding

df['LOS_group']=df['LOS_group'].astype(float)
y = df['LOS_group'].copy() #Storing dependent variable in own
    dataframe

X_train, X_test, y_train, y_test = train_test_split(x, y,
    random_state=0,test_size=0.2) #Creating training and testing
    data

forest = RandomForestClassifier(n_estimators= 10, max_leaf_nodes
    =4) #Applying the algorithm
forest.fit(X_train, y_train) #Determine fit of algorithm

y_pred_test = forest.predict(X_test)
print(forest.score(X_test, y_pred_test)) #Calculates the R squared
    value
```

3.5 Summary

This chapter has provided a comprehensive introduction to the theory underlying the most popular predictive analytical techniques presently used in healthcare. Since this chapter covered general OR approaches, a step-by-step practical example has also been included so that healthcare professionals can quickly apply these strategies to their own departments and data. To enable model adaptation and parameter optimisation, detailed executable Python code has been provided. This allows the code to be applied to any healthcare database. The theory covered in this chapter will be applied to a case study of frail and elderly patient admissions within ABUHB, later on in this thesis.

In the following chapter, Chapter 4, the use of prescriptive analytics within healthcare is discussed, namely deterministic and two-stage stochastic modelling.

Chapter 4

Prescriptive Analytics to Support Capacity Planning for Frail and Elderly Patients

4.1 Introduction

Data analytics is categorised into three main paradigms, each with a correspondingly varying degree of complexity: descriptive analytics, predictive analytics and prescriptive analytics. In both industry and healthcare analytics, the first two stages have been extensively studied and documented. The third paradigm, prescriptive analytics, eliminates the planning risk by bridging the gap between the data that an organisation has and the consequences of implementing new policy. Decision-makers can gain a deeper understanding of how to seize an opportunity or alleviate a problem in the future as a result. As the related work section revealed, there are research gaps in the prescriptive healthcare analytics work which is concerned with mathematical modelling of healthcare services. Literature reviews in this area of OR have only recently been published, making it a relatively new and developing discipline [254, 255, 256]. In all three articles, the application of prescriptive analytics is discussed, with Lopes et al. [255] arguing that successful optimisation of current healthcare resources will, in turn, decrease existing waiting lines and allow a greater ability to treat individuals in need effectively. Islam et al. [254] discovered that just 9% of the publications analysed within their review focused on prescriptive analytics, demonstrating that this is a new and emerging area of healthcare research.

<u>**Research Aim**</u> - This chapter will present the theory in order to address two of the research's objectives: 'How best can specialties be organised among a network

of hospitals to ensure staffing and bed costs are minimised whilst, whilst still meeting the demand for frail and elderly patients?' and 'How can deterministic and two-stage stochastic models be used to plan hospital services for frail and elderly patients?'. These research aims will be answered in Chapter 5.

This research project was funded and carried out in collaboration with Clinical Futures in ABUHB [6]. The organisation are restructuring and organising hospital services in order to improve patient care. One of their primary issues is determining the number of beds required to meet the existing demands for their frail and elderly population. Wards in the UK have minimal staff to patient ratios that are typically computed based on ward bed counts. We can extend this model to incorporate workforce planning and the number of nursing staff the health board should have available by determining the number of beds for each specialty within each hospital to fulfil demand. Hospital admissions are the most utilised resources by frail and elderly, with many inpatient services such as surgical theatres, relying on sufficient hospital beds to be planned.

Two-stage stochastic programming lends itself well to this type of problem because bed numbers and staff must be planned in advance. Only when the demand for both elective and emergency patients is known can it be established whether sufficient beds and staff have been planned. If not, additional beds and staff are needed to meet this demand safely. Because this is not planned in advance, there are typically additional charges, such as relocating patients to different hospitals, opening new wards, and contacting agency and bank personnel. Before deciding on two-stage stochastic modelling for this project, a variety of other stochastic models were considered, However, due to the nature of the problem, this method was found to be the most suitable.

To identify the most cost-effective way to arrange specialised beds and nurses in hospitals, deterministic and two-stage stochastic modelling can be used. The health board has historically planned the number of beds and nursing staff using averages (deterministic). Due to emergency admissions, cancellations, and fluctuating admission LOS's, bed planning can be challenging. Stochastic modelling can help overcome this challenge and provide more informed outcomes. We present a two-stage stochastic model that takes system unpredictability into consideration when scheduling nursing staff and beds. Similar to the chapter on predictive analytics, a worked example will be provided to show how each method operates. Recall the worked example from Chapter 3, where 15 patients with various attributes attended one of two hospitals in South East Wales. Following is a table of these patients:

To ascertain whether there is a benefit to employing the two-stage stochastic model as opposed to the deterministic counterpart, the work by Maggioni and Wallace [257]

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

Table 4.1: Data set of 10 patients that will be used for the worked example.

will be utilised. Within their paper, the authors discuss the quality of the expected value solution in stochastic programming and apply their models to four case studies: A single-sink transportation problem, a production problem, a location routing (network design) problem and a mobile ad-hoc network problem. The authors outline four experiments in which they determine the factors that contributed to the expected value solution. Their work will be expanded upon by applying the theory to a different case study of bed and resource planning. Three of the four experiments (due to relevancy), will be applied to healthcare services for frail and elderly patients. This in turn will determine the factors that contribute to the expected value solution.

Using the deterministic and two-stage stochastic optimisation paradigms this chapter will examine how to effectively organise bed and staffing resources for hospitals. With these models, a user can modify the model to account for the number of hospitals and medical specialisations involved in their particular case. The remainder of the chapter is set out as follows: The prescriptive methods employed in this research and their current use in healthcare are covered in Section 4.2. The development of the deterministic model is shown in Section 4.3, whereas Section 4.4 provides a similar analysis using a two-stage stochastic model. The two approaches are combined in Section 4.5 to assess the performance of the stochastic or deterministic models. Throughout the chapter, a condensed practical example will be given.

4.2 Prescriptive Techniques

Prescriptive analytics is the process of using data and applying it to determine an optimal course of action. It builds upon the work of descriptive and predictive analytics and can be broken down into three approaches:

- Bridging the gap between potential and recommend outcomes

- Turning data into practical strategies
- Providing a clear path forward even with messy data

Prescriptive analytics incorporates large amounts of structured and unstructured data to determine the consequences of making decisions and how the future would be impacted. Furthermore, it can measure the repercussions of a decision based on different possible future scenarios. Organisations will increasingly need to find ways to take advantage of their data, especially as it continues to grow. Prescriptive analytics benefits organisations by allowing them to get the most out of their data and automate key procedures.

Utilising prescriptive analytics allows the user to make more informed, data-driven decisions and eliminates prejudice. It can also determine the likelihood of an organisation's chances of success, which will in turn lower risk and boost productivity. Worst-case scenarios can also be predicted more accurately, enabling organisations to plan accordingly. Implementing these strategies, though, may be challenging and necessitates that organisations have a clear understanding of both the questions to ask and how to respond. If the input assumptions are incorrect, the results will not be reliable. Additionally, prescriptive analytics may be expensive in terms of both time and financial commitment.

4.2.1 Basic Notation

Birge and Louveaux [258] provide a comprehensive introduction to stochastic programming. The authors cover a wide range of topics including the formulation of stochastic programming problems, the theory of stochastic optimisation, and algorithms for solving stochastic programmes. It also includes case studies and applications in finance, inventory management, transportation, and energy systems. Two-stage stochastic modelling has been applied to the area of healthcare previously, with Mestre et al.'s [259] research into location-allocation strategies for hospital network planning, given the uncertain nature of factors such as demand, capacity, and cost. More recently, Maggioni and Wallace [257] evaluated the traditional methods with a series of experiments to demonstrate the effectiveness of their approach. This research will follow the framework as used by Mestre et al. [259] and Maggioni and Wallace's [257]. The following notation and theory are taken from Maggioni and Wallace [257].

Two-stage stochastic modelling uses two stages in order to optimise a solution. Within the first stage, a decision is made without knowledge of what the future is to bring. The second stage sees the realisation of the stochastic elements of the problem. However, we are able to make further decisions to avoid the constraints of the problem becoming infeasible. In other words, to maintain feasibility in the second stage, we have recourse to a further degree of flexibility but at a cost. The second stage decisions will be dependent upon the stochastic elements observed, which can be implemented in standard notation form.

As previously discussed, a decision must be made about some random events prior to the experiment (without complete information). These are referred to as the first stage decision and are denoted by the vector x. Subsequently, we receive the experiment's results after taking into account some random factors $\xi(\omega)$, where ω is the outcome of an uncertain random experiment. The second stage decisions can then be calculated using the vector $y(\omega)$, which assumes that both x and $\xi(\omega)$ are fixed [258, 260, 261]. The remainder of this thesis will refer to $\xi(\omega)$ as $\boldsymbol{\xi}$.

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

$$\min_{x \in X} E_{\boldsymbol{\xi}} z(x, \boldsymbol{\xi}) = \min_{x \in X} \{f_1(x) + E_{\boldsymbol{\xi}}\left[h_2(x, \boldsymbol{\xi})\right]\} \tag{4.1}$$

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

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

$$h_2(x, \xi) = \min_{y \in Y(x, \xi)} f_2(y : x, \xi) \tag{4.2}$$

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

Equation (4.2) reflects the costs associated with the information being revealed through the realisation of ξ from the random vector $\boldsymbol{\xi}$. The term $[h_2(x, \boldsymbol{\xi})]$, 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:

$$RP = E_{\boldsymbol{\xi}} z(x^*, \boldsymbol{\xi}) \tag{4.3}$$

Equation (4.1) 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 problem.

$$EV = \min_{x \in X} z(x, \bar{\xi}) \tag{4.4}$$

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

4.2.2 Evaluation of Measures

Within prescriptive analytics, it is widely recognised that the expected value solution (EV) 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. Maggioni and Wallace [257] set out four tests in their research to determine the success of their models. Within this research, only the first three tests will be considered, since the fourth test is a generalisation of the second and does not provide any additional benefits. Similar to the previous subsection, the following notation is taken from Maggioni and Wallace [257].

4.2.2.1 Test A

The first traditional evaluation measure is to determine the value of the stochastic solution (VSS).

If we let $\bar{x}(\bar{\xi})$ be the optimal solution to Equation (4.4), we can take values and fix these as the first stage, and then allow the second stage of the stochastic model to be performed.

$$EEV = E_{\boldsymbol{\xi}}(z(\bar{x}(\bar{\xi}), \boldsymbol{\xi})) \tag{4.5}$$

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:

$$VSS = EEV - RP \tag{4.6}$$

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

4.2.2.2 Test B

The second test involves fixing the first stage variables which are at zero (or at their lower bound) to zero (or their lower bound) in the EV problem and then solving the stochastic programme. This will determine if the deterministic model produced the correct non-zero variables.

If the original problem is a linear programme, then Test B leads to solving a linear programme but one that is of smaller size. If it is a mixed binary programme, the test calls for fixing all the binary variables to either zero or one, and solving an easier linear programme. When a mixed integer programme (MIP) is involved, a smaller dimension MIP is solved.

Let $\mathcal{J}$ be the set of indices for which the components of the expected value solution $\bar{x}(\bar{\xi})$ are at zero or at their lower bound. Then let $\hat{x}$ be the solution of:

$$\begin{aligned} \min_{x \in X} \quad & E_{\boldsymbol{\xi}} z(x, \boldsymbol{\xi}) \\ \text{s.t.} \quad & x_j = \bar{x}_j(\bar{\xi}), \quad j \in \mathcal{J} \end{aligned} \tag{4.7}$$

We can then compute the expected skeleton solution value (ESSV):

$$ESSV = E_{\boldsymbol{\xi}}(z(\hat{x}, \boldsymbol{\xi})) \tag{4.8}$$

This can then be compared to the RP by means of loss using skeleton solution (LUSS):

$$LUSS = ESSV - RP \tag{4.9}$$

If the LUSS value is close to zero this means that the variables selected by the deterministic solution are good, however their values may be off. Therefore, we have:

$$RP \leq ESSV \leq EEV \tag{4.10}$$

and as a result the following is true:

$$EEV - EV \geq VSS \geq LUSS \geq 0 \tag{4.11}$$

In the case where LUSS = 0 (i.e., ESSV = RP), this corresponds to the perfect skeleton solution as the condition $x_j = \bar{x}_j(\bar{\xi}), \;\; j \in \mathcal{J}$, is satisfied by the stochastic solution x^* even without being enforced by a constraint (i.e., $\hat{x} = x^*$). However, if there exists $j \in \mathcal{J}$ such that x_j^* then $0 < LUSS < VSS$. LUSS = VSS occurs if the stochastic programme, when not allowed to use the variables in $\mathcal{J}$, chooses not to change the value of any of the remaining variables, i.e., $\hat{x} = \bar{x}(\bar{\xi})$.

4.2.2.3 Test C

The third test involves taking the EV solution ($\bar{x}(\bar{\xi})$) as a starting point to the stochastic model (4.1) and then comparing, in terms of the objective functions, to Equation (4.1) without such input. This determines whether the EV solution is upgradeable to become good (if not optimal) in the stochastic setting. This is equivalent to adding in an additional constraint, $x \geq \bar{x}(\bar{\xi})$, to Equation (4.1). As a result, the following problem is solved with solution $\tilde{x}$:

$$\min_{x \in X} \quad E_{\boldsymbol{\xi}} z(x, \boldsymbol{\xi}) \tag{4.12}$$

$$\text{s.t.} \quad x \geq \bar{x}(\bar{\xi}) \tag{4.13}$$

Then the expected input value (EIV) can be calculated:

$$EIV = E_{\boldsymbol{\xi}}(z(\tilde{x}, \boldsymbol{\xi})) \tag{4.14}$$

This is then compared to the RP, by means of the loss of upgrading the deterministic solution (LUDS):

$$LUDS = EIV - RP \tag{4.15}$$

Therefore,

$$RP \leq EIV \leq EEV \tag{4.16}$$

and the following is true:

$$EEV - EV \geq VSS \geq LUDS \geq 0 \tag{4.17}$$

In the case where LUDS = 0, and hence EIV = RP, this corresponds to perfect upgradeability of the deterministic solution. This occurs when the conditions $x \geq \bar{x}(\bar{\xi})$ are satisfied by the stochastic solution x^* without the need to be enforced by the constraints, $\tilde{x} = x^*$, (under the assumptions that the stochastic first-stage decision is unique). If there exists a component $i \in \{1, .., n\}$ such that $x_i^* < \bar{x}_i$, then $\tilde{x}_i = \bar{x}_i$ (case of partial upgradeability) and therefore, $0 < LUDS < VSS$. When LUDS = VSS, this corresponds to the non-upgradeability, in which the condition $x \geq \bar{x}(\bar{\xi})$ is no longer satisfied by any of the components of the solution x^* and then $\tilde{x} = \bar{x}(\bar{\xi})$ (i.e., EIV = EEV).

4.3 Deterministic Model Development

Within this section, the deterministic version of the model will be developed. The aim is to determine the number of beds and nursing staff required to satisfy a given demand, across different specialties within hospitals. All hospital beds and staffing resources are planned based on a fixed demand. If the number of beds within the hospital are insufficient to satisfy the demand and there is no option to open wards or transfer patients to other hospitals, then this results in a non optimal solution given. The number of beds in total must not exceed the bed capacity for that hospital and additionally, satisfactory levels of staff must be deployed to sufficiently open beds.

4.3.1 Sets

Within the deterministic model, we have four sets as given in Table 4.2.

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

Table 4.2: The sets used within the deterministic model where (B, S, H, R) represent the maximum number of nursing bands, specialties, hospitals and regions, respectively.

Each of the specialties must appear in at least one of the hospitals ($\mathcal{S} \in \mathcal{H}$). Similarly, each of the hospitals must appear in one of the regions ($\mathcal{H} \in \mathcal{R}$). Therefore $|\mathcal{H}| \leq |\mathcal{R}|$ is true. The set of nursing bands correspond to different skill levels and experience of nurses [262].

4.3.2 Parameters

Table 4.3 denotes the eight parameters used within the deterministic formulation, along with a brief definition.

4.3.3 Decision Variables

There are two decision variables for the model to calculate, as shown in Table 4.4.

The first decision variable determines the number of hospital beds to be planned for each specialty within each hospital. The second decision variable determines the number of staff required of each band for each specialty within each hospital.

4.3.4 Model

These sets, parameters and decision variables can be utilised within the following deterministic model with the objective function given in Equation (6.8) and the constraints listed between Equations (4.19) to (4.23).

Parameter	Definition
$c^{\text{bed}}_{s,h}$	Cost of bed per day for specialty $s \in \mathcal{S}$, in hospital $h \in \mathcal{H}$
c^{staff}_{b}	Cost of nursing staff per day of band $b \in \mathcal{B}$
$D_{s,r}$	Average daily bed demand for each specialty $s \in \mathcal{S}$, arriving from region $r \in \mathcal{R}$
$R_{s,b}$	Ratio of nursing staff of band $b \in \mathcal{B}$ to patient for specialty $s \in \mathcal{S}$
$K_{s,h}$	Maximum number of beds available to open in each specialty $s \in \mathcal{S}$, in hospital $h \in \mathcal{H}$
$\text{UB}^{\text{max, bed}}_{h}$	Upper bound of the number of beds that are able to be deployed in hospital $h \in \mathcal{H}$
$\text{UB}^{\text{max, staff}}_{b}$	Upper bound of the number of staff that can be deployed

Table 4.3: The parameters used within the deterministic model where b, s, h, r represent the variable number of nursing bands, specialties, hospitals and regions, respectively.

Decision Variable	Definition
$x^{\text{bed}}_{s,h} \in \mathbb{N}$	Number of beds planned for specialty $s \in \mathcal{S}$, in hospital $h \in \mathcal{H}$
$x^{\text{staff}}_{s,b,h} \in \mathbb{N}$	Number of staff planned for band $b \in \mathcal{B}$, for specialty $s \in \mathcal{S}$, in hospital $h \in \mathcal{H}$

Table 4.4: The decision variables used within the deterministic model where b, s, h represent the variable number of nursing bands, specialties and hospitals, respectively.

$$\min \sum_{h\in\mathcal{H}} \sum_{s\in\mathcal{S}} (c^{\text{bed}}_{s,h} x^{\text{bed}}_{s,h} + \sum_{b\in\mathcal{B}} c^{\text{staff}}_{b} x^{\text{staff}}_{s,b,h}) \tag{4.18}$$

subject to:

$$\sum_{h\in\mathcal{H}} x^{\text{bed}}_{s,h} \geq D_{s,r} \qquad \forall s \in \mathcal{S}, r \in \mathcal{R} \tag{4.19}$$

$$\sum_{b'\in\mathcal{B}:b'\geq b} x^{\text{staff}}_{s,b',h} \geq R_{s,b} \cdot x^{\text{bed}}_{s,h} \qquad \forall s \in \mathcal{S}, b \in \mathcal{B}, h \in \mathcal{H} \tag{4.20}$$

$$x^{\text{bed}}_{s,h} \leq K_{s,h} \qquad \forall s \in \mathcal{S}, h \in \mathcal{H} \tag{4.21}$$

$$0 \leq \sum_{s\in\mathcal{S}} x^{\text{bed}}_{s,h} \leq UB^{\text{max, bed}}_{h} \qquad \forall h \in \mathcal{H} \tag{4.22}$$

$$0 \leq \sum_{s\in\mathcal{S}} \sum_{h\in\mathcal{H}} x^{\text{staff}}_{s,b,h} \leq UB^{\text{max, staff}}_{b} \qquad \forall b \in \mathcal{B} \tag{4.23}$$

The first constraint (4.19), ensures the demand for each specialty and region is met by the number of hospital beds deployed. Constraint (4.20) ensures the staff to patient bed ratio is met. Constraint 4.21 requires the number of beds to be deployed to each specialty within each hospital to not exceed the capacity for that specialty and hospital. Equations (4.22) and (4.23) define the bounds for the decision variables of the problem.

4.3.5 Worked Example

Using the matrix defined earlier in the chapter (Table 4.1), we can use these parameters. Recall there are two hospitals within an area, each serving the same two specialties. There is only one region in which patients can arrive from, and each specialty bed in each hospital has a different cost. We assume there are two nursing staff band levels required on the wards for the specialties, with differing staff/bed ratios depending on the specialty. Table 4.5 displays the values of the parameters used within the illustrative example.

Parameters	**Values**
Bed Costs ($c^{\text{bed}}_{s,h}$)	$\begin{bmatrix} 20 & 30 \\ 30 & 40 \end{bmatrix}$
Ratio ($R_{s,b}$)	$\begin{bmatrix} 0.29 & 0.14 \\ 0.14 & 0.29 \end{bmatrix}$
Maximum Specialty Capacity ($K_{s,h}$)	$\begin{bmatrix} 20 & 25 \\ 20 & 25 \end{bmatrix}$
Staff Costs (c^{staff}_b)	[£50, £60]
Upper bed limit ($UB^{\text{max,bed}}_h$)	[20,25]
Upper staff limit ($UB^{\text{max,staff}}_b$)	[15,25]

Table 4.5: The parameter values that will be used within the deterministic model specifically for the worked example.

The matrices are stored in row-major order. The first index of the parameter refers to the row of the matrix. For the parameter $c^{\text{bed}}_{s,h}$, specialty one corresponds to row one, i.e., [20 30], while specialty two corresponds to the second row, i.e., [30 40]. The column of the matrix is referenced by the second index. Accordingly, for the parameter $c^{\text{bed}}_{s,h}$, hospital one refers to column one, i.e., $\begin{bmatrix} 20 \\ 30 \end{bmatrix}$, and hospital two refers to column two, i.e., $\begin{bmatrix} 30 \\ 40 \end{bmatrix}$. Therefore $c^{\text{bed}}_{1,2}$ denotes the top-right corner of the two by two matrix and the value of 30.

The demand from the matrix can be calculated as follows to determine the average daily bed demand, i.e., $D_{s,r}$:

$$\text{Average daily bed demand}_{s,h} = \text{Average LOS}_{s,h} \times \text{Average daily number of admissions}_{s,h} \quad (4.24)$$

$$D_{s,r} = \text{Average daily bed demand}_{s,r} = \sum_{h \in \mathcal{R}} \text{Average daily bed demand}_{s,h} \tag{4.25}$$

The model can be run with this as the medium daily bed demand - we can also artificially create a low (-10%) and high level (+10%) daily bed demand to determine how robust the model is.

Table 4.6 displays the average daily bed demand for the worked example. The average daily bed demands were calculated using Equations (4.24) and (4.25). The high and low demands were then artificially created.

	Low Daily Bed Demand	**Average Daily Bed Demand**	**High Daily Bed Demand**
$D_{0,0}$	15.00001	16.66668	18.33333
$D_{1,0}$	17.10000	19.00002	20.90000

Table 4.6: The daily bed demand figures for each region and specialty ($D_{s,r}$), categorised by low, average and high.

Inputting these parameters into the model, we can determine the value of the deterministic model, for the high, average and low daily bed demand cases. Within Table 4.7, the values for each of the decision variables, $x^{\text{bed}}_{s,h}$ and $x^{\text{staff}}_{s,b,h}$, and the objective function are displayed.

Decision Variable	**Low Daily Bed Demand**	**Average Daily Bed Demand**	**High Daily Bed Demand**
$x^{\text{bed}}_{0,0}$	13	0	19
$x^{\text{bed}}_{0,1}$	3	17	0
$x^{\text{bed}}_{1,0}$	6	20	0
$x^{\text{bed}}_{1,1}$	12	0	21
$x^{\text{staff}}_{0,0,0}$	4	0	6
$x^{\text{staff}}_{0,0,1}$	1	5	0
$x^{\text{staff}}_{0,1,0}$	2	0	3
$x^{\text{staff}}_{0,1,1}$	1	3	0
$x^{\text{staff}}_{1,0,0}$	1	3	0
$x^{\text{staff}}_{1,0,1}$	2	0	3
$x^{\text{staff}}_{1,1,0}$	2	6	0
$x^{\text{staff}}_{1,1,1}$	4	0	7
Objective	£1,950.00	£2,050.00	£2,270.00

Table 4.7: The deterministic results for each of the decision variables x^{bed} and x^{staff} categorised by the low, average and high daily bed demands.

The EV values for the low, average and high daily bed demands were £1,950.00, £2,050.00 and £2,270.00 respectively. The results show there was a larger increase when comparing the average to high demand than the low to average demand (Figure 4.1).

When comparing the low to average level daily bed demands, there is an increase of 10% in the daily demands. In turn, this increases the total costs by 5.12%. Similarly, the average to high level demand increase of 10% shows there is an increase of 10.73% in the objective function. The low to high daily demands cause an additional 20% in demand, with a total increase of 16.41% in costs.

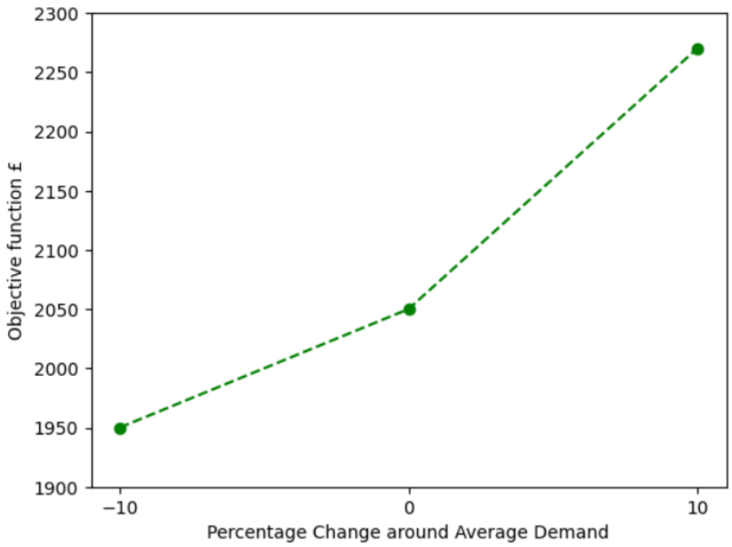


Figure 4.1: Line graph showing the objective function for the deterministic implementation against the different daily demand levels, where -10% indicated the low daily bed demand and +10% indicates the high daily bed demand.

4.4 Two-Stage Stochastic Model Extension

The two-stage stochastic model is an extension of the deterministic model built in Section 4.3. Since this is a two-stage model, Equations (4.1) and (4.2) can be adapted in order to apply these to the case of bed and staffing capacity planning.

All hospital beds are planned in advance before the demand for each specialty and hospital is known. Similarly, nursing staff are deployed to specialties within the hospitals. When the demand is known, there is the option to change some specialty beds to alternative specialty beds with a cost. The two-stage version of this model enables various scenarios to be determined; if either the demand is changed, maximum capacities are changed or costs are changed. If the number of beds within the hospital is insufficient to satisfy the demand, then a patient may be transferred to another hospital with availability, or a new ward will be opened within the original hospital location, with additional associated costs. Ultimately, the total number of beds must not exceed the bed capacity for each individual hospital. Additionally, sufficient nursing staff must be deployed to open beds and maintain the staff/bed ratios, otherwise, bank staff would have to be used, increasing costs.

Recall, the ultimate objective of the model is to determine, for each hospital, the number of beds and staff required for each specialty in order to minimise the total costs, given by the sum of opening specialty beds within the hospital and the transportation costs of moving patients between hospitals.

4.4.1 Sets

In addition to the sets within the deterministic model, there is also the set of scenarios in which different demands and situations can be run through the model.

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

Table 4.8: The sets used within the two-stage stochastic model where (B, S, H, R, K) rep- resent the maximum number of nursing bands, specialties, hospitals, regions and scenarios, respectively

4.4.2 Parameters

Table 4.9 denotes the parameters used within the two-stage stochastic model, along with a definition for each parameter.

Parameter	Definition
$\mathrm{c}_{s,h}^{\text{bed, 1st}}$	Cost of the first stage bed per day for specialty $s \in \mathcal{S}$, in hospital $h \in \mathcal{H}$
$\mathrm{c}_{s,h}^{\text{bed, 2nd}}$	Cost of the second stage bed per day for specialty $s \in \mathcal{S}$, in hospital $h \in \mathcal{H}$
$\mathrm{c}_{b}^{\text{staff, 1st}}$	Cost of the first stage staff per day of band $b \in \mathcal{B}$
$\mathrm{c}_{b}^{\text{staff, 2nd}}$	Cost of the second stage staff per day of band $b \in \mathcal{B}$
p_k	Probability of scenario $k \in \mathcal{K}$
$\mathrm{D}_{s,r,k}$	Average daily bed demand for each specialty $s \in \mathcal{S}$
	arriving from region $r \in \mathcal{R}$, for scenario $k \in \mathcal{K}$
$\mathrm{R}_{s,b}$	Ratio of nursing staff of band $b \in \mathcal{B}$ to patient for each specialty $s \in \mathcal{S}$
$K_{s,h}$	Maximum number of beds available to open in each specialty $s \in \mathcal{S}$
	in hospital $h \in \mathcal{H}$
$\mathrm{UB}_{h}^{\text{max, bed, 1st}}$	Upper bound of the number of beds that are able to be deployed in hospital $h \in \mathcal{H}$ in the first stage
$\mathrm{UB}_{h}^{\text{max, bed, 2nd}}$	Upper bound of the number of beds that are able to be deployed
	in hospital $h \in \mathcal{H}$ in the second stage
$\mathrm{UB}_{b}^{\text{max, staff, 1st}}$	Upper bound of the number of staff that can be deployed in the 1^{st} stage
$\mathrm{UB}_{b}^{\text{max, staff, 2nd}}$	Upper bound of the number of staff that can be deployed in the 2^{nd} stage

Table 4.9: The parameters 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

4.4.3 Decision Variables

The decision variables that were listed in Table 4.4 can be extended to include the second stage decision variables, given in Table 4.10.

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

Table 4.10: The sets used within the two-stage stochastic model where (b, s, h, r, k) represent the value number of nursing bands, specialties, hospitals, regions and scenarios, respectively

4.4.4 Model

The following two-stage stochastic model can be determined, adapted from Equations (6.8) - (4.23), resulting in the Equations (4.26) - (4.35).

$$\begin{aligned}\min \sum_{h\in\mathcal{H}}\sum_{s\in\mathcal{S}}(c_{s,h}^{\text{bed, 1st}} x_{s,h}^{\text{bed}} + \sum_{b\in\mathcal{B}} c_b^{\text{staff, 1st}} x_{s,b,h}^{\text{staff}}) \\ + \sum_{k\in\mathcal{K}}\sum_{h\in\mathcal{H}}\sum_{s\in\mathcal{S}} p_k(c_{s,h}^{\text{bed, 2nd}} u_{s,h,k}^{\text{bed}} + \sum_{b\in\mathcal{B}} c_b^{\text{staff, 2nd}} u_{s,b,k,h}^{\text{staff}})\end{aligned} \tag{4.26}$$

subject to:

$$\sum_{h\in\mathcal{H}}(x^{\text{bed}}_{s,h}+u^{\text{bed}}_{s,h,k}) \geq D_{s,r,k} \qquad \forall s\in\mathcal{S}, r\in\mathcal{R}, k\in\mathcal{K} \tag{4.27}$$

$$\sum_{b'\in\mathcal{B}:b'\geq b} x^{\text{staff}}_{s,b',h} \geq R_{s,b}\cdot x^{\text{bed}}_{s,h} \qquad \forall s\in\mathcal{S}, b\in\mathcal{B}, h\in\mathcal{H} \tag{4.28}$$

$$\sum_{b'\in\mathcal{B}:b'\geq b} u^{\text{staff}}_{s,b',k,h} \geq R_{s,b}\cdot u^{\text{bed}}_{s,h,k} \qquad \forall s\in\mathcal{S}, b\in\mathcal{B}, h\in\mathcal{H}, k\in\mathcal{K} \tag{4.29}$$

$$x^{\text{bed}}_{s,h} \leq K_{s,h} \qquad \forall s\in\mathcal{S}, h\in\mathcal{H} \tag{4.30}$$

$$u^{\text{bed}}_{s,h,k} \leq K_{s,h} \qquad \forall s\in\mathcal{S}, h\in\mathcal{H}, k\in\mathcal{K} \tag{4.31}$$

$$0 \leq \sum_{s\in\mathcal{S}} x^{\text{bed}}_{s,h} \leq UB^{\text{max, bed, 1st}}_{h} \qquad \forall h\in\mathcal{H} \tag{4.32}$$

$$0 \leq \sum_{s\in\mathcal{S}}\sum_{h\in\mathcal{H}} x^{\text{staff}}_{s,b,h} \leq UB^{\text{max, staff, 1st}}_{b} \qquad \forall b\in\mathcal{B} \tag{4.33}$$

$$0 \leq \sum_{s\in\mathcal{S}} u^{\text{bed}}_{s,h,k} \leq UB^{\text{max, bed, 2nd}}_{h} \qquad \forall h\in\mathcal{H}, k\in\mathcal{K} \tag{4.34}$$

$$0 \leq \sum_{s\in\mathcal{S}}\sum_{h\in\mathcal{H}} u^{\text{staff}}_{s,b,k,h} \leq UB^{\text{max, staff, 2nd}}_{b} \qquad \forall b\in\mathcal{B}, k\in\mathcal{K} \tag{4.35}$$

The first sum in the objective function (4.26) 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 region. The first constraint, (4.27), assures the demand for each specialty and region is met by the number of hospital beds deployed. Constraint (4.28) ensures the number of staff deployed meets the minimum requirements for staff on each specialty ward in the first stage, whilst Constraint (4.29) ensures this requirement is met in the second stage. Constraints (4.30) and (4.31) assures the beds deployed do not exceed the maximum number of beds available for each specialty within each hospital. Equations (4.32) - (4.35) define the maximum bounds on the first and second stage decision variables of the problem.

4.4.5 Worked Example

Using the 15 patients discussed previously, these can be applied to the two-stage stochastic model. Table 4.5 can be extended to generate Table 4.11 to include the second stage model parameters.

The matrices are stored in row-major order. For the terms with two indices, the first index of the parameter refers to the row of the matrix, and the second refers to the column of the matrix.

Similarly, Equations (4.24) and (4.25) can be used to determine the average daily bed demand. Three scenarios are studied:

Parameters	**Values**
1^{st} Stage Bed Costs ($c_{s,h}^{\text{bed,1st}}$)	$\begin{bmatrix} 20 & 30 \\ 30 & 40 \end{bmatrix}$
2^{st} Stage Bed Costs ($c_{s,h}^{\text{bed,2nd}}$)	$\begin{bmatrix} 22 & 33 \\ 33 & 44 \end{bmatrix}$
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^{st} Stage Staff Costs ($c_b^{\text{staff, 1st}}$)	[£50, £60]
2^{nd} Stage Staff Costs ($c_b^{\text{staff, 2nd}}$)	[£55, £66]
Upper 1^{st} bed limit ($UB_h^{\text{max,bed,1st}}$)	[20,25]
Upper 2^{nd} bed limit ($UB_{h,k}^{\text{max,bed,2nd}}$)	$\begin{bmatrix} 20 & 20 & 20 \\ 25 & 25 & 25 \end{bmatrix}$
Upper 1^{st} staff limit ($UB_b^{\text{max,staff,1st}}$)	[15,25]
Upper 2^{nd} staff limit ($UB_{b.k}^{\text{max,staff,2nd}}$)	$\begin{bmatrix} 15 & 15 \\ 25 & 25 \end{bmatrix}$
Probability of Scenarios (p_k)	[0.4,0.3,0.3]

Table 4.11: The parameter values that will be used within the two-stage stochastic model specifically for the worked example

- 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%

The model is additionally run with low demand (-10%) and high demand (+10%) with the same scenario percentages. These values can be represented in a matrix.

	Low Demand			**Average Demand**			**High Demand**		
	k=0	**k=1**	**k=2**	**k=0**	**k=1**	**k=2**	**k=0**	**k=1**	**k=2**
$D_{0,0,k}$	15.001	18.002	12.001	16.660	19.992	13.328	18.33	21.996	14.664
$D_{1,0,k}$	17.100	20.520	13.680	19.001	22.80	15.200	20.900	25.080	16.720

Table 4.12: The daily bed demand figures for each region and specialty $D_{s,r,k}$, categorised by low, average and high.

For example the average demand $D_{s,r,k}$ matrix would be represented as:

$$\begin{bmatrix} [[16.660, 19.992, 13.328]] \\ [[19.000, 22.800, 15.200]] \end{bmatrix}$$

The demand term is stored as a three-dimensional array ($D_{s,r,k}$). The first index refers to the row and the second index corresponds to the column. In this case, we only have one region, so only one matrix column is shown. The third index refers to the column inside the sub-matrix. To illustrate, if we refer to the first hospital and region ($D_{0,0,k}$), then, if $k = 0$, we refer to the first element inside the [0,0] matrix, i.e., 16.660. If $k = 1$, we refer to the second element inside the [0,0] matrix, i.e., 19.000.

Table 4.12 displays the demands for each scenario of the low, average and high demands for the two-stage stochastic problem.

These demands can be input into the two-stage stochastic model to obtain the objective function for each case. Table 4.13 shows the values for each of the decision variables, $x^{\text{bed}}_{s,h}$, $x^{\text{staff}}_{s,b,h}$, $x^{\text{bed}}_{s,r,h,k}$ and $u^{\text{staff}}_{s,b,h,k}$, is displayed.

1^{st} **Stage Decision Variables**	**Low Demand**	**Average Demand**	**High Demand**
$x_{0,0}^{\text{bed}}$	13	0	6
$x_{0,1}^{\text{bed}}$	0	0	0
$x_{1,0}^{\text{bed}}$	6	20	13
$x_{1,1}^{\text{bed}}$	6	0	0
$x_{0,0,0}^{\text{staff}}$	4	0	2
$x_{0,0,1}^{\text{staff}}$	0	0	0
$x_{0,1,0}^{\text{staff}}$	2	0	1
$x_{0,1,1}^{\text{staff}}$	0	0	0
$x_{1,0,0}^{\text{staff}}$	1	3	2
$x_{1,0,1}^{\text{staff}}$	1	0	0
$x_{1,1,0}^{\text{staff}}$	2	6	4
$x_{1,1,1}^{\text{staff}}$	2	0	0

2^{nd} **Stage Decision Variables**	**Low Demand**			**Average Demand**			**High Demand**		
	k=1	**k=2**	**k=3**	**k=1**	**k=2**	**k=3**	**k=1**	**k=2**	**k=3**
$u_{0,0,k}^{\text{bed}}$	3	4	3	17	20	14	13	17	10
$u_{0,1,k}^{\text{bed}}$	0	5	0	0	1	0	0	0	0
$u_{1,0,k}^{\text{bed}}$	6	4	0	0	4	0	8	13	4
$u_{1,1,k}^{\text{bed}}$	0	3	0	0	0	0	0	0	0
$u_{0,0,0,k}^{\text{staff}}$	1	3	0	5	6	5	4	5	3
$u_{0,0,1,k}^{\text{staff}}$	0	0	0	0	1	0	0	0	0
$u_{0,1,0,k}^{\text{staff}}$	1	1	0	3	3	2	2	3	2
$u_{0,1,1,k}^{\text{staff}}$	0	0	0	0	1	0	0	0	0
$u_{1,0,0,k}^{\text{staff}}$	1	2	1	0	1	0	2	2	1
$u_{1,0,1,k}^{\text{staff}}$	0	0	0	0	0	0	0	0	0
$u_{1,1,0,k}^{\text{staff}}$	2	3	1	0	2	0	3	4	2
$u_{1,1,1,k}^{\text{staff}}$	0	0	0	0	0	0	0	0	0
Objective Function Value	£1,996.40			£2,185.20			£2,351.60		

Table 4.13: The two-stage stochastic results for each of the decision variables x^{bed}, x^{staff}, u^{bed} and x^{staff}, categorised by low, average and high daily bed demands.

The RP values for the three demand levels are £1,996.40, £2,185.20, and £2,351.60 for the low, average and high levels of demand. All three results show that there is a linear relationship, with the increases being almost directly proportional (Figure 4.2).

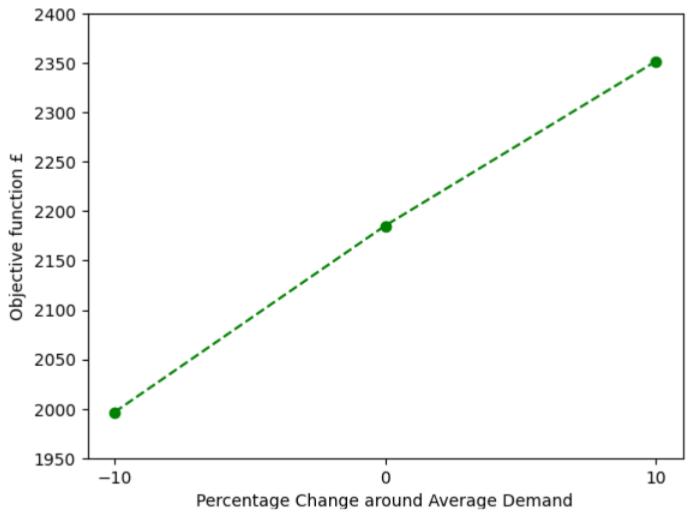


Figure 4.2: Line graph showing the objective function for the two-stage stochastic implementation against the different daily demand levels, where -10% indicates the low daily bed demand and +10% indicated the high daily bed demand.

The low to average level demands show a 9.46% increase which is similar to the 10% increase in demand. Similarly, average to high level demands yield a 7.61% increase. The low to high daily demands cause an additional 20% in costs, with a total increase of 17.79% in demands.

4.5 Evaluation of Measures

This section looks at applying the three tests discussed in Sections 4.2.2.1, 4.2.2.2 and 4.2.2.3 to the worked example of frail and elderly patient services. Furthermore, it will determine how the results can be evaluated and determine where the model is under performing.

4.5.1 Test A

Test A involves calculating the VSS to determine the expected loss when using the deterministic solution. If we recall Equations (4.5) and (4.6), which calculates the EEV and VSS respectively.

$$EEV = E_{\boldsymbol{\xi}}(z(\bar{x}(\bar{\xi}), \boldsymbol{\xi})) \qquad \text{(4.5 revisited)}$$

$$VSS = EEV - RP \qquad \text{(4.6 revisited)}$$

Tables 4.14 - 4.16 display the results for calculating the VSS for the worked example, with low, average and high demand respectively. The deterministic model deploys fewer beds and nursing staff than the stochastic model in each of the three cases, leading to an EV solution that costs approximately two-thirds of the RP. However, the EEV is larger in all cases and this results in the following VSS scores:

$$VSS_{Low} = 2,140.80 - 1,996.40 = £144.40 \ (7.23\%) \tag{4.36}$$

$$VSS_{Average} = 2,240.80 - 2,185.20 = £55.60 \ (5.4\%) \tag{4.37}$$

$$VSS_{High} = 2,668.40 - 2,351.60 = £316.80 \ (13.47\%) \tag{4.38}$$

This demonstrates that utilising the stochastic model rather than the deterministic, can save between 5.4% and 13.47% in bed and staffing costs.

	s=0, h=0	**s=0, h=1**	**s=1, h=0**	**s=1, h=1**	**Objective Value (£)**
Deterministic	[(6), (2, 1)]	[(10), (3, 2)]	[(13), (2, 4)]	[(5), (1, 2)]	1,950.00 = EV
Stochastic	[(20), (7, 3)]	[(0), (0, 0)]	[(16), (3, 5)]	[(6), (1, 2)]	1,996.40 = RP
Test A	[(17), (6, 3)]	[(3), (1,1)]	[(10), (2,4)]	[(12), (2,4)]	2,140.80 = EEV

Table 4.14: Test A results for the worked example using low daily bed demand values, with results recorded in the form [(beds), (staff)].

	s=0, h=0	**s=0, h=1**	**s=1, h=0**	**s=1, h=1**	**Objective Value (£)**
Deterministic	[(0), (0,0)]	[(17), (5, 3)]	[(20), (3, 6)]	[(0), (0, 0)]	2,050.00= EV
Stochastic	[(20), (6, 3)]	[(1), (1, 1)]	[(24), (4, 8)]	[(0), (0, 0)]	2,185.20= RP
Test A	[(4), (2, 1)]	[(17), (5, 3)]	[(24), (4, 8)]	[(0), (0, 0)]	2,240.80 = EEV

Table 4.15: Test A results for the worked example using average daily bed demand values, with results recorded in the form [(beds), (staff)].

	s=0, h=0	**s=0, h=1**	**s=1, h=0**	**s=1, h=1**	**Objective Value (£)**
Deterministic	[(19), (6, 3)]	[(0), (0, 0)]	[(0), (0, 0)]	[(21), (3, 7)]	2,270.00 = EV
Stochastic	[(23, (7, 4)]	[(0), (0, 0)]	[(14), (4, 8)]	[(0), (0, 0)]	2,351.60 = RP
Test A	[(22), (7, 4)]	[(0), (0, 0)]	[(14), (2, 3)]	[(21), (3, 7)]	2,668.40 = EEV

Table 4.16: Test A results for the worked example using high daily bed demand values, with results recorded in the form [(beds), (staff)].

To determine the exact reason why the deterministic model performed poorly, further tests were conducted.

4.5.2 Test B

Test B takes the skeleton solution from the deterministic model. Since the deterministic model did not result in any zero values being produced, the lower bound values instead will be used.

We calculate the ESSV as follows:

$$ESSV = E_{\boldsymbol{\xi}}(z(\hat{x}, \boldsymbol{\xi})) \tag{4.8 revisited}$$

and then use this information to calculate the LUSS:

$$LUSS = ESSV - RP \tag{4.9 revisited}$$

Tables 4.17 - 4.19, present the results after the ESSV calculations have been calculated. For each case, the ESSV is larger than the RP and therefore the LUSS is as follows:

$$LUSS_{Low} = 1,996.40 - 1,996.40 = £0 \tag{4.39}$$

$$LUSS_{Average} = 2,240.80 - 2,185.20 = £55.6 \tag{4.40}$$

$$LUSS_{High} = 2,408.40 - 2,351.60 = £56.80 \tag{4.41}$$

Since the model is fixed by how many beds it can deploy in the first stage, the demand is not satisfied within the first stage. The model reacts by opening beds in the second stage, in the facilities where this is restricted in the first stage. This occurs because the second stage costs for some bed specialties are in fact less expensive than first stage costs in other bed specialties.

	s=0, h=0	**s=0, h=1**	**s=1, h=0**	**s=1, h=1**	**Objective Value (£)**
Deterministic	[(6), (2, 1)]	[(10), (3, 2)]	[(13), (2, 4)]	[(5), (1, 2)]	1,950.00 = EV
Stochastic	[(20), (7, 3)]	[(0), (0, 0)]	[(16), (3, 5)]	[(6), (1, 2)]	1,996.40 = RP
Test A	[(17), (6, 3)]	[(3), (1,1)]	[(10), (2,4)]	[(12), (2,4)]	2,140.80 = EEV
Test B	[(20), (7, 3)]	[(0), (0, 0)]	[(16), (3, 5)]	[(6), (1, 2)]	1,996.40 = ESSV = RP

Table 4.17: Test B results for the worked example using low daily bed demand values, with results recorded in the form [(beds), (staff)].

	s=0, h=0	**s=0, h=1**	**s=1, h=0**	**s=1, h=1**	**Objective Value (£)**
Deterministic	[(0), (0,0)]	[(17), (5, 3)]	[(20), (3, 6)]	[(0), (0, 0)]	2,050.00= EV
Stochastic	[(20), (6, 3)]	[(1), (1, 1)]	[(24), (4, 8)]	[(0), (0, 0)]	2,185.20= RP
Test A	[(4), (2, 1)]	[(17), (5, 3)]	[(24), (4, 8)]	[(0), (0, 0)]	2,240.80 = EEV
Test B	[(4), (2, 1)]	[(17), (5, 3)]	[(24), (4, 8)]	[(0), (0, 0)]	2,240.80 = ESSV = EEV

Table 4.18: Test B results for the worked example using average daily bed demand values, with results recorded in the form [(beds), (staff)].

4.5.3 Test C

Test C determines the upgradeability of the model. The number of beds and staff deployed in the deterministic solution are added as constraints, where these

	s=0, h=0	s=0, h=1	s=1, h=0	s=1, h=1	Objective Value (£)
Deterministic	[(19), (6, 3)]	[(0), (0, 0)]	[(0), (0, 0)]	[(21), (3, 7)]	2,270.00 = EV
Stochastic	[(23, (7, 4)]	[(0), (0, 0)]	[(14), (4, 8)]	[(0), (0, 0)]	2,351.60 = RP
Test A	[(22), (7, 4)]	[(0), (0, 0)]	[(14), (2, 3)]	[(21), (3, 7)]	2,668.40 = EEV
Test B	[(23), (7,4)]	[(0), (0, 0)]	[(9), (2, 3)]	[(17), (3, 5)]	2,408.40 = ESSV

Table 4.19: Test B results for the worked example using high daily bed demand values, with results recorded in the form [(beds), (staff)].

minimum numbers have to be met in the first stage. Equation (4.14) calculates the expectant result from this test:

$$EIV = E_{\boldsymbol{\xi}}(z(\tilde{x}, \boldsymbol{\xi})) \qquad \text{(4.14 revisited)}$$

$$LUDS = EIV - RP \qquad \text{(4.15 revisited)}$$

The LUDS values for each case are as follows:

$$LUDS_{Low} = 2,235.40 - 1,996.40 = £239 \qquad (4.42)$$

$$LUDS_{Average} = 2,240.80 - 2,185.20 = £55.6 \qquad (4.43)$$

$$LUDS_{High} = 2,565.40 - 2,351.60 = £213.80 \qquad (4.44)$$

Tables 4.20 - 4.22 display the results for the EIV. For all cases, the LUDS value is greater than zero but less than the VSS. This demonstrates partial upgradeability.

	s=0, h=0	s=0, h=1	s=1, h=0	s=1, h=1	Objective Value (£)
Deterministic	[(6), (2, 1)]	[(10), (3, 2)]	[(13), (2, 4)]	[(5), (1, 2)]	1,950.00 = EV
Stochastic	[(20), (7, 3)]	[(0), (0, 0)]	[(16), (3, 5)]	[(6), (1, 2)]	1,996.40 = RP
Test A	[(17), (6, 3)]	[(3), (1,1)]	[(10), (2,4)]	[(12), (2,4)]	2,140.80 = EEV
Test B	[(20), (7, 3)]	[(0), (0, 0)]	[(16), (3, 5)]	[(6), (1, 2)]	1,996.40 = ESSV = RP
Test C	[(17), (6, 3)]	[(3), (1,1)]	[(9), (2,3)]	[(13), (2,3)]	2,235.40 = EIV

Table 4.20: Test C results for the worked example using low daily bed demand values, with results recorded in the form [(beds), (staff)].

	s=0, h=0	s=0, h=1	s=1, h=0	s=1, h=1	Objective Value (£)
Deterministic	[(0), (0,0)]	[(17), (5, 3)]	[(20), (3, 6)]	[(0), (0, 0)]	2,050.00= EV
Stochastic	[(20), (6, 3)]	[(1), (1, 1)]	[(24), (4, 8)]	[(0), (0, 0)]	2,185.20= RP
Test A	[(4), (2, 1)]	[(17), (5, 3)]	[(24), (4, 8)]	[(0), (0, 0)]	2,240.80 = EEV
Test B	[(4), (2, 1)]	[(17), (5, 3)]	[(24), (4, 8)]	[(0), (0, 0)]	2,240.80 = ESSV = EEV
Test C	[(4), (2, 1)]	[(17), (5, 3)]	[(24), (4, 8)]	[(0), (0, 0)]	2,240.80 = EIV = ESSV = EEV

Table 4.21: Test C results for the worked example using average daily bed demand values, with results recorded in the form [(beds), (staff)].

4.5.4 Worked Example Conclusion

The worked example provided three low, average and high demand scenarios to determine how beds and staff can be planned for two specialties within two

	s=0, h=0	**s=0, h=1**	**s=1, h=0**	**s=1, h=1**	**Objective Value (£)**
Deterministic	[(19), (6, 3)]	[(0), (0, 0)]	[(0), (0, 0)]	[(21), (3, 7)]	2,270.00 = EV
Stochastic	[(23, (7, 4)]	[(0), (0, 0)]	[(14), (4, 8)]	[(0), (0, 0)]	2,351.60 = RP
Test A	[(22), (7, 4)]	[(0), (0, 0)]	[(14), (2, 3)]	[(21), (3, 7)]	2,668.40 = EEV
Test B	[(23), (7,4)]	[(0), (0, 0)]	[(9), (2, 3)]	[(17), (3, 5)]	2,408.40 = ESSV
Test C	[(23), (7, 4)]	[(0), (0, 0)]	[(5), (1, 2)]	[(21), (3, 7)]	2,565.40 = EIV

Table 4.22: Test C results for the worked example using high daily bed demand values, with results recorded in the form [(beds), (staff)].

hospitals. To replicate both real world planning and real world events, these cases were performed in both a deterministic and stochastic environment.

Figure 4.3 provides a visual for the results, showing the differences between the deterministic and stochastic models for all three cases.

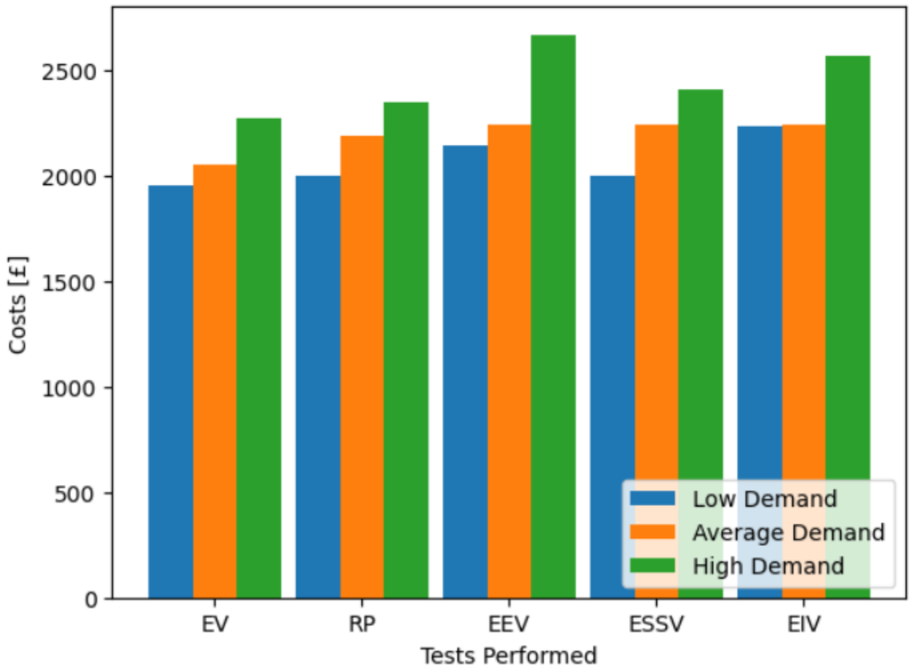


Figure 4.3: Clustered bar chart showing the objective function value for the deterministic and two-stage stochastic models. Included within this is the EEV, ESSV and EIV results for the low, average and high daily bed demands.

The results show the deterministic solution does not perform well in a stochastic environment, because of the too low number of beds and staff deployed at the first stage. By performing Test A, this showed there may be value in considering the deterministic solution as a lower bound for the stochastic case. Test B has shown that the deterministic solution performed poorly because it deployed the incorrect number of beds and staff. Finally, Test C, showed that even if some beds and staff are planned using the deterministic model, the stochastic model provides options on how to upgrade or improve the solution. It also informed as to where shortages are likely to be expected.

4.6 Summary

Prescriptive analytics offer healthcare professionals the opportunity to optimise outcomes by recommending the best course of action for patients or providers. Often, healthcare decisions are made using simple tools or are based on staff intuition, potentially leading to less than ideal outcomes by subjecting patients to unnecessary risks. Using more evidence-based and unbiased approaches, prescriptive analytics can be driven by real-time data which is routinely collected. Due to the ability to perform numerous scenario analysis, the impact of selecting different actions can be evaluated prior to implementation.

This chapter has provided an introduction into two prescriptive methodologies, deterministic and two-stage stochastic modelling. Expanding on the two-stage stochastic programming paradigm and building on the tests introduced by Maggioni and Wallace [257], this chapter has gone further by creating two-dimensional decision variables which are dependent on each other along with the application to a different field of research, namely healthcare. To demonstrate the procedure and the robustness of the models, a simplified worked example including two hospitals and two specialties was used. The tests discussed in [257], have also been employed, applied and evaluated to each of the examples.

In the following chapter, Chapter 5, the theory from both this and the previous chapter will be applied to a case study of frail and elderly patients. The deterministic and two-stage stochastic models developed will be expanded to a case of 14 hospitals and 29 specialties, to determine how to effectively organise bed numbers and staffing requirements to fulfil current and future demands, whilst being mindful of increasing medical costs.

Chapter 5

Experimental Analysis of the Predictive and Prescriptive Models

5.1 Introduction

The field of data science and OR has witnessed a significant evolution in recent years, with the development of predictive and prescriptive models emerging as an important aspect of the discipline. Predictive models aim to forecast future events or trends based on historical data, while prescriptive models focus on providing recommendations for actions to be taken based on the predictions made by predictive models. Both types of models have been widely adopted across various domains, including business, healthcare, and social media. This chapter discusses the experimental results for the predictive and prescriptive theory discussed in Chapters 3 and 4 respectively.

Research Aim - This chapter will utilise the methods discussed in the previous two chapters to apply to data from ABUHB in order to answer the following two research questions:

1. How do the clinical and demographical attributes of frail and elderly patients effect their length of stay within hospital? - Section 5.3
2. How best can specialties be organised among a network of hospitals to ensure staffing and bed costs are minimised whilst, whilst still meeting the demand for frail and elderly patients? - Section 5.4

The remainder of the Chapter is structured as follows: Section 5.2 introduces three years' worth of data from ABUHB used within this research, highlighting key insights. Section 5.3 applies predictive analytics to the ABUHB data, discussing linear and logistic regression and CART models. Section 5.4 will develop the deterministic and two-stage stochastic models within Microsoft Excel OpenSolver and Python PuLP.

5.2 Data Introduction

This section provides an overview of the data received from ABUHB, to gain a deeper understanding into the current practice and trends within ABUHB. Three years' worth of data was analysed ranging from April 2017 to March 2020, with two data sets being amalgamated to show an insight into the overall pathway. The first data set is from Myrddin [263], the patient administration system (PAS). Myrddin stores all patient contact details, outpatient appointments, generates letters for patients, and specifically for this research, inpatient information. The second data set is the Welsh Radiological Information System (RadIS), [263]. The RadIS IT system records and keeps track of which patients have received scans as well as the data that is gathered in conjunction with this.

Prior to any data analysis occurring, data cleansing was performed to ensure that all incorrect and incomplete entries were removed. Additionally, to make sure the patients were pertinent to the study, additional criteria were established. The following criteria were set:

1. Only complete patient information files were included i.e., no missing entries. If a patient did not have an NHS number or they had a missing admission and or discharge date, these were removed as the patient could not be tracked across multiple attendances and LOS could not be calculated. Diagnosis was excluded from this, since patients could be admitted to hospital and discharged with no formal diagnosis.

2. Patients were only included if they were aged 65 and over, in accordance with our elderly definition [264].

3. For the RadIS data set, patients were required to be admitted within hospital.

In total, 165,118 patients, having met the admission criteria, were included within the study. There were 15,483 scan records present from the admitted patients. Figure 5.1 provides an overview of the data cleansing process for both data sets.

Within the data there was a total of 24 different data headings Table 5.1 provides a brief definition of the column headings within the data. A full list of data items with data types and attributes can be found within Table C.1 of the appendix.

Data Item	Definition
Admission Date	The exact date on which a patient is formally admitted to a healthcare facility, marking the beginning of their stay for medical evaluation, treatment, or other necessary healthcare services.
Admission Method	The process or means by which patients are admitted to a healthcare facility. These are national codes as defined in [265].
Admission Source	The origin or specific location from which a patient comes to the healthcare facility. These are national codes as defined in [266]
Admission Time	The exact time of day when a patient is formally admitted to a healthcare facility.
Borough	The specific area within a region where a patient resides, providing geographical information that is relevant for healthcare planning, resource allocation, and demographic analysis.
Date of Birth	The date of birth of the patient.
Diagnosis	Categorisation of a patient's medical condition or illness, which is determined through medical examinations, tests, and evaluations carried out by healthcare professionals, enabling appropriate treatment and care planning.
Discharge Date	The specific date on which a patient is formally released or discharged from a healthcare facility, marking the end of their stay.
Discharge Destination	The specific location or facility to which a patient is transferred or sent upon being formally discharged from a healthcare facility.
Discharge Time	The exact time of day when a patient is formally discharged from a healthcare facility.
Hospital	The hospital in which a patient has been admitted to.
NHS Number	A unique number which enables healthcare staff and service providers to identify you correctly and match your details to your health records.
Postcode	The specific geographical area or location used to identify the patient's residential address.
Registered GP	The primary healthcare provider or family doctor with whom the patient is formally registered.
Registered GP Practice	The specific geographical area or location where the patient's primary healthcare provider or registered GP's practice is situated.
Scan Attendance Date	The specific date on which a patient undergoes a medical scan or diagnostic imaging procedure.
Scan Attendance Time	The specific time of day when a patient undergoes a medical scan or diagnostic imaging procedure.
Scan Exam	The specific type of diagnostic imaging procedure to be performed and its corresponding location.
Scan Exam Code	A unique code assigned to a specific type of diagnostic imaging procedure to be performed.
Scan Location Name	The specific name of the healthcare facility or imaging centre where a medical scan or diagnostic imaging procedure was conducted.
Scan Procedure Group	A categorisation or grouping of related diagnostic imaging procedures based on certain criteria, such as the body system or medical condition being investigated, the imaging technology used, or the purpose of the scan
Scan Requested Date	The specific date on which a medical scan or diagnostic imaging procedure is requested by a healthcare professional.
Scan Specialty Code	The referring consultant's specialty which has requested the scan.
Specialty	The type of ward a patient is admitted to, referring to the specific area of medicine of the patient's medical condition or illness. These are national codes as defined in [267].

Table 5.1: Definitions for each of the data items within the merged ABUHB data set.

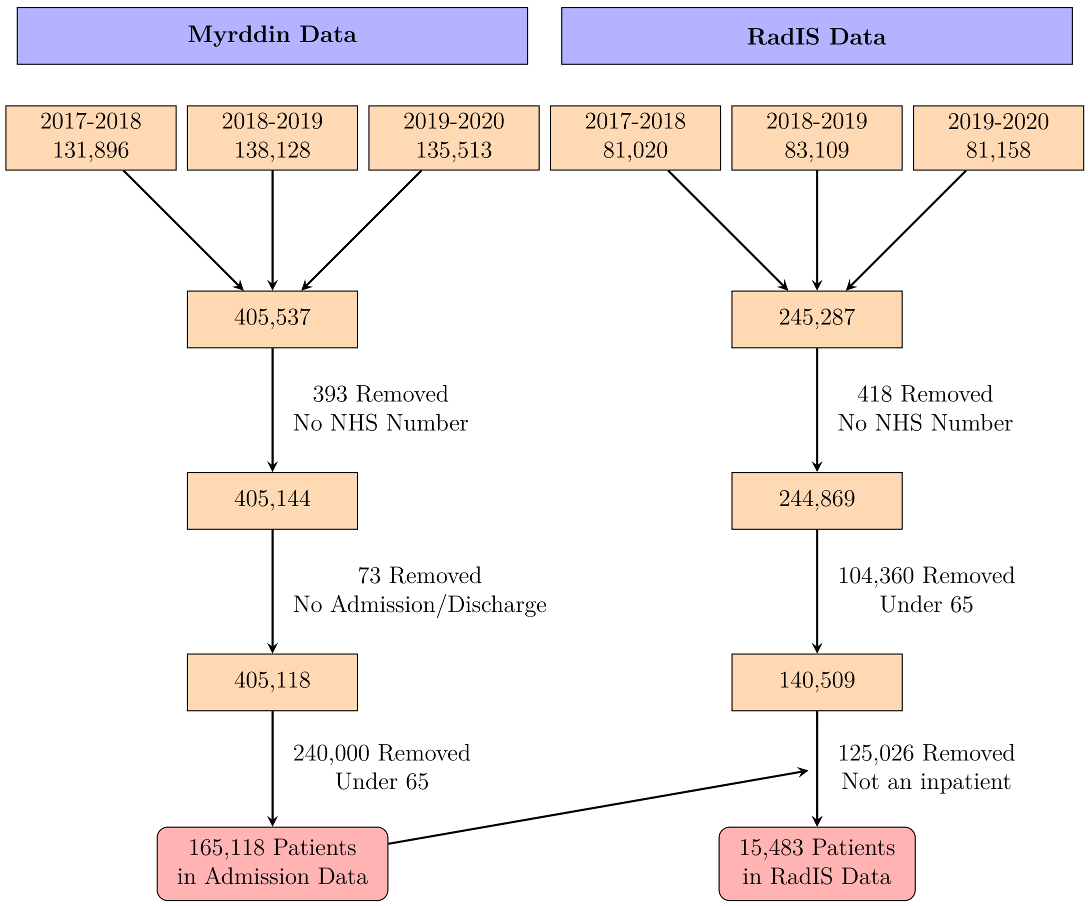


Figure 5.1: Flow chart of data cleansing process resulting in 165,118 patient admissions and 15,483 patient scans within ABUHB for the period April 2017 to March 2020.

5.2.1 Data Trends

Three years' worth of data was analysed to gain an understanding as to the demands faced by the elderly population within ABUHB. Figure 5.2 displays the daily count of admissions into ABUHB over this time period. The fluctuations within the data suggest seasonality is present as is often found within healthcare data [268]. Analysis on a year-to-year basis (running from April to March, also known as the fiscal year), showed the number of patients remained fairly consistent over the three years:

- 2017-2018: 53,256 (32.25%)
- 2018-2019: 56,050 (33.95%)
- 2019-2020: 55,812 (33.80%)

Due to the Covid-19 pandemic and the skewed effects it would have had on admission statistics, data from April 2020 was excluded from the study [269]. The effect of Covid-19 can be seen from March 2020, where admissions for patients aged 65 and older started to decrease.

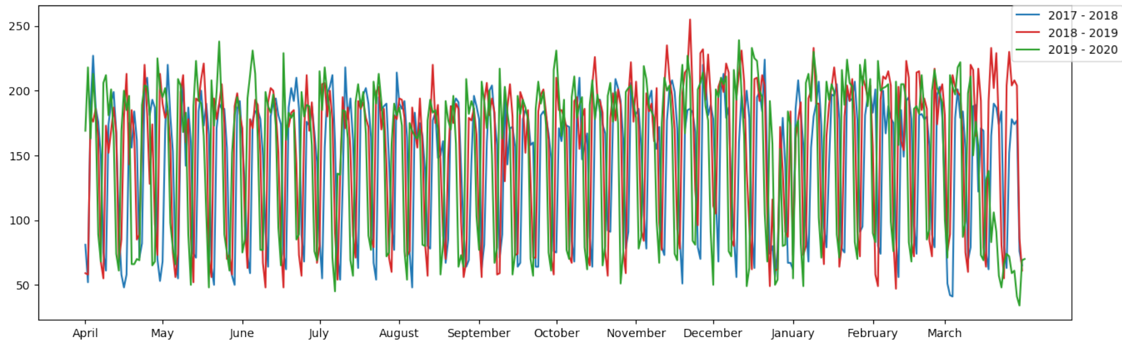


Figure 5.2: Line graph showing the trend of the number of patients admitted per day into ABUHB, split by fiscal year.

Within the data there were 66,289 unique NHS numbers, meaning 98,829 stays were either part of a care spell or independent admissions. The data did not give any indication regarding patient care spells, i.e., where they had been transferred between hospitals. Therefore, it was decided that if a patient had been discharged and subsequently readmitted on the same day, then this would fall into the 'care spell' category. In total, there were 8,826 care spell episodes. Table 5.2 displays the breakdown of the number of transfers.

No. of Transfers	**1**	**2**	**3**	**4**	**5**	**6**	**7**
Total	8,435	313	50	22	4	1	1

Table 5.2: The total number of readmissions/transfers for the 9,332 care spell episodes within the ABUHB data set. Number of patient transfers in ABUHB.

The patient who underwent seven transfers and readmissions, spent a total of 219 days in hospital, primarily moving between Royal Gwent Hospital (RGH) and County Hospital before being discharged to a non-NHS care home. This is also known as step-up and step-down care.

Within ABUHB there are 29 different specialties offered by the health board. The specialty relates to the ward a patient has been admitted to and is directly related to their treatment required. The most common specialty is general surgery with nearly 12% of admissions (Table 5.3). The top 11 specialties accounted for a total of 81.49% of admissions. In order to account for at least 95% of admissions, the top 15 specialties should be included.

The age of patients had a lower bound of 65, with the oldest being 107 years old. The mean age was calculated to be 77 years with a standard deviation of eight years. Due to the 42 year age range, patients were also grouped into five-year age categories to determine if this would be a more accurate predictor (Figure 5.3).

The highest frequency of admissions occurred on a Tuesday (18.04%), followed by

Specialty	Count	Proportion
General Surgery	19,782	11.98%
Care of the Elderly	18,797	11.38%
Gastroenterology	16,499	9.99%
Trauma & Orthopaedic	16,471	9.98%
General Medicine	12,913	7.82%
Urology	12,574	7.62%
Ophthalmology	11,557	6.99%
Dermatology	9,466	5.73%
Rehabilitation	8,466	5.12%
Respiratory	8,039	4.86%

Table 5.3: The top 11 admitting specialties by count within the ABUHB data set.

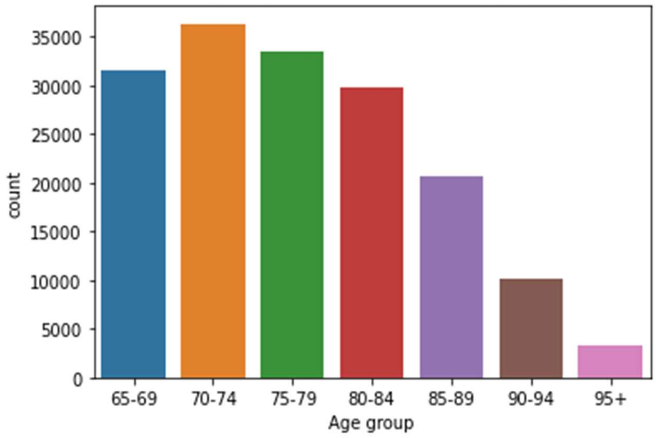


Figure 5.3: Bar chart showing the number of patients within the ABUHB data set, falling into each of the seven age groups.

a Thursday (17.84%) (Figure 5.4). Patients who were admitted over a weekend accounted for 14.08% of total admissions. Weekday admissions peaked between 7am and 10am, with a secondary peak between 12pm and 1pm.

The admission source of a patient determined where a patient was directly prior to admission. Although there are 25 different admission sources listed in total, the top six admission sources accounted for 98.63% of admissions (Table 5.4). If this were to be extended to include the top 10, 99.88% of admissions would be accounted for. The top two admission sources, ‘Usual Place of Residence’ and ‘Own Home’ have a combined patient count of 147,294 (89.21%).

The admission method followed a similar trend to the admission source, with a small number of methods accounting for the majority of patients. Although there were 16 distinct methods, 99.23% of patients were accounted for through the top seven (Table 5.5). ‘Elective - waiting list’ was the most common admission method and is one that had been arranged in advance of admission. The patient had been admitted via a waiting list, where at the point of being put onto the waiting list,

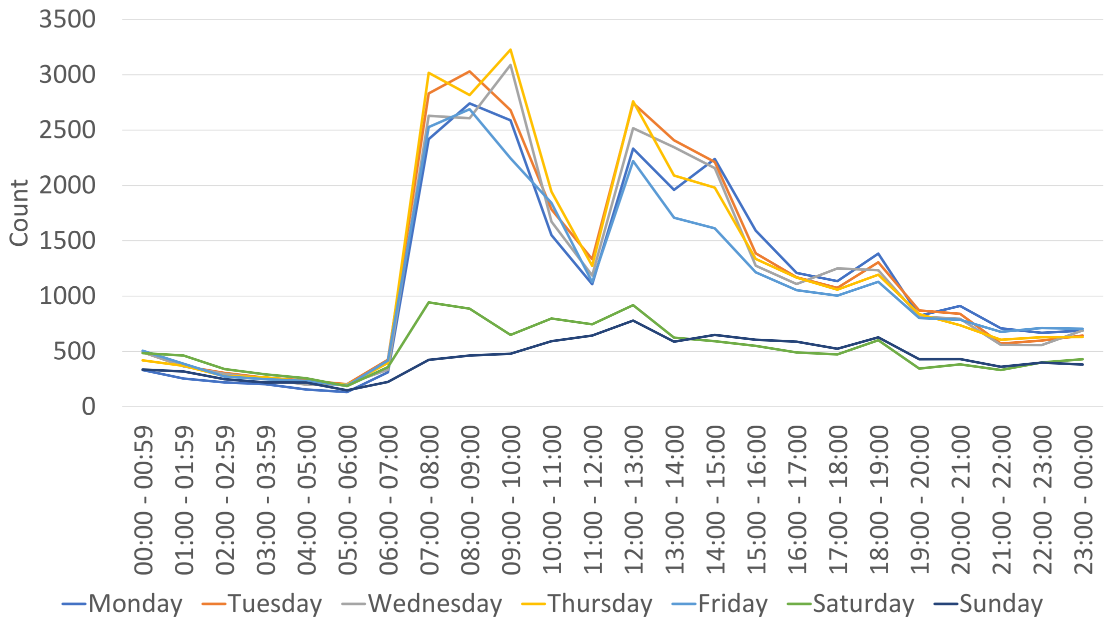


Figure 5.4: Line graph showing the number of admissions per day by hourly interval, for the ABUHB data set.

Admission Source	**Count**	**Proportion**
Usual Place of Residence	99,464	60.24%
Own Home	47,830	28.97%
Same Trust-General or young phys.disabled	7,711	4.67%
Patient transfer within the same health board/trust	4,350	2.63%
Non NHS (other than L.A.) run res.care home	1,755	1.06%
Non NHS (other than L.A.) run nursing home	1,314	0.80%

Table 5.4: The top six admitting methods by count within the ABUHB data set.

did not know their date of admission. The period in which the patient had to wait was dependant on the demand for hospital resources and facilities.

Admission Method	**Count**	**Proportion**
Elective - waiting list	73,482	44.50%
Emergency - casualty	39,437	23.88%
Emergency - GP	28,169	17.06%
Other - transferred from another hospital	13,460	8.15%
Elective - booked	5,950	3.60%
Elective - planned	1,696	1.02%
Emergency - other means	1,658	1.04%

Table 5.5: The top seven admission methods by count within the ABUHB data set.

The LOS of the patient was determined by the time between admission and hospital discharge. There was a large range of LOS's, from 0 to 413 days. The LOS can be modelled in two ways, in hours or in days. The first method of calculating LOS, used the number of hours that patients had been admitted for. It was determined using the admission and discharge dates as well as specific times. The average LOS was calculated to be 155.23 hours (6.47 days). The second method of calculating LOS, meant that if a patient was admitted overnight, an additional day was added to their LOS i.e., if admitted Monday evening and discharged Tuesday morning, their LOS was one. Additional analysis of LOS in hours revealed the mean to be 6.47 days and a 75^{th} percentile of seven days. The 90^{th} percentile was 18 days increasing to 30 days with the 95^{th} percentile. This implies that some long LOS's are skewing the mean. Patients were released from the hospital in 81,538 (49.38%) cases within 24 hours and 114,015 (69.05%) cases within five days. A LOS longer than 30 days was present in 5% of patients.

A patient's LOS changed depending on the day they were admitted (Table 5.6). Again, patients who were admitted on a weekend have a different variation to those patients admitted during the week. If hospitalised on a weekend, the average LOS was at least an additional day longer.

Day of week	**Mean LOS**	**Standard Deviation**
Monday	6.22	12.77
Tuesday	6.12	13.54
Wednesday	6.11	13.25
Thursday	5.97	13.02
Friday	6.78	12.97
Saturday	7.88	13.78
Sunday	8.10	12.97

Table 5.6: The mean and standard deviation of patient LOS in hospital by the admission day.

As discussed in Section 1.1, two approaches, each employing ICD10 codes, were used to determine the frailty score. Regarding the patient's diagnosis and the cause for admission, there were 2,758 unique codes in total. The average frailty score was 0.5, with a standard deviation of 0.98. The maximum score was 8.1, with the lowest score being zero.

Depending on their health and course of treatment, patients who were admitted to the hospital may undergo a number of scans. In total, 12,350 patient scans were recorded, of which 9,863 patients had just one scan (Table 5.7). There were 2,487 patients who had at least two scans, of which 1,957 patients only had two scans. One patient had a total of eight scans.

Scan Number	**1**	**2**	**3**	**4**	**5**	**6**	**8**
Count	9,863	1,957	437	77	11	4	1

Table 5.7: The total number of scans for patients who had a scan during their inpatient admission within the ABUHB data set.

The data set comprised of a diverse range of scans, encompassing various modalities, including MRI and ultrasound. Figure 5.5 displays the frequency of each of the procedure codes, showing that X-ray (R) is the most common scan type, with 63.76% of all scans. For patients who have multiple scans, there are duplications of the same scan in the same region, however, this can be due to different angles required to be taken. Additionally there are different scan types on the same body region as well as different scans on different body regions.

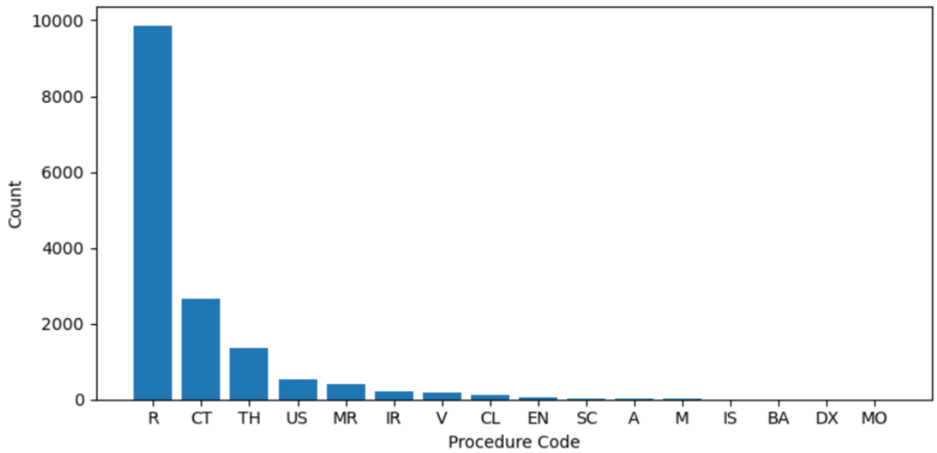


Figure 5.5: Bar chart showing the number of patients undertaking each type of scan within the ABUHB data set.

5.2.2 Data Insights

The information from the merged data sets was comprised of 165,118 patient records over a three year period, with admissions remaining constant throughout this time. There were more arrivals from South East Wales' more populated towns and cities, like Caerphilly and Newport, than from the region's more rural parts. The acute hospitals with 24/7 services in the area, which are similarly situated in Caerphilly and Newport, tended to have greater patient attendance rates. Peak admissions occurred on a Tuesday, with admissions generally being greater during the week. For all arrivals, there were typically peaks in the late morning, early afternoon, and a smaller peak in the evening. The patient demographics were varied amongst all hospitals, however, the LOS varied by age and hospital attended. The trends indicated that shorter LOS were more typical in acute hospitals while longer LOS were more prevalent in community-based hospitals. The patients ages range widely, although the majority are between 65 and 85. The frailty scores of the patients provide some indication of the severity of the illness. Higher frailty scores were correlated with longer hospital stays and older patient populations in some hospitals. Among the admitted patients, there was a small subset who required at least one scan.

5.3 Predictive Analytics Results

This section will examine how the patient data from ABUHB can be used to apply the notion of predictive analytics covered in Chapter 3 in order to ascertain how clinical and demographic factors affect hospital LOS.

5.3.1 Linear Regression

Linear regression was performed on the 14 variables to determine the highest influence on LOS. The LOS was converted into how many nights the patient spent in hospital, for example, if admitted on 1^{st} January 2020 and discharged on the 3^{rd} January 2020, the LOS would be two. This produced a higher R^2 value for linear regression for all cases than using the continuous hourly LOS. The R^2 and adjusted R^2 values for each variable against this LOS are shown in Table 5.8.

Continuous Variables	**R^2 Value**	**Adjusted R^2 Value**
Age	0.050	0.050
Frailty Score	0.028	0.028
No. of Scans	0.003	0.003
Categorical Variables	**R^2 Value**	**Adjusted R^2 Value**
Age Group	0.051	0.051
Admission Method	0.282	0.282
Admission Source	0.195	0.195
Day of Admission	0.002	0.002
Diagnosis	0.273	0.261
Frailty Group	0.028	0.028
Hospital	0.182	0.181
ICD10 - First Letter	0.092	0.092
Scan Y/N	0.002	0.002
Month of Admission	0.000	0.000
Specialty	0.288	0.288

Table 5.8: The results for the linear regression when run against the continuous and categorical variables within the ABUHB data set. The R^2 and Adjusted R^2 values are given.

The variables can be considered as two different data types, continuous and categorical. Within the continuous variables, age produced the highest R^2 value of 0.05. This means 5% of the LOS variation is explained by age. The model can be denoted by:

$$Y = 0.375x - 22.635 \tag{5.1}$$

where x is the age of the patient. Therefore, for each one year increment in age, the LOS will increase by 0.375 days. Table C.2 within the Appendix displays the equations for the three continuous variables.

Similarly, we can calculate the linear regression model for categorical variables within Table 5.8. The month of admission produced an R^2 value of zero, which indicates that the month does not account for any variation in the response data around its mean. Specialty provided the largest R^2 value of 0.288. There are 29 subcategories of different specialties within the specialty category. We can further analyse each variable to be able to forecast LOS. The corresponding linear regression values for specialty are shown in Table 5.9. When one x variable is selected, the corresponding value is the LOS. For instance, if the specialty of A&E is chosen, the LOS will be 2.2673 days. This can be directly compared to other disciplines; for instance, the LOS in anaesthetics is 6.5 times greater than that in A&E. The value column displays the variation in the average LOS between each of the specialties. For all categorical variables, excluding diagnosis, the coefficients can be seen in Table C.2 within the Appendix. Diagnosis was excluded due to the large number of variables.

Specialty Type	x Variable	Value
Accident & Emergency (A&E)	x_1	2.2673
Anaesthetics	x_2	14.9517
Cardiology	x_3	4.6470
Care of the Elderly	x_4	12.1101
Community Medicine	x_5	34.235
Dermatology	x_6	0.2616
Diabetes & Endocrinology	x_7	11.6161
Ear, Nose & Throat	x_8	2.7500
GP Other	x_9	39.2603
Gastroenterology	x_{10}	2.1382
General Medicine	x_{11}	8.4519
General Surgery	x_{12}	3.7149
Gynaecology	x_{13}	1.6536
Haematology	x_{14}	0.7974
Infectious Diseases	x_{15}	11.6289
Intermediate Care	x_{16}	14.3725
Maxillo-Facial	x_{17}	0.6018
Neurology	x_{18}	5.6131
Ophthalmology	x_{19}	0.1307
Pain	x_{20}	0.0080
Plastic Surgery	x_{21}	0.1128
Radiology	x_{22}	0.3548
Radiotherapy & Oncology	x_{23}	13.6667
Rehabilitation	x_{24}	28.7732
Respiratory	x_{25}	7.7985
Restorative Dentistry	x_{26}	0.0000
Rheumatology	x_{27}	2.3333
Trauma & Orthopaedic	x_{28}	6.6658
Urology	x_{29}	0.9932

Table 5.9: The linear regression results for the specialty variable, where each x has a corresponding LOS value.

5.3.2 Logistic Regression

Logistic regression was performed to determine the effect of grouping LOS. Grouped LOS categories were determined by grouping patients into whether they were discharged on the same day as arrival or admitted overnight. This was investigated as it was a particular interest of managers within ABUHB, as they wanted to determine the characteristics of patients who should be discharged on the same day but ultimately required overnight admission. This resulted in 75,216 patients falling into the '0' category (45.55%), where they were discharged the same day and 89,902 patients who fell into the '1' category (54.45%) where their LOS was at least two days. This is beneficial for bed and staff planning to determine the turnaround time of patients.

Table 5.10 displays the four scoring measures against each of the variables. In three cases, precision is given the value 'N/A', this is due to both the TP and FP rates being zero (i.e., the model only predicted negative results). Therefore, a result cannot be calculated. Similarly, because the precision cannot be calculated, then the F1 score cannot be calculated.

Continuous Variables	**Accuracy**	**Precision**	**Recall**	**F1 Score**
Age	0.6037	0.5721	0.5164	0.5428
Frailty Score	0.5860	0.5283	0.8503	0.6517
No. of Scans	0.5445	N/A	0.0	N/A
Categorical Variables	**Accuracy**	**Precision**	**Recall**	**F1 Score**
Admission Method	0.8764	0.8378	0.9036	0.8694
Admission Source	0.5829	1.0	0.0844	0.1557
Age Group	0.6037	0.5721	0.5164	0.5428
Day of Admission	0.5474	0.5085	0.1991	0.2862
Diagnosis	0.8496	0.8360	0.8607	0.8481
Frailty Group	0.5871	0.5291	0.8498	0.6522
Hospital	0.6171	0.6923	0.2871	0.4059
ICD10 - First Letter	0.7554	0.7813	0.6430	0.7055
Scan Y/N	0.4555	N/A	0.0	N/A
Month of Admission	0.4555	N/A	0.0	N/A
Specialty	0.8008	0.7645	0.8131	0.7881

Table 5.10: The results for the logistic regression when run against the continuous and categorical variables within the ABUHB data set. The accuracy, precision, recall and F1 scores are given.

We can determine the likelihood of falling into one of the two LOS groupings using the logit function. Equation (5.2) displays the general formula using age, a continuous variable, as an example.

$$logit(\pi(x)) = \ln\left[\frac{\pi(x)}{(1-\pi(x))}\right] = -5.0746 + (0.0681 \times Age) \tag{5.2}$$

If the age is set to be 80, then the conditional logit of being admitted overnight is:

$$\ln\left[\frac{\pi(x)}{(1-\pi(x))}\right](Age = 80) = -5.0746 + (0.0681 \times 80) \tag{5.3}$$

Then the effect of a one-unit increase in age can be examined. When the age of a patient is 81, the following is calculated:

$$\ln\left[\frac{\pi(x)}{(1-\pi(x))}\right](Age = 81) = -5.0746 + (0.0681 \times 81) \tag{5.4}$$

Taking the difference of the two equations, we are left with the following result:

$$\ln\left[\frac{\pi(x)}{(1-\pi(x))}\right](Age=81) - \ln\left[\frac{\pi(x)}{(1-\pi(x))}\right](Age=80) = 0.1381 \tag{5.5}$$

Therefore, the coefficient for age is the difference in the log odds, and as such, for one unit increase in age, the expected change in log odds is 0.1381. Exponentiating both sides, results in a value of 1.1481:

$$e^{\ln\left[\frac{\pi(x)}{(1-\pi(x))}\right](Age=81)-\ln\left[\frac{\pi(x)}{(1-\pi(x)}\right](Age=80)} = e^{0.1381} = 1.1481 \tag{5.6}$$

Therefore, we can say for one-unit increase in age, there is a 14.81% increase in the likelihood of being admitted overnight. The 14.81% of increase is not dependent on the value age is held at. For the $logit(\pi(x))$ equations for all three continuous variables, see Table C.4 in the Appendix.

Similarly, we can calculate the log odds for a categorical variable, e.g., age group. Table 5.11 describes the relationship between the LOS group and age group.

Age Group	**Log Odds Ratio**	**Odds compared to 65-69**
Intercept	-0.3582	-
70-74	0.1263	13.46%
75-79	0.4389	55.10%
80-84	0.7609	114.02%
85-89	1.2366	244.39%
90-94	1.8219	518.36%
95+	2.2298	829.80%

Table 5.11: The log odds ratios for age group category after running the logistic regression model. The odds compared to the 65-69 age group are also provided.

The intercept is also known as the reference category, which in this instance is the age group '65-69'. If we compare the reference group to the '70-79' category and perform, $e^{0.1263}$, a result of 1.1346 is produced. This shows that patients in the '70-74' age group have a 13.46% higher chance of being admitted overnight. There is an 829.80% higher chance of 95+ year old patient being admitted compared to those aged between 65 and 69. Table C.5 in the Appendix contains the log odds ratios for all categorical variables with the exception of diagnosis due to the large number of variables.

5.3.3 Classification and Regression Trees

The variables analysed within subsections 5.3.1 and 5.3.2 can then be inputted into CART models to predict patients LOS within hospitals. Diagnosis will be used instead of the 'ICD10 - first letter' since both linear and logistic regression results produced a higher result. Similarly, the number of scans rather than whether a person had a scan (Scan Y/N), will be used. Age and age group will be investigated for their impact on the CART model. To determine the effect of using a frailty measure, continuous and grouped frailty will be compared against not using a frailty score within the model.

5.3.3.1 Regression Trees

This section will look at the development and results of the regression trees, analysing continuous LOS. A total of nine variables will be used within the model, listed as follows:

- Admission Method
- Admission Source
- Age (Continuous and Grouped)
- Day
- Diagnosis
- Frailty (None, Continuous and Grouped)
- Hospital
- Number of Scans
- Specialty

Month was excluded from the regression model since in the linear regression model the R^2 was calculated to be zero and therefore did not account for any of the variability in patients LOS.

A 20% test set was used, meaning the data is trained on 80% of the data. Using the Python algorithm discussed in Section 3.4.3, Table 5.12 displays the parameters inputted into the model. The parameters 'min_samples_leaf' and 'max_leaf_nodes' will undergo testing to determine the trade off between R^2 score and computation time.

Parameters	**DecisionTreeRegressor**
criterion	"squared_error"
splitter	"best"
max_depth	None
min_samples_split	2
min_weight_fraction_leaf	0
max_features	None
random_state	None
min_impurity_decrease	0
ccp_alpha	0

Table 5.12: The parameters used within the regression tree model using the 'DecisionTreeRegressor' algorithm within Python.

Tables 5.13a - 5.18b display the R^2 score and computation time for a range of 'max_leaf_nodes' and 'min_samples_leaf' variables. The variable 'max_leaf_nodes' was evaluated on a range from five to 30 leaf nodes. Larger values were not selected to ensure a usable number of groups were identified. The variable 'min_samples_leaf', was investigated from one sample to 500 samples. By having a minimum number of samples per leaf, in theory, will reduce the likelihood of overfitting. Computation time was also collected to determine if there was a trade off between R^2 and time to run the model.

The highest R^2 score of 34.28% was attained in the regression tree using grouped age and continuous frailty (Table 5.17a). This was achieved using 100 minimum samples per leaf and 30 maximum leaf nodes. However, in comparison to other models, it produced a longer computation run time of 26.9786 seconds (Table 5.17b). If analysing the grouped age and continuous frailty models, by reducing R^2 by 0.05%, approximately 6.9 seconds can be saved by increasing the number of minimum samples per leaf to 200.

Within Table 5.18a, 30 maximum leaf nodes and 100 minimum samples per leaf also produced an R^2 score of 34.23%, with a computational time of 18.5107 seconds. This combination was selected as the optimum and will be used going forward as it produces a large R^2 score with low computation time.

The R^2 scores all range between 29.25% to 34.28%, due to the large range of LOS's within the data. The range of LOS was between zero days and 417 days, and an R^2 score of 34%, shows 34% of the time we are able to correctly assign patients to the correct node, despite this large range. These leaf nodes will be able to be used to group patients accordingly to LOS.

max_leaf_nodes	min_samples_leaf: 1	100	200	300	400	500
5	0.2925	0.2925	0.2925	0.2925	0.2925	0.2925
10	0.3224	0.3224	0.3224	0.3190	0.3190	0.3190
15	0.3328	0.3328	0.3328	0.3322	0.3322	0.3250
20	0.3362	0.3384	0.3384	0.3374	0.3374	0.3286
25	0.3377	0.3408	0.3408	0.3393	0.3392	0.3301
30	0.3375	0.3416	0.3410	0.3395	0.3407	0.3304

(a) R^2 score.

max_leaf_nodes	min_samples_leaf: 1	100	200	300	400	500
5	9.5668	9.8238	9.7276	8.5625	8.5016	8.8017
10	13.9514	12.0725	11.9836	11.9235	13.9066	12.6438
15	14.4875	14.2050	16.7100	14.2829	14.5327	14.8251
20	16.7751	17.9756	16.5835	15.8724	15.6849	14.9515
25	17.8356	18.2084	16.8472	16.8951	17.8183	18.5711
30	19.5540	19.1484	19.1552	18.5376	19.4121	19.6245

(b) Computational time in seconds (s)..

Table 5.13: The regression tree results for the R^2 (a) and computational time (b) for continuous age and no frailty.

max_leaf_nodes	min_samples_leaf: 1	100	200	300	400	500
5	0.2925	0.2925	0.2925	0.2925	0.2925	0.2925
10	0.3224	0.3224	0.3224	0.3190	0.3190	0.3190
15	0.3328	0.3328	0.3328	0.3322	0.3322	0.3250
20	0.3362	0.3381	0.3381	0.3381	0.3381	0.3294
25	0.3389	0.3412	0.3412	0.3391	0.3396	0.3308
30	0.3383	0.3420	0.3422	0.3393	0.3410	0.3308

(a) R^2 score.

max_leaf_nodes	min_samples_leaf: 1	100	200	300	400	500
5	8.9660	9.1705	8.7471	8.4058	8.7663	8.4657
10	12.9853	12.1247	11.7582	11.6642	15.0279	12.1618
15	13.4786	13.3782	16.3377	14.3459	15.3793	17.0997
20	16.8013	18.5770	17.1018	15.8003	16.1066	20.7639
25	18.3142	17.3414	17.3936	17.6433	25.1619	17.0871
30	19.5150	18.2041	19.6589	18.7183	20.5239	19.4184

(b) Computational time in seconds (s).

Table 5.14: The regression tree results for the R^2 (a) and computational time (b) for continuous age and continuous frailty.

max_leaf_nodes	min_samples_leaf: 1	100	200	300	400	500
5	0.2925	0.2925	0.2925	0.2925	0.2925	0.2925
10	0.3224	0.3224	0.3224	0.3190	0.3190	0.3190
15	0.3328	0.3328	0.3328	0.3322	0.3322	0.3250
20	0.3362	0.3384	0.3384	0.3379	0.3379	0.3291
25	0.3387	0.3409	0.3409	0.3394	0.3393	0.3305
30	0.3380	0.3416	0.3421	0.3395	0.3408	0.3306

(a) R^2 score.

max_leaf_nodes	min_samples_leaf: 1	100	200	300	400	500
5	9.5461	13.3577	10.0347	14.3069	15.3515	10.3998
10	12.8929	17.2952	13.9642	13.3573	19.6081	17.3906
15	15.0100	20.8186	15.0045	14.3649	17.6949	19.1605
20	18.6414	17.8904	17.3192	18.0861	18.5902	15.8897
25	24.3819	19.3318	18.1902	19.8152	19.1655	17.1851
30	30.6406	24.9834	19.9453	27.4008	24.1919	19.1556

(b) Computational time in seconds (s).

Table 5.15: The regression tree results for the R^2 (a) and computational time (b) for continuous age and grouped frailty.

max_leaf_nodes	min_samples_leaf: 1	100	200	300	400	500
5	0.2925	0.2925	0.2925	0.2925	0.2925	0.2925
10	0.3208	0.3208	0.3208	0.3194	0.3194	0.3194
15	0.3320	0.3208	0.3320	0.3320	0.3320	0.3236
20	0.3347	0.3379	0.3379	0.3369	0.3369	0.3292
25	0.3360	0.3407	0.3407	0.3392	0.3388	0.3297
30	0.3364	0.3418	0.3420	0.3403	0.3400	0.3309

(a) R^2 score.

max_leaf_nodes	min_samples_leaf: 1	100	200	300	400	500
5	11.6459	10.0844	12.8503	11.3010	12.5443	9.3685
10	14.0654	17.3290	12.4793	17.8037	17.7200	14.2252
15	15.5847	19.5658	13.5358	19.3913	15.1067	18.1730
20	22.4814	27.2284	20.0218	24.0735	19.1546	19.2974
25	20.0251	22.6745	23.6449	23.4709	24.0338	22.2256
30	23.8315	24.2733	23.7499	26.3434	20.2900	23.4050

(b) Computational time in seconds (s).

Table 5.16: The regression tree results for the R^2 (a) and computational time (b) for grouped age and no frailty.

max_leaf_nodes	min_samples_leaf: 1	100	200	300	400	500
5	0.2925	0.2925	0.2925	0.2925	0.2925	0.2925
10	0.3208	0.3208	0.3208	0.3194	0.3194	0.3194
15	0.3320	0.3208	0.3320	0.3320	0.3320	0.3236
20	0.3286	0.3384	0.3384	0.3377	0.3377	0.3294
25	0.3321	0.3417	0.3417	0.3403	0.3397	0.3309
30	0.3314	0.3428	0.3423	0.3402	0.3409	0.3322

(a) R^2 score.

max_leaf_nodes	min_samples_leaf: 1	100	200	300	400	500
5	10.3076	13.7601	13.3941	10.3015	10.1157	10.4414
10	18.1097	18.2357	20.2272	21.7858	12.6390	13.9807
15	21.0192	23.0711	22.7380	18.9120	17.9297	23.8506
20	21.2881	28.4333	23.5638	18.9010	20.4316	25.3724
25	22.5401	25.8047	25.5739	22.0894	20.1581	23.0146
30	28.5686	26.9786	20.1018	20.8679	22.5167	25.4133

(b) Computational time in seconds (s).

Table 5.17: The regression tree results for the R^2 (a) and computational time (b) for grouped age and continuous frailty.

max_leaf_nodes	min_samples_leaf: 1	100	200	300	400	500
5	0.2925	0.2925	0.2925	0.2925	0.2925	0.2925
10	0.3208	0.3208	0.3208	0.3194	0.3194	0.3194
15	0.3320	0.3208	0.3320	0.3320	0.3320	0.3236
20	0.3342	0.3381	0.3381	0.3374	0.3374	0.3291
25	0.3318	0.3414	0.3414	0.3399	0.3393	0.3306
30	0.3311	0.3423	0.3421	0.3401	0.3405	0.3313

(a) R^2 score.

max_leaf_nodes	min_samples_leaf: 1	100	200	300	400	500
5	11.4307	16.4057	12.4373	15.8219	15.7179	9.3838
10	13.6344	18.4423	24.6463	12.5033	19.6957	18.3607
15	16.6546	19.4606	20.0820	13.6390	19.6694	15.4206
20	18.2777	22.8429	23.1844	16.9840	25.4143	16.9038
25	20.9142	19.2554	24.9062	18.5972	20.5371	17.8328
30	25.1081	18.5107	27.9524	26.3307	20.3707	23.5998

(b) Computational time in seconds (s).

Table 5.18: The regression tree results for the R^2 (a) and computational time (b) for grouped age and grouped frailty.

Figure 5.6 displays the regression tree visualisation with R^2 of 34.23%. The results display the most important determination of LOS is 'admission_method_Other - transferred from another hospital'. Due to the one-hot encoding of variables, if this value is zero, the patient had a different admission method and the reader would move to the left hand side of the tree. The 'value' on the nodes denotes the predicted LOS in days.

The model produced a total of 30 leaf nodes and therefore contains 30 groupings of patients with different LOS's. The LOS's are given in days, with the number of samples that fall into this category. The average demand can then be calculated using Equation (5.7):

$$\text{Average daily bed demand} = \frac{\text{Average LOS} \times \text{count}}{\text{Total number of days}} \tag{5.7}$$

For example, if we take the patients who have the admission method of being transferred from another hospital and are admitted to NHH. To determine how a patient's LOS varies, the specialty rehabilitation results in a difference of 17.835 days. Using Equation (5.7), an average of 6.2 beds should be available per day. For the remaining specialties offered by NHH, 5.94 beds should be planned in total per day. This highlights the small but important difference in bed numbers. It should be noted that additional beds will be added to NHH, through the left hand side of the tree. However, adding these individual nodes together will create a more precise picture of the actual demands faced by the health board.

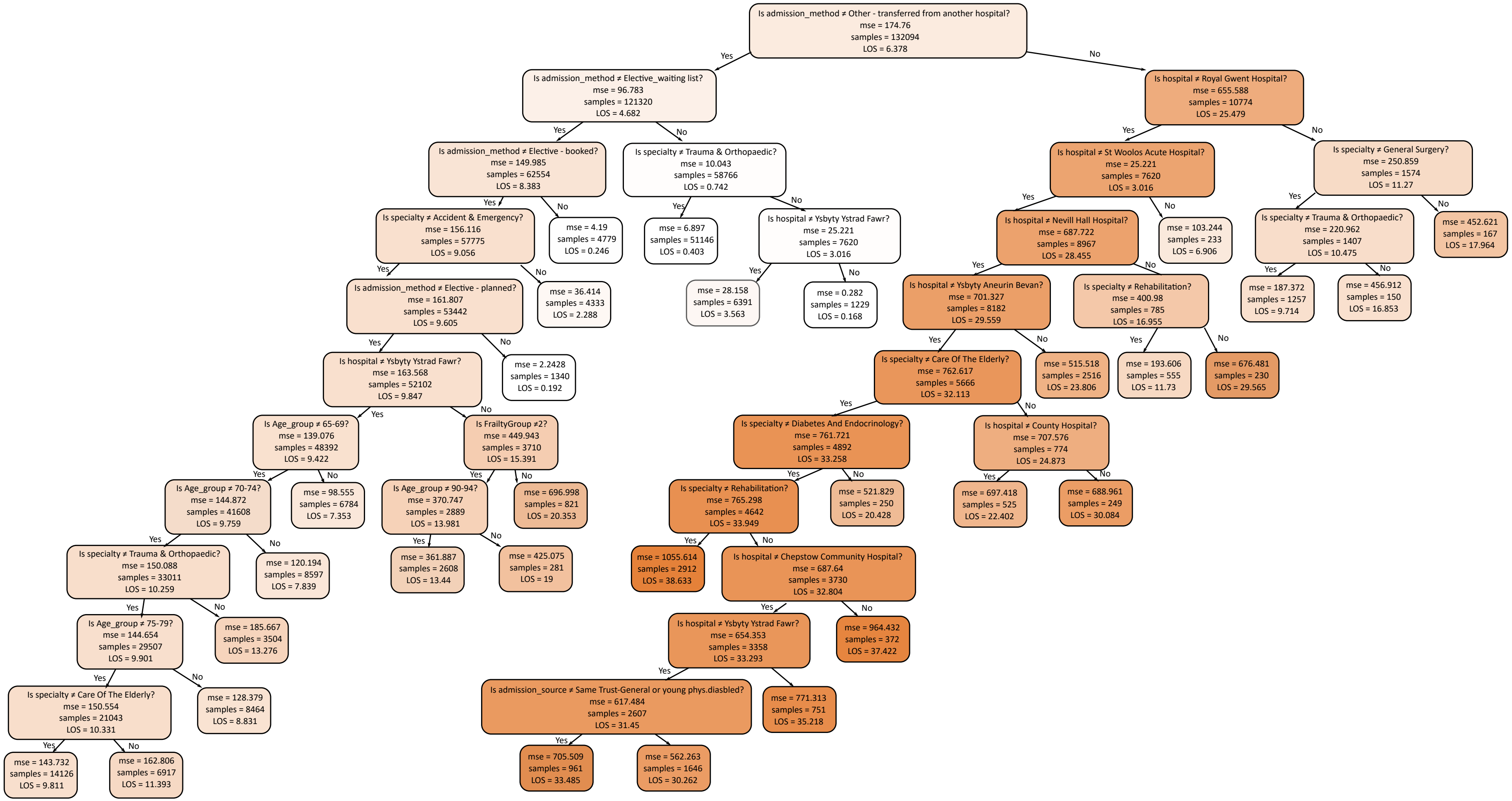


Figure 5.6: Regression tree with an R^2 of 34.23% for predicting continuous LOS for patients within ABUHB, consisting of 30 terminal nodes.

5.3.3.2 Classification Trees

This section discusses the development and results of the classification trees, analysing grouped LOS into patients who are discharged on the same day as arrival and those who are admitted overnight. A total of 10 variables will be used within the model, listed as follows:

- Admission Method
- Admission Source
- Age (Continuous and Grouped)
- Day
- Diagnosis
- Frailty (None, Continuous and Grouped)
- Hospital
- Month
- Number of Scans
- Specialty

Unlike regression trees, the month was included within the model as the accuracy score produced was greater than zero and could provide some benefit in being used within the model.

Again, an 80% training set and a 20% test set were used to build and develop the model. Using the 'DecisionTreeClassifier' algorithm within Python as discussed in Section 3.4.4, the following parameters were selected for the model (Table 5.19). The parameters 'min_samples_leaf' and 'max_leaf_nodes' will be investigated to determine the combination which yields the highest accuracy, precision and recall scores against the computational time.

Parameters	**DecisionTreeClassifier**
criterion	"gini"
splitter	"best"
max_depth	None
min_samples_split	2
min_weight_fraction_leaf	0
max_features	None
random_state	None
min_impurity_decrease	0
class_weight	None
ccp_alpha	0

Table 5.19: The parameters used within the classification tree using the 'DecisionTreeClassifier' algorithm within Python.

The accuracy scores for each of the six experiments range from 88.46% to 89.89%, a difference between 1.43% (Tables 5.20a - 5.25d). This shows regardless of frailty and age classification, the difference is minimal between a range of different minimum samples per leaf and the maximum number of leaf nodes. The precision has a range of 4.41%, from 84.75% to 89.16%. Comparing the accuracy to precision result, the highest values for each do not occur in the same 'min_samples_leaf' and 'max_leaf_nodes' combination. For example, comparing Table 5.20a with Table 5.20b, accuracy yielded the highest result with one minimum sample per leaf and 30 maximum leaf nodes. The highest precision results were recorded with 15 maximum leaf nodes and either 400 or 500 minimum samples per leaf. The recall scores produced the largest result in the same combination as the accuracy scores (largest maximum leaf nodes and smallest minimum samples per leaf), however, it produced the lowest score in the same location as the precision. Comparing the 400 minimum samples per leaf and 15 maximum leaf nodes suggests that FN > FP. In terms of healthcare, this means patients are predicted to have short LOS's whereas, in reality, they have longer LOS's. Therefore, even though more costly as extra resources will be planned, it is beneficial to plan beds and not require them. This means the optimal solution should prioritise the recall score over the precision.

The optimum combination was selected to be one minimum sample per leaf with 30 maximum leaf nodes, using continuous age and continuous frailty (Tables 5.21a - 5.21d). This produced the highest accuracy and recall score, whilst also producing the lowest computational time out of the combination.

max_leaf_nodes	min_samples_leaf 1	100	200	300	400	500
5	0.8846	0.8846	0.8846	0.8846	0.8846	0.8846
10	0.8913	0.8913	0.8913	0.8913	0.8913	0.8913
15	0.8950	0.8950	0.8950	0.8950	0.8925	0.8925
20	0.8956	0.8956	0.8950	0.8950	0.8936	0.8936
25	0.8977	0.8976	0.8962	0.8963	0.8936	0.8939
30	0.8989	0.8987	0.8968	0.8968	0.8939	0.8937

(a) Accuracy score.

max_leaf_nodes	min_samples_leaf 1	100	200	300	400	500
5	0.8475	0.8475	0.8475	0.8475	0.8475	0.8475
10	0.8645	0.8645	0.8645	0.8645	0.8645	0.8645
15	0.8643	0.8643	0.8643	0.8643	0.8916	0.8916
20	0.8641	0.8641	0.8643	0.8643	0.8844	0.8844
25	0.8637	0.8617	0.8598	0.8594	0.8844	0.8804
30	0.8570	0.8617	0.8611	0.8635	0.8804	0.8731

(b) Precision score.

max_leaf_nodes	min_samples_leaf 1	100	200	300	400	500
5	0.8944	0.8944	0.8944	0.8944	0.8944	0.8944
10	0.8941	0.8941	0.8941	0.8941	0.8941	0.8941
15	0.8941	0.9019	0.9019	0.9019	0.8755	0.8755
20	0.9033	0.9033	0.9019	0.9019	0.8829	0.8829
25	0.9080	0.9094	0.9081	0.9088	0.8829	0.8865
30	0.9163	0.9117	0.9083	0.9061	0.8865	0.8919

(c) Recall score.

max_leaf_nodes	min_samples_leaf 1	100	200	300	400	500
5	9.1109	8.3105	8.2769	8.7439	13.1439	10.0509
10	11.5346	11.6610	11.6360	12.0482	18.4817	13.7823
15	14.8981	13.3688	14.2424	14.0589	16.0018	17.7209
20	17.1809	16.0233	16.2691	21.6318	18.7793	19.5191
25	17.3070	17.6209	17.1829	22.8581	21.2422	20.4255
30	18.0397	17.7756	18.9454	21.1663	21.4531	22.0665

(d) Computational time in seconds (s).

Table 5.20: The classification tree results for accuracy score (a), precision score (b), recall score (c) and computational time (d) for continuous age and no frailty.

Figure 5.7 displays the classification tree visualisation with an accuracy score of 89.89%, a precision score of 85.70% and a recall score of 91.63%. The tree shows the most important factor to determine whether a patient will be admitted overnight is the 'admission_method_elective - waiting list'. If a patient was admitted via this method, then they are more likely to be discharged on the same day. The class displayed on the node represents the highest quantity of patients. The colours

	min_samples_leaf					
max_leaf_nodes	1	100	200	300	400	500
5	0.8846	0.8846	0.8846	0.8846	0.8846	0.8846
10	0.8913	0.8913	0.8913	0.8913	0.8913	0.8913
15	0.8950	0.8950	0.8950	0.8950	0.8925	0.8925
20	0.8956	0.8956	0.8950	0.8950	0.8925	0.8925
25	0.8977	0.8976	0.8962	0.8963	0.8936	0.8939
30	0.8989	0.8987	0.8968	0.8968	0.8941	0.8939

(a) Accuracy score.

	min_samples_leaf					
max_leaf_nodes	1	100	200	300	400	500
5	0.8475	0.8475	0.8475	0.8475	0.8475	0.8475
10	0.8645	0.8645	0.8645	0.8645	0.8645	0.8645
15	0.8643	0.8643	0.8643	0.8643	0.8916	0.8916
20	0.8641	0.8641	0.8643	0.8643	0.8916	0.8916
25	0.8637	0.8617	0.8598	0.8594	0.8844	0.8804
30	0.8570	0.8617	0.8611	0.8635	0.8812	0.8804

(b) Precision score.

	min_samples_leaf					
max_leaf_nodes	1	100	200	300	400	500
5	0.8944	0.8944	0.8944	0.8944	0.8944	0.8944
10	0.8941	0.8941	0.8941	0.8941	0.8941	0.8941
15	0.9019	0.9019	0.9019	0.9019	0.8755	0.8755
20	0.9033	0.9033	0.9019	0.9019	0.8755	0.8755
25	0.9080	0.9094	0.9081	0.9088	0.8829	0.8865
30	0.9163	0.9117	0.9083	0.9061	0.8863	0.8865

(c) Recall score.

	min_samples_leaf					
max_leaf_nodes	1	100	200	300	400	500
5	7.8660	7.8724	7.9036	10.9756	7.9045	8.0315
10	10.7175	11.0817	10.9410	13.6632	11.7000	11.1358
15	12.9169	12.8437	12.6717	16.5188	13.4148	12.7473
20	15.1352	14.7659	14.7613	16.1299	15.6667	14.9961
25	16.1773	15.0654	15.7308	16.7057	18.3250	16.5215
30	16.7448	16.9692	17.1850	17.7688	16.7887	17.6103

(d) Computational time in seconds (s).

Table 5.21: The classification tree results for accuracy score (a), precision score (b), recall score (c) and computational time (d) for continuous age and continuous frailty.

	min_samples_leaf					
max_leaf_nodes	1	100	200	300	400	500
5	0.8846	0.8846	0.8846	0.8846	0.8846	0.8846
10	0.8913	0.8913	0.8913	0.8913	0.8913	0.8913
15	0.8950	0.8950	0.8950	0.8950	0.8925	0.8925
20	0.8956	0.8956	0.8950	0.8950	0.8925	0.8925
25	0.8977	0.8976	0.8962	0.8963	0.8936	0.8936
30	0.8989	0.8987	0.8968	0.8968	0.8941	0.8941

(a) Accuracy score.

	min_samples_leaf					
max_leaf_nodes	1	100	200	300	400	500
5	0.8475	0.8475	0.8475	0.8475	0.8475	0.8475
10	0.8645	0.8645	0.8645	0.8645	0.8645	0.8645
15	0.8643	0.8643	0.8643	0.8643	0.8916	0.8916
20	0.8641	0.8641	0.8643	0.8643	0.8916	0.8916
25	0.8637	0.8617	0.8598	0.8594	0.8844	0.8844
30	0.8570	0.8617	0.8611	0.8635	0.8810	0.8844

(b) Precision score.

	min_samples_leaf					
max_leaf_nodes	1	100	200	300	400	500
5	0.8944	0.8944	0.8944	0.8944	0.8944	0.8944
10	0.8941	0.8941	0.8941	0.8941	0.8941	0.8941
15	0.9019	0.9019	0.9019	0.9019	0.8755	0.8755
20	0.9033	0.9033	0.9019	0.9019	0.8755	0.8755
25	0.9080	0.9094	0.9081	0.9088	0.8829	0.8829
30	0.9163	0.9117	0.9083	0.9061	0.8865	0.8829

(c) Recall score.

	min_samples_leaf					
max_leaf_nodes	1	100	200	300	400	500
5	8.3217	8.5695	8.8101	8.8153	11.3366	25.3592
10	11.3205	11.6050	11.8029	12.3844	14.7207	17.2309
15	13.5908	13.5596	13.6742	13.7826	20.0681	16.4461
20	15.5900	15.6141	15.9532	15.9093	20.6382	20.5908
25	16.5625	16.1451	16.9353	17.2625	19.2398	19.6446
30	17.8176	17.5732	18.1772	19.2464	25.3592	22.3077

(d) Computational time in seconds (s).

Table 5.22: The classification tree results for accuracy score (a), precision score (b), recall score (c) and computational time (d) for continuous age and grouped frailty.

of the node symbolise the weighting within the group, with the darkest orange representing discharge on the same day (<1) and the darkest blue representing admittance overnight (≥1). There are 16 leaf nodes representing the '≥ 1' class, with one node generating a gini of zero, and therefore is a perfect classification.

The model produced 30 leaf nodes, displaying the class of patient, the majority fall into. The user is then able to analyse the individual results within the nodes to determine the average LOS's. They are able to achieve this by taking the individual patient clusters from the nodes and analysing them separately. Additionally, when new patients are admitted to a ward, they are able to determine at the point of arrival with their characteristics, what their expected LOS should be. This is also beneficial to determine where targets of being discharged on the same day of admission are not being met, for example, surgical cases. It also displays influencing

max_leaf_nodes	min_samples_leaf 1	100	200	300	400	500
5	0.8846	0.8846	0.8846	0.8846	0.8846	0.8846
10	0.8913	0.8913	0.8913	0.8913	0.8913	0.8913
15	0.8950	0.8950	0.8950	0.8950	0.8925	0.8925
20	0.8956	0.8956	0.8956	0.8956	0.8936	0.8939
25	0.8977	0.8976	0.8962	0.8963	0.8936	0.8939
30	0.8989	0.8987	0.8968	0.8968	0.8939	0.8937

(a) Accuracy score.

max_leaf_nodes	min_samples_leaf 1	100	200	300	400	500
5	0.8475	0.8475	0.8475	0.8475	0.8475	0.8475
10	0.8645	0.8645	0.8645	0.8645	0.8645	0.8645
15	0.8643	0.8643	0.8643	0.8643	0.8916	0.8916
20	0.8641	0.8641	0.8643	0.8643	0.8944	0.8944
25	0.8637	0.8617	0.8598	0.8594	0.8844	0.8804
30	0.8570	0.8617	0.8611	0.8635	0.8804	0.8731

(b) Precision score.

max_leaf_nodes	min_samples_leaf 1	100	200	300	400	500
5	0.8944	0.8944	0.8944	0.8944	0.8944	0.8944
10	0.8941	0.8941	0.8941	0.8941	0.8941	0.8941
15	0.9019	0.9019	0.9019	0.9019	0.8755	0.8755
20	0.9033	0.9033	0.9019	0.9019	0.8829	0.8829
25	0.9080	0.9094	0.9081	0.9088	0.8829	0.8865
30	0.9163	0.9117	0.9083	0.9061	0.8865	0.8819

(c) Recall score.

max_leaf_nodes	min_samples_leaf 1	100	200	300	400	500
5	9.0003	9.1338	8.0463	7.9693	8.0229	7.9960
10	11.8935	12.6970	11.2224	11.1074	11.3092	11.2887
15	13.8198	13.0589	13.0067	13.5755	13.3384	13.1068
20	16.3988	15.6034	14.9843	15.0497	15.4588	15.3504
25	20.8890	15.9613	16.0429	17.0717	16.3030	16.8512
30	24.0505	18.3618	17.5238	17.6762	17.7605	17.3051

(d) Computational time in seconds (s).

Table 5.23: The classification tree results for accuracy score (a), precision score (b), recall score (c) and computational time (d) for grouped age and no frailty.

max_leaf_nodes	min_samples_leaf 1	100	200	300	400	500
5	0.8846	0.8846	0.8846	0.8846	0.8846	0.8846
10	0.8913	0.8913	0.8913	0.8913	0.8913	0.8913
15	0.8950	0.8950	0.8950	0.8950	0.8925	0.8925
20	0.8956	0.8956	0.8950	0.8950	0.8925	0.8925
25	0.8977	0.8976	0.8962	0.8963	0.8936	0.8939
30	0.8989	0.8987	0.8968	0.8968	0.8941	0.8939

(a) Accuracy score.

max_leaf_nodes	min_samples_leaf 1	100	200	300	400	500
5	0.8475	0.8475	0.8475	0.8475	0.8475	0.8475
10	0.8645	0.8645	0.8645	0.8645	0.8645	0.8645
15	0.8643	0.8643	0.8643	0.8643	0.8916	0.8916
20	0.8641	0.8641	0.8643	0.8643	0.8916	0.8916
25	0.8637	0.8617	0.8598	0.8594	0.8844	0.8804
30	0.8570	0.8617	0.8611	0.8635	0.8812	0.8804

(b) Precision score.

max_leaf_nodes	min_samples_leaf 1	100	200	300	400	500
5	0.8944	0.8944	0.8944	0.8944	0.8944	0.8944
10	0.8941	0.8941	0.8941	0.8941	0.8941	0.8941
15	0.9019	0.9019	0.9019	0.9019	0.8755	0.8755
20	0.9033	0.9033	0.9019	0.9019	0.8755	0.8755
25	0.9080	0.9094	0.9081	0.9088	0.8829	0.8865
30	0.9163	0.9117	0.9083	0.9061	0.8863	0.8865

(c) Recall score.

max_leaf_nodes	min_samples_leaf 1	100	200	300	400	500
5	9.0269	8.7352	9.0878	8.5125	8.1122	8.6958
10	12.0635	12.2949	12.1766	11.5033	12.1410	12.7141
15	13.8758	14.2554	14.4314	13.3940	14.7385	15.0438
20	16.0829	15.9826	15.1769	15.3548	17.0093	17.0980
25	16.9550	17.4665	16.3701	16.8606	18.1321	20.0174
30	18.0005	18.0719	18.6015	18.0826	18.6726	19.3604

(d) Computational time in seconds (s).

Table 5.24: The classification tree results for accuracy score (a), precision score (b), recall score (c) and computational time (d) for grouped age and continuous frailty.

max_leaf_nodes	min_samples_leaf 1	100	200	300	400	500
5	0.8846	0.8846	0.8846	0.8846	0.8846	0.8846
10	0.8913	0.8913	0.8913	0.8913	0.8913	0.8913
15	0.8950	0.8950	0.8950	0.8950	0.8925	0.8925
20	0.8956	0.8956	0.8950	0.8950	0.8925	0.8925
25	0.8977	0.8976	0.8962	0.8963	0.8936	0.8936
30	0.8989	0.8987	0.8968	0.8968	0.8941	0.8936

(a) Accuracy score.

max_leaf_nodes	min_samples_leaf 1	100	200	300	400	500
5	0.8475	0.8475	0.8475	0.8475	0.8475	0.8475
10	0.8645	0.8645	0.8645	0.8645	0.8645	0.8645
15	0.8643	0.8643	0.8643	0.8643	0.8916	0.8916
20	0.8641	0.8641	0.8643	0.8643	0.8916	0.8916
25	0.8637	0.8617	0.8598	0.8594	0.8844	0.8844
30	0.8570	0.8617	0.8611	0.8635	0.8810	0.8844

(b) Precision score.

max_leaf_nodes	min_samples_leaf 1	100	200	300	400	500
5	0.8944	0.8944	0.8944	0.8944	0.8944	0.8944
10	0.8941	0.8941	0.8941	0.8941	0.8941	0.8941
15	0.9019	0.9019	0.9019	0.9019	0.8755	0.8755
20	0.9033	0.9033	0.9019	0.9019	0.8755	0.8755
25	0.9080	0.9094	0.9081	0.9088	0.8829	0.8829
30	0.9163	0.9117	0.9083	0.9061	0.8865	0.8829

(c) Recall score.

max_leaf_nodes	min_samples_leaf 1	100	200	300	400	500
5	8.9117	8.9096	10.8248	9.9785	8.8351	10.1257
10	12.1086	12.1666	13.6021	14.9230	12.3419	12.3234
15	14.0532	17.2649	16.1410	16.0983	14.5993	13.8329
20	16.5178	17.3444	21.1871	17.1917	17.0247	15.6259
25	17.5062	19.6158	19.7483	17.1917	17.5036	16.7736
30	17.9821	20.0913	19.3814	19.3519	18.2937	17.5289

(d) Computational time in seconds (s).

Table 5.25: The classification tree results for accuracy score (a), precision score (b), recall score (c) and computational time (d) for grouped age and grouped frailty.

factors in LOS. For example, a patient who is admitted through the elective waiting list to the specialty gynaecology, will have their LOS dependent on the day they are admitted. If they are admitted on a Monday, they fall into the '$\geq$ 1' class, otherwise, they fall into the '$<$ 1' class. This can be used to help practitioners plan elective admission and when they should be admitted in order to reduce unexpected and prolonged LOS's.

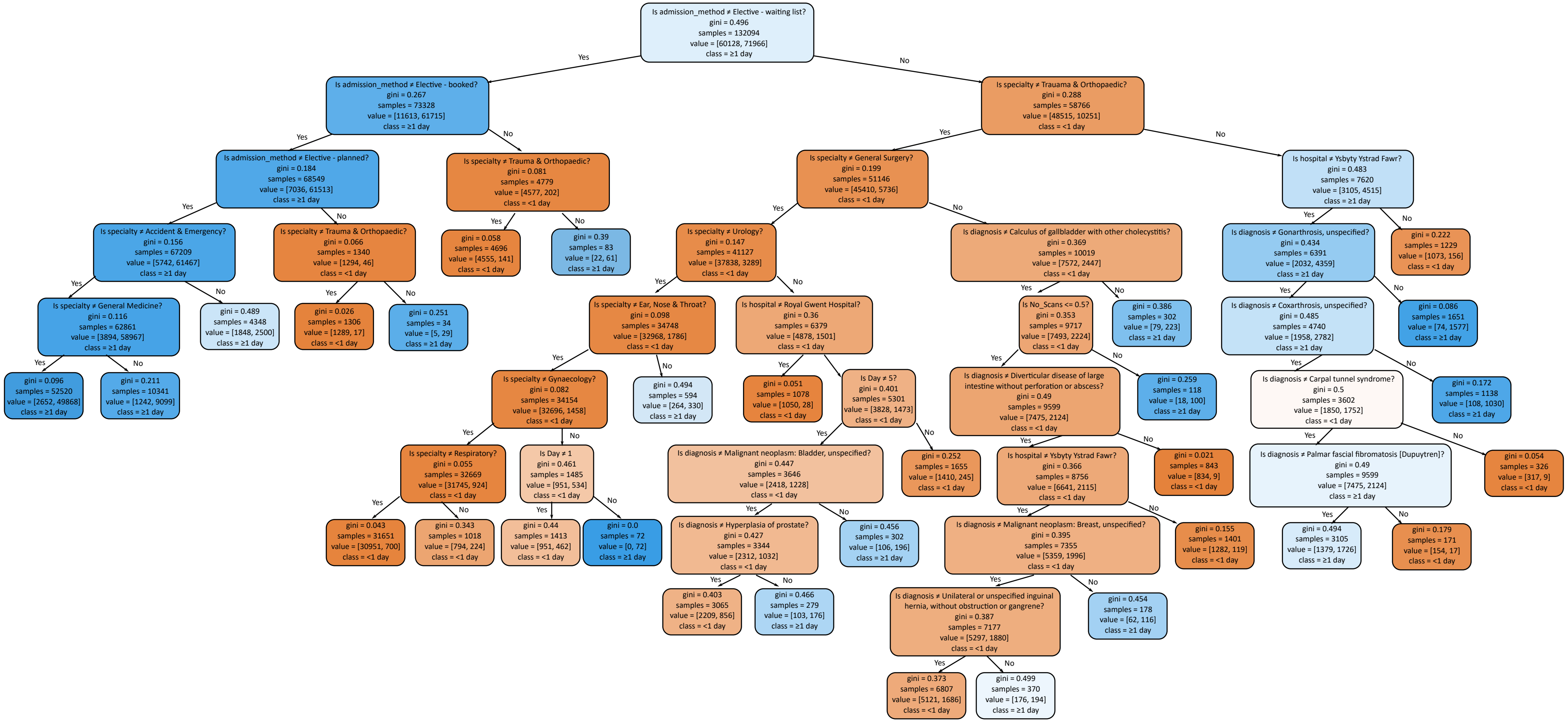


Figure 5.7: Classification tree with an accuracy score of 89.89% for predicting grouped LOS for patients within ABUHB, consisting of 30 terminal nodes.

5.3.3.3 Random Forests

The CART model parameters that yielded the best results, were then implemented into a random forest model to determine if there was an improvement in these scores. The highest scoring CART model parameters were used, as these CART models will be used for the remainder of the research and therefore a direct comparison could be made in terms of accuracy and computational time.

Parameter	**RandomForestRegressor**	**RandomForestClassifier**
criterion	"squared_error"	"gini"
max_depth	None	None
min_samples_split	2	2
min_samples_leaf	100	1
min_weight_fraction_leaf	0	0
max_features	None	None
random_state	None	None
max_leaf_nodes	30	30
min_impurity_decrease	0	0
bootstrap	True	True
oob_score	False	False
n_jobs	None	None
warm_start	False	False
ccp_alpha	0	0
max_samples	None	None
class_weight	N/A	None

Table 5.26: The parameters used within the random forests using the 'RandomForestRegressor' and 'RandomForestClassifier' algorithms within Python.

The variable 'n_estimators' underwent hyperparameter tuning to determine the value that produced the highest scores. This variable determines the number of trees to be used within the model. Typically in random forests, the larger the number of trees, the better the result produced, however, the higher the computational cost [270]. A range from 10 to 50 estimators were investigated, in increments of 10, to determine R^2 and accuracy against computational time.

	n_estimators	10	20	30	40	50
Regression	R^2 Score	0.342912	0.342917	0.343148	0.343291	0.343171
	Computational Time (s)	130.80	260.13	390.63	535.95	655.92
Classification	Accuracy	0.883539	0.887809	0.8892296	0.890443	0.890928
	Computational Time (s)	16.58	19.37	27.72	36.87	37.88

Table 5.27: The accuracy and R^2 score for the regression and classification random forest models, measured against the n_estimator parameter, respectively. Each model's computational runtime was additionally provided.

Table 5.27 displays the random forest results for the regression and classification experiments. The regression random forest increased the regression tree R^2 by

0.0491% from 34.28%. This was achieved by setting 'n_estimators' to 40, in a computational time of 535.95 seconds. Since there is only an increase of 0.0491% and a computational increase of 508.9714 seconds, the trade off is not worth the additional increase in accuracy.

Additionally, the classification random forest did not improve the accuracy of the model compared to the classification tree. There is a difference of 0.8% in accuracy levels. One explanation for this phenomenon could be the optimal parameters for the random forest are not the same as those for the classification tree.

5.3.4 Predictive Analytics Summary

This section has discussed the application of predictive analytics to data within ABUHB. Linear and logistic regressions were firstly performed, detailing the influence of a variable for the variability in the LOS. Admission method, diagnosis and specialty categories were the most influencing factors in the LOS.

Regression trees were developed to identify groupings of patients in order to predict continuous LOS. Figure 5.6 displays the different groupings and classifications of patients LOS, with the average LOS for each group. A total of 30 groups were identified due to the 'max_leaf_nodes' being equal to 30. This resulted in an R^2 value of 34.23%.

Classification trees were then built with the groupings of discharged on the day of arrival or admitted overnight (Figure 5.7). The best trade off solution between accuracy, precision, recall and computational time, was 30 'max_leaf_nodes' and one 'min_samples_leaf', yielding an accuracy score of 89.89%. Comparing the classification to regression trees showed there to be a 55.66% improvement by grouping LOS into two groups.

CART models excel in face validity and usefulness, primarily due to their interpretability compared to other prediction methods [271, 272]. The tree structure allows straightforward visualisation of decision-making pathways, making it easier for healthcare practitioners and managers to comprehend and trust the model's output. Furthermore, CART offers the advantage of forced splits, allowing collaboration with healthcare practitioners to jointly determine clinical and statistically meaningful groupings. This ensures the model aligns with real-world scenarios and policy levers. Another benefit of CART is the ability to create targeted groupings, providing valuable insights into specific demographics. If the user wanted fewer numbers of patient groupings to work with, the technique of pruning could be used to selectively remove branches or nodes from the initial tree to simplify its structure.

CART was utilised as a predictive tool by applying it solely to historical data. In this context, the model was designed to predict outcomes based on patterns and relationships found within the past data. By using the historical data as the training set, the CART model identified decision rules and splits that best separated different groups or categories, enabling it to make predictions of LOS for new instances in the future. This approach ensured that the CART model served as a valuable predictive tool for understanding and anticipating events based on past trends. Chapter 6 will involve future scenario analysis, where the CART LOS prediction tool will be utilised to assess the impact of potential increases in certain patient grouping demands on LOS, bed requirements, and staffing. By applying the CART model to project future patient groupings and LOS, the study aims to evaluate how changing demands could influence resource needs within the healthcare system.

The consideration of runtime in building a CART model is justified by the need for computational efficiency, especially when the model is intended for frequent use or real-time decision-making scenarios. By prioritising faster run times, organisations can obtain timely predictions and optimise resource allocation efficiently, which is crucial in time-sensitive applications, such as healthcare. The runtime of a CART model can be compared to more complex models, i.e., random forests. While random forests can often achieve higher accuracy due to their ensemble nature, they come at the cost of increased computational requirements. Random forests consist of multiple decision trees, and the process of building and combining these trees can be computationally intensive, particularly for large data sets. This computational issue has the implication that they are less suitable for real-time or repeated use applications. The random forests performed in this section were both computationally heavier in performance time and did not cause a large enough increase in the overall accuracy and R^2 scores, compared to the CART models.

The CART models showed improvement in the simple linear and logistic regressions and therefore highlighted the benefits of using these techniques to predict patient classifications. The CART models will be linked with the prescriptive models (introduced in Section 5.4), in Chapter 6. Here, an evaluation will take place into the benefits of using CART models over traditional averages.

5.4 Prescriptive Analytics Results

This section will look at the deterministic and two-stage stochastic models developed in Chapter 4. The demographics of ABUHB will be used in order to determine the most efficient way to organise specialties and nursing staff amongst a network of hospitals.

5.4.1 Model Data

The deterministic and two-stage stochastic models require user inputs to generate results. In total, the deterministic model requires 11 variables, whereas the two-stage stochastic model requires 17 variables. The collaboration with senior and clinical partners within ABUHB has provided justification around the assumptions and values used within these examples. Discussion around the selection of these variables will take place in Sections 5.4.1.1 to 5.4.1.3. A complete list of the variables used within the models can be found in Table C.6 for the deterministic model and Table C.7 for the two-stage stochastic model in the Appendix.

5.4.1.1 Hospitals and Regions

Within ABUHB, there are 10 hospitals located within the five regions as discussed in Section 1.2. In addition, the data contained an additional four medical sites where patients receive treatment. It is important to include patients attending these sites within the model to ensure the entire demand is included and sufficient beds and staff are planned. Table 5.28 displays the six regions included in the model and their associated hospitals.

Region	**Hospitals**
Region 1 (Newport)	Royal Gwent Hospital (RGH), St Woolos Acute Hospital (STWAH), St Woolos Community Hospital (STWCH)
Region 2 (Caerphilly)	Ysbyty Ystrad Fawr (YYF), Rhymney Integrated Health and Social Care Centre (RIHSC)
Region 3 (Blaenau Gwent)	Ysbyty Aneurin Bevan (YAB)
Region 4 (Torfaen)	County Hospital (CH)
Region 5 (Monmouthshire)	Nevill Hall Hospital (NHH), Chepstow Community Hospital (CCH), Monnow Vale Integrated Health and Social Care Centre (MVHSCF)
Other	University Hospital of Wales, Offsite, Outsource, Outsource - CareUK

Table 5.28: List of the 14 care locations within ABUHB and their associated regions.

The parameter $\mathrm{UB}_h^{\text{max, bed, 1st}}$ was determined from online publicly available data recorded by the Welsh Government [273]. The Welsh Government record, over a year period, the average daily beds available for each specialty in each hospital. Within the 10 main hospitals, a total of 1,704 beds were available per day to the entire population. Therefore, the maximum number of beds available will be scaled to represent the proportion of elderly admitted. For the four additional hospitals, which are either not at a hospital site or outside the trust, a fixed value of 20 was given. For the second stage maximum number of beds, $\mathrm{UB}_h^{\text{max, bed, 2nd}}$, an additional 10% of beds will be able to be made available, either by opening additional wards, transferring patients to other hospitals in the region, or to temporarily have patients waiting in corridors for permanent beds.

5.4.1.2 Hospitals and Specialties

ABUHB provides 29 different specialties among the 14 hospital and care locations, resulting in a combination of 406 unique hospital and specialty combinations. However, in practice, there are 90 combinations of hospital and specialty locations, as shown in Figure 1.5 and Appendix A. These locations will determine the value of $K_{s,h}$. If a specialty is able to open in a hospital, then the value will be equal to the $\mathrm{UB}_h^{\mathrm{max,bed}}$, otherwise the value is zero. This therefore has the assumption that if a specialty can open, the hospital can choose to open all their beds to that specialty. This assumption was derived from the current practice within certain ABUHB hospitals. For example, the hospitals, Rhymney Integrated Health and Social Care (RIHSC) and Monnow Vale Health and Social Care Facility (MVHSCF), only provide beds for the specialty 'GP Other'. Looking forward, the health board could consider an innovative approach to enhance its services by consolidating the locations where they offer care, leading to the establishment of specialty-focused hospitals. This strategic revamp would enable them to streamline their resources and provide more targeted and specialised medical care to the community. Within the specialties, some may not conventionally be viewed as distinct ones. However, within ABUHB, these areas are recognised as having their unique capacity to admit patients and provide dedicated care with their own dedicated beds and staff, e.g., GP and anaesthetics [273], and therefore will be included in the overall list of 29 specialties.

Online publicly available data was used for the costings per specialty from Public Health Scotland. This data can be viewed within 'Table 3: Hospital cost breakdown R040 - Specialty costs and activity - inpatients in all specialties (excluding long stay), by specialty (Excel file, 269KB)' [11]. Within the Excel spreadsheet, the 'Direct Cost per Case' is given, and to make this comparable to all entries, only the 'Medical and Dental' costs will be included. The assumption was made that this was the daily cost per specialty. It has been assumed that the cost of scans required whilst an inpatient has been absorbed into these costs. For pain specialty, a different data set was utilised from Public Health Scotland, namely, 'Table 3: Hospital cost breakdown R040LS - Specialty costs and activity - day cases, by specialty (Excel file, 201KB)' [11]. The cost is listed as the 'Direct Cost per Case' and the same 'Medical and Dental' column was taken. As this file specifically mentions daycases, it is assumed that this is the daily specialty cost. The Scottish population follows a similar demographic to those in Wales and has similar operational running costs within hospitals. Additionally, the health boards in Scotland contain a variety of community and acute hospitals. Table 5.29 displays the minimum, maximum and weighted average cost for each specialty in the 2019-2020 financial year. In order to determine similar costings for hospitals within ABUHB, specialty cost values per hospital were randomly generated within the range of the Scottish data. Additionally, the values produced the same weighted average as the Scottish data. One limitation of using the Scottish data over Welsh data is that it is not representative of the current practice within ABUHB. Scottish data was chosen because it is located within the UK and costing values are expected to be similar. The availability of more specific data tailored to ABUHB could potentially yield more reliable and accurate results. There is a large range of costs between hospitals potentially due to a number of reasons such as the type of diagnosis and care required [274]. Our objective is to introduce variation into ABUHB by utilising randomly generated values within the range of minimum and maximum costs observed in the NHS Scotland data. This approach enables us to capture diverse scenarios and account for potential fluctuations in costs within ABUHB's context while maintaining the same average cost as observed in the NHS Scotland data. In general in the NHS, there is no fixed tariff for bed days for specific specialties. Some specialties in ABUHB did not overlap with the Scottish data, and therefore the category, 'Medical Other', was selected for these specialties (anaesthetics, community medicine, diabetes and endocrinology and radiology). To determine the second stage hospital costs, $\mathrm{c}_{s,h}^{\text{bed, 2nd}}$, an additional 20% was added to each of the $\mathrm{c}_{s,h}^{\text{bed, 1st}}$ values. This percentage was fixed across all specialties and hospitals and was selected to provide a significant penalty for not having sufficient demand.

5.4.1.3 Staffing

Within the NHS, there are different levels of experience among nursing staff [262]. Typically on a ward, there will be a mixture of different bands of nurses to make up the team of nurses. It is also important to ensure there is a mixture of different skill sets on a ward at a time [275]. Within the NHS there are different nurse to patient ratios, ranging from 1:1 to 1:10 [276]. The most critical patients require more direct care from nurses, i.e., intensive care units often have required ratios of one nurse to either one or two patients [277]. Within general inpatient wards this ratio varies. The ratio of nurses to patients will vary between different specialties, with the more acute specialties requiring more nurses to patients. It was decided for more low-need based wards, e.g., community medicine, a ratio of 1:10 would be required. For more acute wards, a ratio of 1:4 would be necessary, e.g., A&E. Finally, the remaining wards would be assigned a required ratio of 1:8. It will be assumed that an equal number of nurses across the bands will be required on the wards, with each band required to meet a given ratio.

After discussion with the senior staff in the health board, two NHS nursing bands would be considered since this is their core members of staff required on a ward, namely bands five and six nurses [278]. These nursing bands are crucial in ensuring

Specialty	Minimum Cost (£)	Weighted Average (£)	Maximum Cost (£)
Accident & Emergency	22	247	924
Anaesthetics	34	1,021	2,370
Cardiology	4	614	1,513
Care of the Elderly	131	577	6,021
Community Medicine	34	1021	2,370
Dermatology	203	1,381	2,446
Diabetes & Endocrinology	34	1,021	2,370
Ear, Nose & Throat	77	491	1,436
Gastroenterology	112	656	1,472
General Medicine	2	290	1,418
General Surgery	11	541	1,517
GP Other	6	325	1,117
Gynaecology	180	517	2,526
Haematology	411	1,208	5,214
Infectious Diseases	438	711	1,188
Intermediate Care	0	118	361
Maxillo-Facial	96	1,410	8,637
Neurology	620	1,273	3,260
Ophthalmology	166	729	10,895
Pain	6	128	865
Plastic Surgery	179	902	1,399
Radiology	34	1,021	2,370
Radiotherapy and Oncology	48	1,089	2,182
Rehabilitation	58	1,455	30,305
Respiratory	75	448	1,818
Restorative Dentistry	1	140	178
Rheumatology	155	596	1,256
Trauma and Orthopaedic	4	703	1,633
Urology	77	379	17,899

Table 5.29: Specialty cost data displaying the minimum, weighted average and the maximum daily bed cost per specialty, taken from NHS Scotland [11].

efficient and quality patient care, making them a priority for staffing considerations. The hourly pay for each band varies with experience, with Table 5.30 displaying the intermediate salary [12]. When there are insufficient staffing levels, bank or agency staff are required. These nurses have a higher hourly wage, due to the unreliability of shifts [13].

	Band 5	**Band 6**
$c_b^{\text{staff,1st}}$	£14.21	£17.48
$c_b^{\text{staff,2nd}}$	£18.95	£23.36

Table 5.30: Hourly nursing staffing costs per nursing band level within the NHS [12, 13].

As of June 2022, ABUHB employed approximately 1,935 registered adult and general nurses [279]. This value includes all bands of nurses across all specialties. For the purpose of the model, it will be assumed that 400 nurses are available, 200 for each band. Similarly, for the variable $\text{UB}_b^{\text{max, staff, 2nd}}$, it will be assumed that up to 200 additional nurses for each can be requested from the bank for each band.

5.4.2 Model Development

These variables will be implemented into models developed in both Microsoft Excel using the OpenSolver add-in and Python using the PuLP packages as a solver. Microsoft Excel was selected as a primary tool since it is familiar with the health board and open-source. The OpenSolver add-in was used over the built in Excel solver, due to the limitations on the number of decision variables and constraints. OpenSolver has previously been used within healthcare planning for both bed planning [280] and staff allocation [281]. A Python tool was additionally developed as it allows for flexibility within the number of hospitals, specialties and band levels of nurses. This adaptability ensures the model is dynamic to the changing needs and demographics of the health board. Python is also open-source, adaptable and easy to learn [282], which will enable future development by senior staff at ABUHB.

Two optimiser engines were used within the project: COIN-OR (Computational Infrastructure for Operations Research) and Gurobi. COIN-OR is a project managed by the COIN-OR Foundation with the aim to provide an “open-source community for operations research software” [283, 284]. The COIN-OR project consists of numerous smaller initiatives, including the development of various software for a variety of problems, methods, and coding languages. The COIN-OR community has developed two main linear programming solvers: CLP (COIN-OR Linear Programming), which primarily uses the simplex method as its core algorithm, and CBC (COIN-OR Branch and Cut), a mixed integer linear program-based (MILP) branch and cut library that also makes use of CLP. Although both of these solvers are designed in C++, they can be used with different languages through readily available packages such as PuLP. Since the deterministic and two-stage stochastic models developed require the decision variables to be integer, the CBC solver is the most suitable choice.

COIN-OR solvers are free and open source which is vital for ABUHB as the minimisation of cost is a necessity for the organisation. There exists more advanced commercial software including CPLEX [285] and Gurobi [286]. Gurobi is a widely used optimization solver that provides powerful tools for solving various mathematical optimization problems. The Gurobi Optimization, Inc. company focuses on delivering high-performance solvers for a range of applications in operations research and related fields. One of the distinguishing features of Gurobi is its ability to handle complex optimization models with speed and precision. Gurobi's interface supports multiple programming languages, including Python, C++, Java, and MATLAB, making it accessible to a broad audience of developers and researchers. A free academic license was obtained for Gurobi and provided us with unrestricted access to the solver's optimisation tools. This enabled the model to be solved quickly and efficiently, whilst also validating the CBC results.

Within both the OpenSolver and implementations offer the flexibility to adjust specific parameters, enabling users to enhance accuracy and control the execution time of their models. The first is the maximum length of time in which the programme is allowed to run before the model is stopped. Within Excel, this is denoted as 'Maximum Solution Time (seconds)' and within Python as 'maxSeconds'. Prior to applying any restrictions, these models would execute indefinitely until the optimal solution was discovered. However, by setting the appropriate value, users can now limit the runtime of the program, ensuring efficient execution and timely results while still achieving the best possible solution. For the purpose of this research, no restriction on runtime was set. The second option is the relative gap tolerance for the solver to stop, i.e., the solver will stop if it has found a feasible integer solution whose objective function is within the given percentage of the true integer optimal solution. In Excel, this is denoted as the 'Branch and Bound Tolerance (%)' and the 'fracGap' value in Python. In order to reach optimal values, this was set to 0, and the solvers will always reach the optimal values. Within Excel, there is the option to change the maximum number of iterations, which determines how many iterations the solver will use, however, this was left unbounded. The PuLP package does not have this built in as default.

The Microsoft Excel OpenSolver and the Python PuLP tools have been provided on Github [287], to allow the models to be usable by other researchers and healthcare specialists. As the parameters previously discussed are consistent between the Excel and Python models, the same objective function and decision variables are achieved. As a result, users have the flexibility to opt for either implementation as the same results will be achieved. This aspect highlights the interchangeability of the models, empowering users to utilise the most suitable platform while achieving identical outcomes.

5.4.3 Results

The following subsection will look at implementing the ABUHB data into the deterministic and two-stage stochastic optimisation models. Either Microsoft Excel or Python implementations can be used. Traditionally, the health board plan using averages, and therefore the deterministic model is meant to replicate their current planning process within ABUHB.

5.4.3.1 Experiment 1 - Three Years' Worth of Data

The first experiment examined the entire three years' worth of data from April 2017 to March 2020, to determine the daily average number of beds and staff required.

Recall the following equations discussed in Chapter 4, which were used to calculate the daily demand for each region.

$$\text{Average daily bed demand}_{s,h} = \text{Average LOS}_{s,h} \times \text{Average daily number of admissions}_{s,h} \quad \text{(4.24 revisited)}$$

$$D_{s,r} = \text{Average daily bed demand}_{s,r} = \sum_{h \in \mathcal{R}} \text{Average daily bed demand}_{s,h} \quad \text{(4.25 revisited)}$$

Using these equations against the whole data set, resulted in Table 5.31 which displayed the average daily demand for each specialty in each of the six regions. The demands were rounded to four decimal places, however, since beds have to be integer, the model would round these to the nearest integer.

Specialty	Region 1	Region 2	Region 3	Region 4	Region 5	Region 6
Accident & Emergency	2.1081	0	0	0	9.1846	0
Anaesthetics	4.6079	0	0	0	0	0
Cardiology	16.0947	0	0	0	9.8809	0.0002
Care of the Elderly	94.5387	57.7380	0.7489	8.7416	46.4786	0
Community Medicine	0	6.9952	0	0.3121	12.9756	0
Dermatology	2.4192	0	0	0	0	0
Diabetes & Endocrinology	14.5635	21.2838	0	0	17.2387	0
Ear Nose & Throat	3.2480	0.0041	0	0	0	0
Gastroenterology	13.0208	0.6065	0	0	20.1331	0.0725
General Medicine	84.9712	0.9846	0.0115	0	14.1695	0
General Surgery	46.5808	0.5390	0	0	21.8943	0.0006
GP Other	0	9.8734	0	0	15.2880	0
Gynaecology	2.1194	0.1222	0	0	1.1716	0.0002
Haematology	3.0718	0.0320	0	0	1.8339	0.0002
Infectious Diseases	7.1817	0	0	0	0	0
Intermediate Care	0	0	0.3426	0.3248	0	0
Maxillo-Facial	1.1831	0	0	0	0.0243	0
Neurology	1.5828	0	0	0	0	0
Ophthalmology	2.5538	0.0028	0	0	0.1968	0.5232
Pain	0.0586	0.0055	0	0.0037	0.0131	0
Plastic Surgery	0	0	0	0	0.0336	0.0002
Radiology	0.0146	0	0	0	0.0026	0
Radiotherapy & Oncology	0.2265	0	0	0	0	0
Rehabilitation	62.8610	32.9653	68.9710	31.9917	24.4659	0
Respiratory	29.8010	0	0	0	27.8372	0
Restorative Dentistry	0.0001	0	0	0	0	0
Rheumatology	0.0004	0.001	5 0	0	0.0128	0
Trauma & Orthopaedic	60.3126	0.6665	0	0	41.5186	0
Urology	12.3397	0.0521	0	0	0.2034	0.0702

Table 5.31: The daily bed demands for each specialty grouped by regions within ABUHB for three years' worth of patient admissions, rounded to four decimal places.

5.4.3.1.1 Deterministic Model

In order to meet the demand and satisfy the constraints, the deterministic model utilised the first stage variables only (Table C.6). The results yielded a daily cost of £904,280.80. In total, 1,026 beds across the health board were deployed with Figure 5.8 displaying the precise locations of these beds. In order to satisfy demand a total of 414 NHS nurses across a 24 hour period were required. The maximum total number of beds available to deploy across the health board is 1,510, showing that the elderly and frail patients would be running at a 68% occupancy level, with the demand level being 974.78 beds. The number of staff deployed is larger than the maximum 1:4 ratio of nurses to patients. This is due to the requirement of an integer number of nursing staff and therefore the model is rounding up.

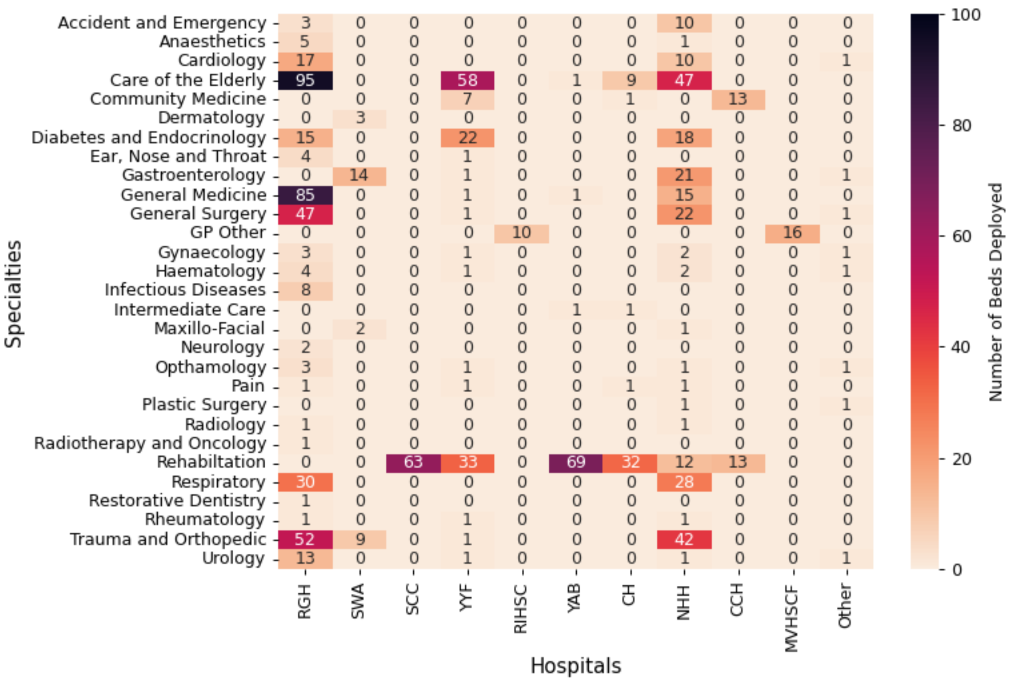


Figure 5.8: Heatmap of bed locations for each specialty within each hospital for the deterministic model for Experiment 1. Note that a darker colour indicates a larger number of beds are deployed.

5.4.3.1.2 Two-Stage Stochastic Model

The two-stage stochastic model was considered with three different scenarios. Table 5.32 displays each of the three scenarios and their associated probabilities. The average of all scenarios is equal to the deterministic daily demands (Table 5.31). The variable values can be seen within Table C.7.

Scenario	Probability
Demand remains the same	$33.3\dot{3}\%$
Demand increases by 20%	$33.3\dot{3}\%$
Demand decreases by 20%	$33.3\dot{3}\%$

Table 5.32: The three scenarios and their associated probabilities of occurring which will be used within the two-stage stochastic model.

Table 5.33 compares the results for the deterministic and two-stage stochastic models.

The two-stage stochastic model deployed an additional 197 beds compared to the deterministic model, deploying 862 in the first stage and a maximum of 361 in the second stage. Similarly, 70 additional nurses were deployed. The objective value increased by 4.56%, to a daily cost of £945,500.48.

	Total Beds		Total Staff		Objective Function Value (£)
	x^{bed}	u^{bed}	x^{staff}	u^{staff}	
Deterministic	1,026	-	414	-	904,280.80 = EV
Stochastic	862	361	348	136	945,500.48 = RP

Table 5.33: The EV and RP values for the x^{bed}, x^{staff}, u^{bed} and u^{staff} decision variables and the objective function value for Experiment 1.

The location of each of the 1,223 beds can be seen within Figure 5.9.

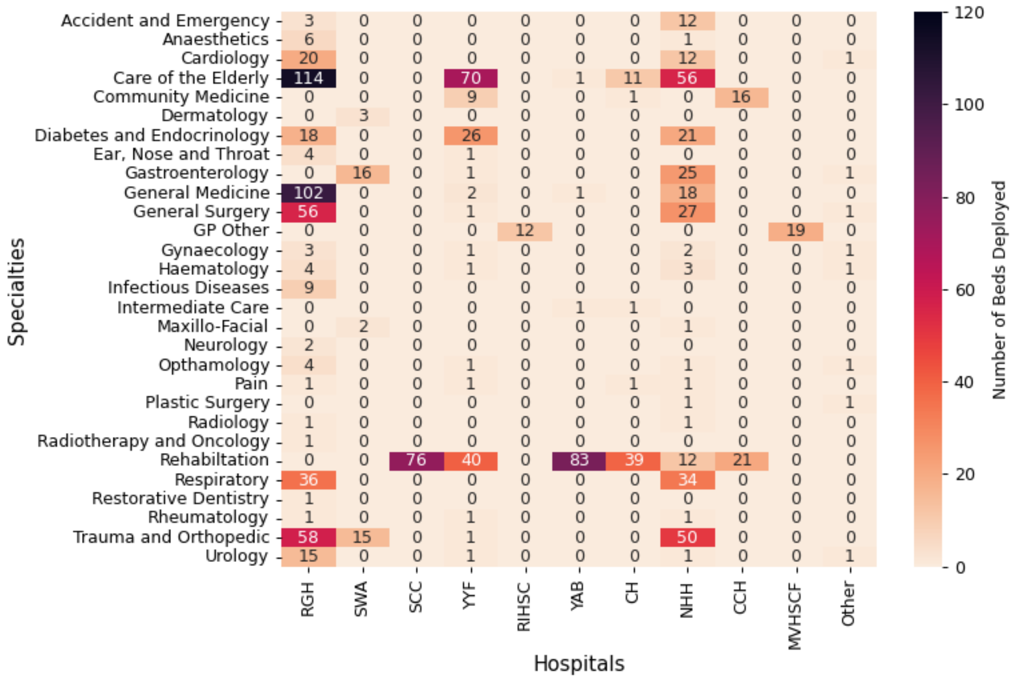


Figure 5.9: Heatmap of bed locations for each specialty within each hospital for the two-stage stochastic model for Experiment 1. Note that a darker colour indicates a larger number of beds are deployed.

5.4.3.1.3 Test A

The first test as discussed in Section 4.2.2.1 involved using the results from the deterministic model and fixing these as the first stage variables in the two-stage stochastic model. Table 5.34 displays the results for the VSS.

The VSS can be calculated to be 3.38%, equating to £31,956.24 per day. Whilst a value of 3.38% is low, the saving over a year period equates to £11,664,027.60. This

	Total Beds		**Total Staff**		**Objective Function Value (£)**
	x^{bed}	u^{bed}	x^{staff}	u^{staff}	
Deterministic	1,026	-	414	-	904,280.80 = EV
Stochastic	862	361	348	136	945,500.48 = RP
Test A	1,026	194	414	98	977,456.72 = EEV

Table 5.34: The EV, RP and EEV values for the x^{bed}, x^{staff}, u^{bed} and u^{staff} decision variables and the objective function value for Experiment 1.

showed that there is a benefit in using the stochastic solution over the deterministic solution.

To understand why the deterministic solution was performing poorly, we calculated the LUSS and LUDS values. This would determine whether the model was too optimistic or the locations of beds and staff were incorrect.

5.4.3.1.4 Test B

The second test discussed in Section 4.2.2.2, involved fixing the first stage variables which are at zero or the lower bound in the deterministic problem, and then to compute in the stochastic programme. In this case, there were hospitals, which even though beds could be deployed, the value was zero. This would determine if the deterministic model produced the correct, non-zero variables.

Recalling Equation (4.9), the RP result can then be compared to the ESSV.

$$LUSS = ESSV - RP \tag{4.9 revisited}$$

	Total Beds		**Total Staff**		**Objective Function Value (£)**
	x^{bed}	u^{bed}	x^{staff}	u^{staff}	
Deterministic	1,026	-	414	-	904,280.80 = EV
Stochastic	862	361	348	136	945,500.48 = RP
Test A	1,026	194	414	98	977,456.72 = EEV
Test B	862	361	348	136	945,500.48 = ESSV

Table 5.35: The EV, RP, EEV and ESSV values for the x^{bed}, x^{staff}, u^{bed} and u^{staff} decision variables and the objective function value for Experiment 1.

The results in Table 5.35 show that LUSS value was calculated to be £0 since the ESSV and RP values are equal. Therefore this equates to the perfect skeleton solution and suggests the variables selected by the first stage of the solution are robust.

5.4.3.1.5 Test C

The final test determined the upgradeability of the model by adding the number of beds and staff deployed in the deterministic model as a constraint, for the stochastic model (Section 4.2.2.3).

	Total Beds		**Total Staff**		**Objective Function Value (£)**
	x^{bed}	u^{bed}	x^{staff}	u^{staff}	
Deterministic	1,026	-	414	-	904,280.80 = EV
Stochastic	862	361	348	136	945,500.48 = RP
Test A	1,026	194	414	98	977,456.72 = EEV
Test B	862	361	348	136	945,500.48 = ESSV
Test C	1,035	185	414	84	976,601.16 = EIV

Table 5.36: The EV, RP, EEV, ESSV and EIV values for the x^{bed}, x^{staff}, u^{bed} and u^{staff} decision variables and the objective function value for Experiment 1.

The LUDS value was calculated by the difference between EIV and RP, which using Table 5.36, determined this value to be £31,100.68. Since LUDS < VSS, this demonstrated partial upgradeability and additional beds and staff were deployed in the first stage compared to the deterministic solution.

5.4.3.1.6 Experiment 1 Summary

This section has analysed bed and staffing requirements based on three years' worth of data, using the daily demand average. The VSS has shown that there is a 3.38% saving per day using costings from NHS Scotland [11], equalling £7,681,877.60 per year. Any additional potential benefits that can be utilised by the NHS are critically important.

In conclusion, the deterministic solution did not perform well in a stochastic environment because too few beds and staff were deployed (1,026 beds and 414 staff compared to 1,223 beds and 484 staff). As the LUSS was equal to zero meaning the deterministic solution has the perfect skeleton solution, but plans on deploying too many beds and staff for the demand. Within ABUHB, planning currently takes place based on averages, i.e., the deterministic model. Although ABUHB costing figures have not been used, the differences between the deterministic and two-stage stochastic model have been shown, with the VSS being calculated to demonstrate the benefit of the second method. Therefore, this has shown there is evidence for the NHS to move away from simply planning on averages and use more sophisticated techniques.

5.4.3.2 Experiment 2

The second experiment analysed the beds on a year-to-year basis, to determine if there were yearly differences in the number of beds and staff that should be deployed. Since the yearly demand of patients contained little variation (Section 5.2.1), from a high level, it could be assumed that the beds and staff required would contain little variation. This experiment would therefore determine if there was a need to plan on a smaller scale horizon.

Tables 5.37 and 5.38 display the regional demands for each specialty for each year. The same first stage (Table C.6) and second stage (Table C.7) variables would be used as those in Experiment 1 to allow direct comparisons.

Specialty	**Region 1**			**Region 2**			**Region 3**		
	2017-2018	**2018-2019**	**2019-2020**	**2017-2018**	**2018-2019**	**2019-2020**	**2017-2018**	**2018-2019**	**2019-2020**
Accident & Emergency	2.2703	2.2025	1.8524	0	0	0	0	0	0
Anaesthetics	4.5049	4.4998	4.8186	0	0	0	0	0	0
Cardiology	17.2662	14.6579	16.3594	0	0	0	0	0	0
Care of the Elderly	76.9707	88.5761	118.0052	55.7326	58.4843	58.9937	0	0.3977	1.8461
Community Medicine	0	0	0	7.1593	6.0895	7.7349	0	0	0
Dermatology	2.8287	2.1355	2.2939	0	0	0	0	0	
Diabetes & Endocrinology	16.7264	11.5043	15.4574	24.7432	19.8417	19.2722	0	0	0
Ear, Nose & Throat	3.0791	3.7301	2.9358	0.0084	0.0027	0.0013	0	0	0
Gastroenterology	12.1823	10.4803	16.3906	0.6416	0.5303	0.6478	0	0	0
General Medicine	106.1909	96.5455	52.2668	0.2873	1.1528	1.5126	0	0	0.0344 0
General Surgery	48.8745	46.5117	44.3624	0.4163	0.521	0.6794	0	0	0
GP Other	0	0	0	9.6797	10.1153	9.8254	0	0	0
Gynaecology	2.5425	1.8482	1.968	0.1391	0.1231	0.1046	0	0	0
Haematology	3.4524	2.6894	3.0736	0	0	0.0959	0	0	0
Infectious Diseases	8.1529	6.4296	6.9635	0	0	0	0	0	0
Intermediate Care	0	0	0	0	0	0	0	0.0078	1.0184
Maxillo-Facial	1.2256	1.0697	1.2539	0	0	0	0	0	0
Neurology	1.4497	1.565	1.7336	0	0	0	0	0	0
Ophthalmology	2.5831	2.4957	2.5828	0	0	0.0086	0	0	0
Pain	0.0555	0.066	0.0545	0	0.0118	0.0047	0	0	0
Plastic Surgery	0	0	0	0	0	0	0	0	0
Radiology	0.0102	0.0196	0.0142	0	0	0	0	0	0
Radiotherapy & Oncology	0.0031	0.0023	0.673	0 0	0	0	0	0	
Rehabilitation	64.5703	62.9902	61.0276	34.7587	34.72	29.427	69.5287	64.8741	72.5009
Respiratory	30.1523	30.7641	28.4902	0	0	0	0	0	0
Restorative Dentistry	0.0002	0	0.0002	0	0	0	0	0	0
Rheumatology	0	0	0.0011	0	0.0044	0	0	0	0
Trauma & Orthopaedic	57.1547	61.7714	62.0072	0.7173	0.6536	0.6288	0	0	0
Urology	11.4221	13.5649	12.033	0.057	0.0541	0.0455	0	0	0

Table 5.37: The daily bed demands for each specialty within regions one, two and three of ABUHB for three individual years' worth of patient admissions, rounded to four decimal places.

Specialty	**Region 4**			**Region 5**			**Region 6**		
	2017-2018	**2018-2019**	**2019-2020**	**2017-2018**	**2018-2019**	**2019-2020**	**2017-2018**	**2018-2019**	**2019-2020**
Accident & Emergency	0	0	0	8.5642	10.0826	8.9078	0	0	0
Anaesthetics	0	0	0	0.423	0.7812	1.1152	0	0	0
Cardiology	0	0	0	10.7899	10.2858	8.5707	0	0	0.0008
Care of the Elderly	12.3557	6.4882	7.3849	53.7813	44.1656	41.5028	0	0	0
Community Medicine	0.3653	0.1386	0.4321	16.6726	13.9982	8.2691	0	0	0
Dermatology	0	0	0	0	0	0	0	0	0
Diabetes & Endocrinology	0	0	0	17.8653	18.4154	15.4405	0	0	0
Ear, Nose & Throat	0	0	0	0	0	0	0	0	0
Gastroenterology	0	0	0	18.3652	21.7402	20.2937	0.0006	0	0.0008
General Medicine	0	0	0	12.0771	11.8051	18.6143	0	0	0
General Surgery	0	0	0	21.656	23.1851	20.8447	0.0003	0.0003	0.0008
GP Other 0	0	0	14.7579	13.86	17.2409	0	0	0	
Gynaecology	0	0	0	1.2598	1.5459	0.7105	0	0	0.0008
Haematology	0	0	0	1.8601	1.7505	1.8911	0.0003	0	0
Infectious Diseases	0	0	0	0	0	0	0	0	0
Intermediate Care	0	0.6014	0.3729	0	0	0	0	0	0
Maxillo-Facial	0	0	0	0.0271	0.0141	0.0319	0	0	0
Neurology	0	0	0	0	0	0	0	0	0
Ophthalmology	0	0	0	0.1748	0.2065	0.2093	0	0.0003	0.0012
Pain	0.0086	0.0008	0.0016	0.0145	0.0156	0.0094	0	0	0
Plastic Surgery	0	0	0	0.0212	0.0199	0.06	0	0	0.0008
Radiology	0	0	0	0.006	0.0003	0.0017	0	0	0
Radiotherapy & Oncology	0	0	0	0	0	0	0	0	0
Rehabilitation	29.5590	32.9368	33.4754	18.3286	25.2346	29.8199	0	0	0
Respiratory	0	0	0	30.7903	29.5149	23.2192	0	0	0
Restorative Dentistry	0	0	0	0	0	0	0	0	0
Rheumatology	0	0	0	0	0.0386	0	0	0	0
Trauma & Orthopaedic	0	0	0	42.5161	41.2805	40.7614	0	0	0
Urology	0	0	0	0.1685	0.1933	0.2485	0.0028	0.0027	0.0035

Table 5.38: The daily bed demands for each specialty within regions four, five and six of ABUHB for three individual years' worth of patient admissions, rounded to four decimal places.

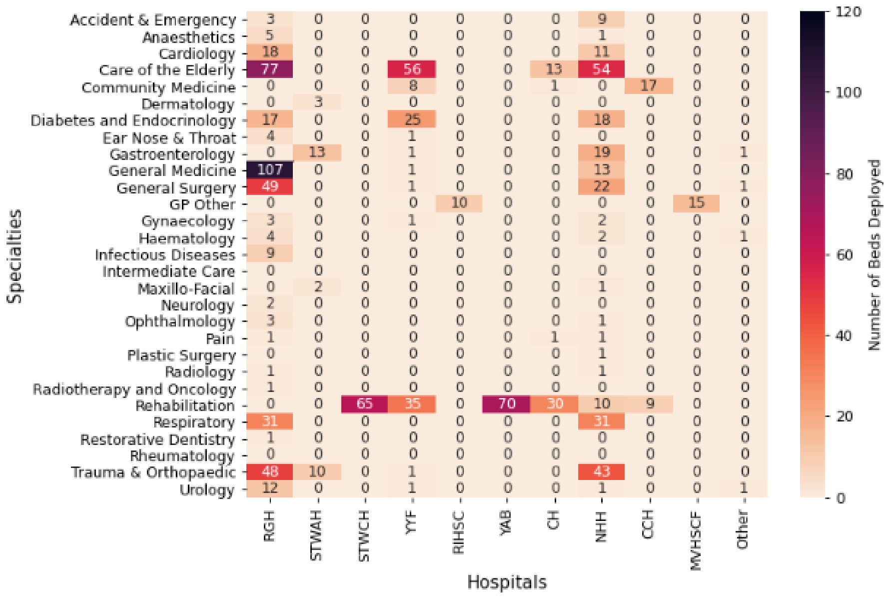
5.4.3.2.1 Deterministic Model

Using the average daily demands as the minimum number of beds that were required to be met, the deterministic model ran on each of the three years investigated. Table 5.39 displays the number of beds and staff required each year with the expected daily cost per year. The years 2017-2018, had the largest expected cost, planning the largest number of beds and staff. By planning year-by-year, this had the potential for up to £13,312.60 to be saved per day.

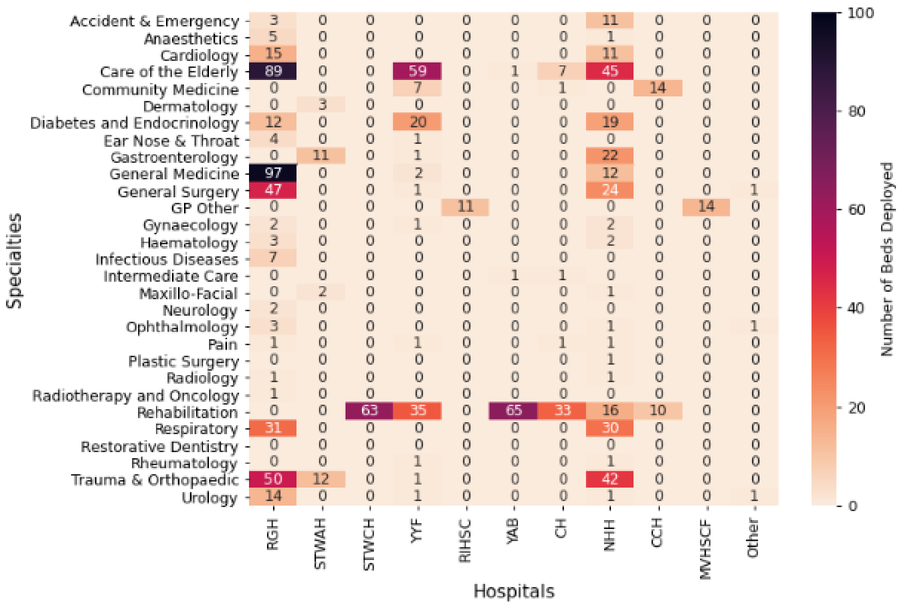
	Year	**Total Beds**		**Total Staff**		**Objective Function Value (£)**
		x^{bed}	u^{bed}	x^{staff}	u^{staff}	
	2017-2018	1,031	-	396	-	898,254.20 = EV_{17-18}
Deterministic	2018-2019	1,015	-	396	-	890,968.20 = EV_{18-19}
	2019-2020	1,010	-	396	-	893,712.20 = EV_{19-20}

Table 5.39: The EV values for the x^{bed}, u^{bed}, x^{staff} and u^{staff} decision variables and the objective function value per year for Experiment 2.

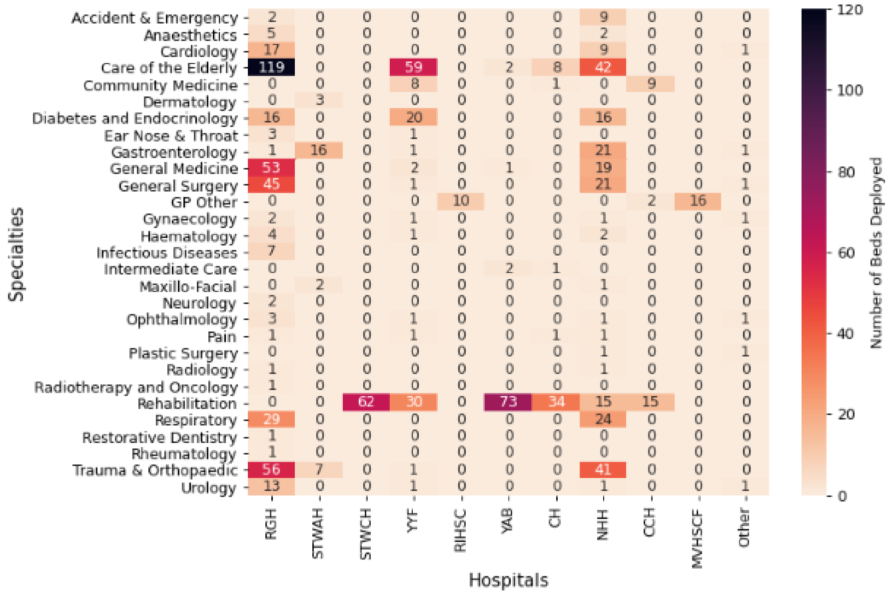
In order to visualise how these beds should be planned, Figure 5.10 displays each of the three heatmaps for hospital and specialty locations. For the majority of locations, if hospital beds are opened then in the following years the beds remain open. The results also display, the number of patients who are required to be transferred to other non NHS site locations or hospitals in other health boards through the “Other” category. Therefore if hospital managers wanted to limit the number of patients falling into the “Other’ category, they could determine how much additional demand they need to make available. In all three cases, a maximum of one additional bed was required per day for three to six specialties.



(a) 2017-2018



(b) 2018-2019



(c) 2019-2020

Figure 5.10: Heatmaps of bed locations for each specialty within each hospital for the deterministic model for the years 2017-2018 (a), 2018-2019 (b) and 2019-2020 (c), for Experiment 2. Note that a darker colour indicates a larger number of beds are deployed.

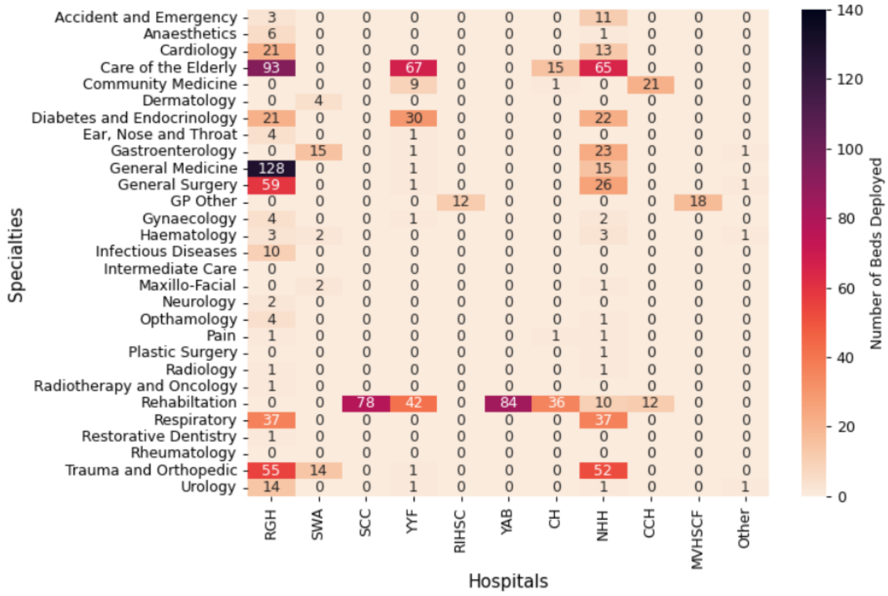
5.4.3.2.2 Two-Stage Stochastic Model

The two-stage stochastic model was considered with the same four scenarios as in Table 5.32. Table 5.40 shows that in the first stage, fewer beds and staff were deployed compared to the deterministic result for all three years. This in turn, increased the objective value ranging from 4.64% to 4.97%, with a maximum of 218 and 72 additional beds and staff deployed respectively.

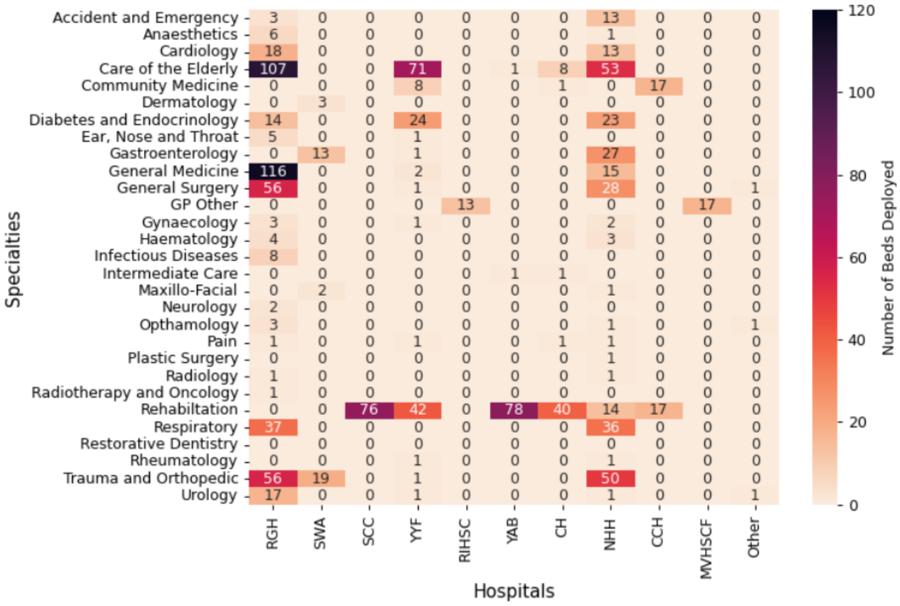
	Year	Total Beds		Total Staff		Objective Function Value (£)
		x^{bed}	u^{bed}	x^{staff}	u^{staff}	
Deterministic	2017-2018	1,031	-	396	-	898,254.20 = EV_{17-18}
	2018-2019	1,015	-	396	-	890,968.20 = EV_{18-19}
	2019-2020	1,010	-	396	-	893,712.20 = EV_{19-20}
Stochastic	2017-2018	849	379	326	142	941,764.64 = RP_{17-18}
	2018-2019	847	362	330	136	935,335.28 = RP_{18-19}
	2019-2020	852	350	336	134	935,171.84 = RP_{19-20}

Table 5.40: The EV and RP values for the x^{bed}, u^{bed}, x^{staff} and u^{staff} decision variables and the objective function value per year for Experiment 2.

Figure 5.11 presents the locations of bed deployment for each specialty in each year. The largest differences can be seen when planning beds for RGH in the COTE and general medicine wards. A total of 93 and 107 beds are each planned for COTE wards in 2017-2018 and 2018-2019 respectively, however, the daily demand increased

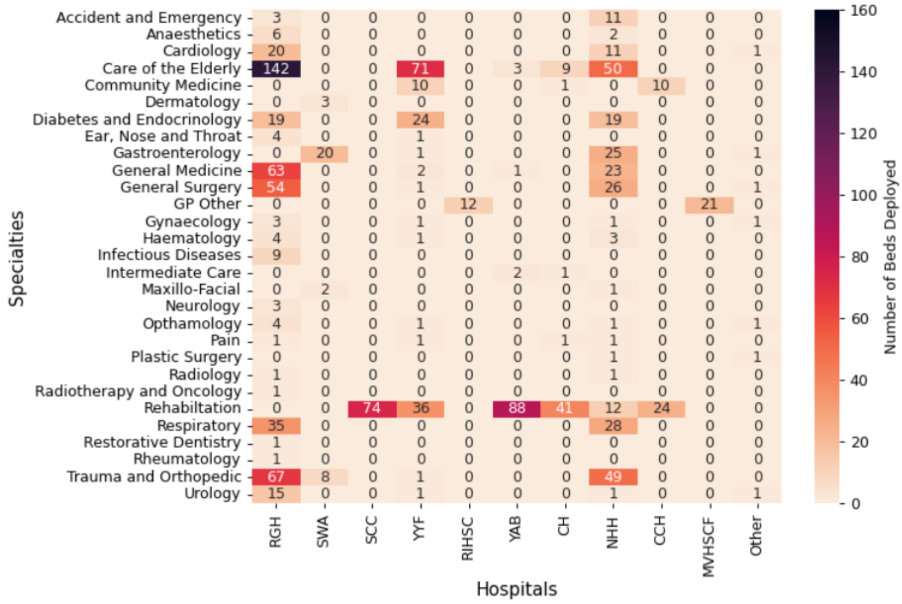


(a) 2017-2018



(b) 2018-2019

to 142 beds in the 2019-2020 period. Conversely, the general medicine daily bed requirement decreased from 128 and 116 in 2017-2018 and 2018-2019 respectively, to 63 in 2019-2020.



(c) 2019-2020

Figure 5.11: Heatmaps of bed locations for each specialty within each hospital for the two-stage stochastic model for the years 2017-2018 (a), 2018-2019 (b) and 2019-2020 (c), for Experiment 2. Note that a darker colour indicates a larger number of beds are deployed.

5.4.3.2.3 Test A

Test A calculates the VSS to determine the benefit of using the stochastic solution over the deterministic solution. Table 5.41 displays the EEV, for each year. By fixing the deterministic variables, the stochastic nature of healthcare can be realised and the model can determine the additional number of beds that would be required if the deterministic values were used.

	Year	Total Beds		Total Staff		Objective Function Value (£)
		x^{bed}	u^{bed}	x^{staff}	u^{staff}	
Deterministic	2017-2018	1,031	-	396	-	898,254.20 = EV_{17-18}
	2018-2019	1,015	-	396	-	890,968.20 = EV_{18-19}
	2019-2020	1,010	-	396	-	893,712.20 = EV_{19-20}
Stochastic	2017-2018	849	379	326	142	941,764.64 = RP_{17-18}
	2018-2019	847	362	330	136	935,335.28 = RP_{18-19}
	2019-2020	852	350	336	134	935,171.84 = RP_{19-20}
Test A	2017-2018	1,031	197	396	106	973,265.64 = EEV_{17-18}
	2018-2019	1,015	190	396	92	964,969.40 = EEV_{18-19}
	2019-2020	1,010	192	396	104	968,081.36 = EEV_{19-20}

Table 5.41: The EV, RP and EEV values for the x^{bed}, u^{bed}, x^{staff} and u^{staff} decision variables and the objective function value per year for Experiment 2.

The VSS ranges from between 3.17% to 3.52%, showing that the total cost is increasing by up to 3.52% by implementing the deterministic solution.

$$\text{VSS}_{17-18} = \text{EEV}_{17-18} - \text{RP}_{17-18} = \pounds 31,501.00\ (3.34\%) \tag{5.8}$$

$$\text{VSS}_{18-19} = \text{EEV}_{18-19} - \text{RP}_{18-19} = \pounds 29,634.12\ (3.17\%) \tag{5.9}$$

$$\text{VSS}_{19-20} = \text{EEV}_{19-20} - \text{RP}_{19-20} = \pounds 32,909.52\ (3.52\%) \tag{5.10}$$

There were a number of reasons why the deterministic solution was considered poor. The model may be too optimistic on the randomness leading to insufficient beds and staff being deployed or the model could be planning the wrong beds and staff. This could be determined by calculating the LUSS and LUDS values.

5.4.3.2.4 Test B

The ESSV was calculated by fixing the zero variables produced in the deterministic result and allowing the stochastic model to run. After performing the experiments, it was determined that the ESSV was equal to the EEV in all cases (Table 5.42).

	Year	**Total Beds**		**Total Staff**		**Objective Function Value (£)**
		x^{bed}	u^{bed}	x^{staff}	u^{staff}	
Deterministic	2017-2018	1,031	-	396	-	898,254.20 = EV_{17-18}
	2018-2019	1,015	-	396	-	890,968.20 = EV_{18-19}
	2019-2020	1,010	-	396	-	893,712.20 = EV_{19-20}
Stochastic	2017-2018	849	379	326	142	941,764.64 = RP_{17-18}
	2018-2019	847	362	330	136	935,335.28 = RP_{18-19}
	2019-2020	852	350	336	134	935,171.84 = RP_{19-20}
Test A	2017-2018	1,031	197	396	106	973,265.64 = EEV_{17-18}
	2018-2019	1,015	190	396	92	964,969.40 = EEV_{18-19}
	2019-2020	1,010	192	396	104	968,081.36 = EEV_{19-20}
Test B	2017-2018	849	379	326	142	941,764.64 = ESSV_{17-18}
	2018-2019	847	362	330	136	935,335.28 = ESSV_{18-19}
	2019-2020	852	355	336	134	935,171.84 = ESSV_{19-20}

Table 5.42: The EV, RP, EEV and ESSV values for the x^{bed}, u^{bed}, x^{staff} and u^{staff} decision variables and the objective function value per year for Experiment 2.

The LUSS was calculated to be £0 for all three cases. The LUSS was equal to the RP and therefore corresponded to the perfect skeleton solution. This suggested that the variables selected by the first stage variables of the solution are robust.

$$\text{LUSS}_{17-18} = \text{ESSV}_{17-18} - \text{RP}_{17-18} = \pounds 0 \tag{5.11}$$

$$\text{LUSS}_{18-19} = \text{ESSV}_{18-19} - \text{RP}_{18-19} = \pounds 0 \tag{5.12}$$

$$\text{LUSS}_{19-20} = \text{ESSV}_{19-20} - \text{RP}_{19-20} = \pounds 0 \tag{5.13}$$

5.4.3.2.5 Test C

The final test involved taking the decision variables determined by deterministic solution and adding this as a minimum constraint. These results can be seen within Table 5.43.

	Year	Total Beds		Total Staff		Objective Function Value (£)
		x^{bed}	u^{bed}	x^{staff}	u^{staff}	
Deterministic	2017-2018	1,031	-	396	-	898,254.20 = EV_{17-18}
	2018-2019	1,015	-	396	-	890,968.20 = EV_{18-19}
	2019-2020	1,010	-	396	-	893,712.20 = EV_{19-20}
Stochastic	2017-2018	849	379	326	142	941,764.64 = RP_{17-18}
	2018-2019	847	362	330	136	935,335.28 = RP_{18-19}
	2019-2020	852	350	336	134	935,171.84 = RP_{19-20}
Test A	2017-2018	1,031	197	396	106	973,265.64 = EEV_{17-18}
	2018-2019	1,015	190	396	92	964,969.40 = EEV_{18-19}
	2019-2020	1,010	192	396	104	968,081.36 = EEV_{19-20}
Test B	2017-2018	849	379	326	142	941,764.64 = ESSV_{17-18}
	2018-2019	847	362	330	136	935,335.28 = ESSV_{18-19}
	2019-2020	852	355	336	134	935,171.84 = ESSV_{19-20}
Test C	2017-2018	1,038	190	396	96	972,719.84 = EIV_{17-18}
	2018-2019	1,019	190	396	92	964,786.08 = EIV_{18-19}
	2019-2020	1,021	181	396	86	967,397.44 = EIV_{19-20}

Table 5.43: The EV, RP, EEV, ESSV and EIV values for the x^{bed}, u^{bed}, x^{staff} and u^{staff} decision variables and the objective function value per year for Experiment 2.

$$\text{LUDS}_{17-18} = \text{EIV}_{17-18} - \text{RP}_{17-18} = £31,006.24 \tag{5.14}$$

$$\text{LUDS}_{18-19} = \text{EIV}_{18-19} - \text{RP}_{18-19} = £29,450.80 \tag{5.15}$$

$$\text{LUDS}_{19-20} = \text{EIV}_{19-20} - \text{RP}_{19-20} = £32,225.60 \tag{5.16}$$

Therefore, $EEV - EV \geq VSS \geq LUDS \geq 0$ is satisfied. This result corresponds to partial upgradeability, where the deterministic results are upgraded in first stage results by the stochastic model.

5.4.3.2.6 Experiment 2 Summary

This section has analysed the bed and staffing requirements based on a year-to-year basis and determined where there was variation in the system. The largest VSS of 3.52% showed the benefit of using the two-stage stochastic model for planning in the first instance rather than using the deterministic implementation. Similar to the previous experiment, the deterministic solution did not perform well in a stochastic environment since insufficient beds and staff were deployed across all three years. This showed the deterministic model was too optimistic in terms of the demand.

5.4.3.3 Prescriptive Analytics Summary

Microsoft Excel OpenSolver and Python PuLP are two optimisation tools generated for this research, both of which have been able to be applied to generate the given results.

The results from the two experiments can be compared to determine whether ABUHB should plan on a year-to-year basis, or plan on a three-year horizon. Although the Public Health Scotland data are not accurate to ABUHB, the values given will still provide insight into the potential savings that could be made. Table 5.44 presents the VSS results per year and the total savings if implemented over the three years. The third year (2019-2020), was a leap year with 366 days and therefore impacts the difference in VSS. The findings indicate minimal variation in VSS, suggesting that planning on a year-to-year basis might not be necessary. Instead, it may be more beneficial to focus on longer-term horizons for planning purposes.

	Experiment 1	**Experiment 2**
Year 1	£11,664,027.60	£11,497,865.00
Year 2	£11,664,027.60	£10,816,453.80
Year 3	£11,695,983.84	£12,044,884.32
Total Saving	£35,024,039.04	£34,359,203.12

Table 5.44: The total yearly VSS values for each of the three years by experiment. Note that year 3 was a leap year and therefore contains 366 days which accounts for the variation between years.

This section has discussed the research aims, 'How best can specialties be organised among a network of hospitals to ensure staffing and bed costs are minimised, whilst still meeting the demand for frail and elderly patients?'. The heatmaps produced visualised the number of beds to locate across different specialties in order to minimise costs. The models also determined the number of staff to deploy based on the number of beds. These models work under the assumption of current hospital locations, where a hospital cannot open wards if they do not have the resources for these specialties. The models perform by deploying beds to the least expensive location first, determining the trade off between different combinations. Traditionally, ABUHB has planned on averages (deterministic solution), using historic data to determine the locality and quantity of these beds. By implementing a two-stage stochastic model, with different levels of demand, it provides savings for the NHS from approximately 3.52% per day. This saving has the potential to impact patient care by reallocating this money to additional resources, more staff training or improving community care schemes to reduce the pressures on hospitals. Whilst costing figures from ABUHB were not available, the model's utility and ability to provide cost savings have been demonstrated using NHS Scotland values [11].

5.5 Summary

This section has presented the findings of the predictive and prescriptive analytical models. Section 5.2 provided an overview of the current data and trends within ABUHB and within the frail and elderly community. Section 5.3 has considered the improved results by using CART models over traditional linear and logistic regression methods. These CART models have also enabled patient groupings of similar attributes to be generated, as shown in Appendices C.2 and C.3. Section 5.4 has applied the deterministic and two-stage stochastic models generated in Chapter 4 to ABUHB data determining how beds should be organised and staff deployed based on figures from Public Health Scotland.

Predictive and prescriptive models are increasingly being used in healthcare to improve patient outcomes and optimise resource utilisation. Predictive models can be used to identify patients who are at high risk of longer LOS and put in place appropriate interventions to reduce this. Prescriptive models can be used to determine capacity planning and staff requirements in order to reduce the likelihood of not having sufficient capacity. By applying predictive and prescriptive models, healthcare organisations can improve patient outcomes while also maximising the value of healthcare resources. The following chapter will discuss the ability to link the predictive and prescriptive methods together.

Chapter 6

Linking Predictive and Prescriptive Analytics for Healthcare Services

6.1 Introduction

Predictive and prescriptive modelling are two powerful techniques in OR that have the ability to extract insights and guide decision-makers. Predictive models are used to forecast future outcomes based on historical data and patterns, while prescriptive models provide recommendations on how to optimise those outcomes based on certain constraints and objectives. While both techniques are valuable in their own right, they become even more powerful when linked together. By integrating predictive and prescriptive models, organisations can predict future outcomes and also make informed decisions on how to optimise those outcomes in the most effective and efficient ways possible. This can lead to more accurate and impactful decision making, and ultimately improve business and healthcare performance. The concept of linking these two methods is still relatively novel, [3, 255], especially within the healthcare field and has great potential to drive significant value for organisations across a wide range of industries.

Research Aim - This Chapter aims to link the CART results with deterministic and two-stage stochastic models together to ensure the results are consistent. It will seek to present the results in order to address the following two research objectives:

1. Can linking predictive and prescriptive analytics provide improvements for decision making for frail and elderly services? - Section 6.2
2. How can deterministic and two-stage stochastic models be used to plan

hospital services for frail and elderly patients within Aneurin Bevan University Health Board? - Section 6.3

The remainder of the Chapter is structured as follows: Section 6.2 discusses linking the predictive and prescriptive paradigms together. Section 6.3 determines the robustness of the models by applying a number of different scenarios. Section 6.4 discusses the flexibility within the models which allows users to apply them to other healthcare situations.

6.2 Linking Paradigms

To investigate linking these two paradigms, two methods have been explored and applied to both the classification tree and regression tree results. The first method calculated the number of patients of each specialty and the overall average LOS for each end node. The second method used each end node and the specific LOS for each specialty and hospital within the node. These were then summed together to form the $D_{s,r}$ parameter. These two methods were run on both the regression and the classification trees, using the Microsoft Excel implementation. The results have been compared on a year-to-year basis, as well as the three year range. The VSS was calculated using the deterministic and two-stage stochastic models. For each example, a deterministic and two-stage stochastic heatmap has been provided within the Appendix D.

Using the predicted LOS from the CART models to work out demands can be more beneficial than simply using average demands due to several reasons. Firstly, predicted LOS accounts for individual patient characteristics and medical histories, allowing for more personalised estimates of resource demands, unlike average demands that treat all patients the same. Secondly, CART models can capture complex relationships between variables, resulting in more accurate predictions compared to simplistic average calculations that might overlook the impact of specific patient attributes on resource requirements. Moreover, predicted LOS adapts to changes in patient profiles and other factors affecting LOS, providing more up-to-date and flexible estimates. The model can handle outliers and extreme cases more effectively, ensuring robust capacity planning. By incorporating various patient features and clinical parameters, CART offers valuable data-driven insights into factors influencing resource demands, aiding healthcare providers in identifying areas for improvement.

6.2.1 Regression Tree and Average LOS

The first method utilised the regression tree generated in Section 5.3.3.1 (Figure 5.6). This tree generated 30 patient groupings determined by the 30 leaf nodes. The average LOS determined by each node was used to calculate the demand for each node. Using Equation (6.1), the average demand was calculated as follows:

$$D_{s,r} = \sum_{h \in r} D_{s,h} = \frac{\text{Number of Patients}_{s,h} * \text{Node Average LOS}}{\text{Total Number of Days in Data Set}} \tag{6.1}$$

The procedure for calculating the average bed demand for node two is shown in Table 6.1. For reference, in Figure 5.3.3.1, node two is the second left leaf node, and it indicates if a patient meets the 11 criteria listed below:

1. Admission method ≠ Other - transferred from another hospital
2. Admission method ≠ Elective - waiting list
3. Admission method ≠ Elective - booked
4. Specialty ≠ Accident & Emergency
5. Admission method ≠ Elective - planned
6. Hospital ≠ Ysbyty Ystrad Fawr
7. Age Group ≠ 65 - 69
8. Age Group ≠ 70 - 74
9. Specialty ≠ Trauma & Orthopaedic
10. Age Group ≠ 75 - 79
11. Specialty = Care Of The Elderly

For this node, there were three hospitals included, all of which were the COTE specialty accounting for 8,776 patients. The final column of Table 6.1 refers to the average daily bed demand across three years' worth of data. Each node's demands are consolidated for each specialty and hospital and then used for the overall demand, and are shown within Table D.1.

Hospital	**Specialty**	**Count**	**Average LOS**	**Average Daily Demand**
Nevill Hall Hospital	Care Of The Elderly	2696	11.393	28.0251168
Royal Gwent Hospital	Care Of The Elderly	6076	11.393	63.1604635
Ysbyty Aneurin Bevan	Care Of The Elderly	4	11.393	0.0415803

Table 6.1: The count of admissions and the associated average LOS for each hospital and specialty within the second node of the regression tree. The average daily bed demand has additionally been calculated.

Table 6.2 presents the results from the demands generated using the average LOS. These demands can be found within Table D.1 in the Appendix. The same four scenarios, as listed in Table 5.32, were applied to the demand figures. The VSS can be calculated to be 3.34% with a saving of £31,052.04. In comparison to Table 5.34, there was a difference in the deterministic solution of approximately 1.83%, with the regression tree deploying fewer numbers of beds and nurses.

	Total Beds		**Total Staff**		**Objective Function Value (£)**
	x^{bed}	u^{bed}	x^{staff}	u^{staff}	
Deterministic	997	-	406	-	887,845.20 = EV
Stochastic	836	359	342	130	929,725.40 = RP
Test A	997	192	406	106	960,777.44 = EEV

Table 6.2: The EV, RP and EEV values for the x^{bed}, u^{bed}, x^{staff}, u^{staff} decision variables and objective function using the regression tree and the average LOS across all three years.

Regression tree nodes were also used to calculate the year-to-year planning to see how the model performed annually. Equation (6.2) illustrates the process by which each annual demand was produced.

$$D_{s,r,year} = \sum_{h \in r} D_{s,h,year} = \frac{\text{Number of Patients}_{s,h} \times \text{Node Average LOS}_{year}}{\text{Number of Days in Year}} \tag{6.2}$$

Table 6.3 displays the EV, RP and EEV values for each year, with the demands given in Tables D.2 and D.3. The results reveal that the model employing the average LOS for the nodes predicted approximately the same number of beds and staff when compared to Table 5.43. The deterministic difference, which ranged from 0.65% to 1.13 demonstrated that the average LOS for all three years produced results that are comparable. The locations of the bed placements can be seen within Figures D.1 and D.2.

The third and final experiment has taken the average LOS for each end node and specialty, hospital and year combination. For the year 2019-2020, the total capacity had to be increased for the YAB (by 10%), meaning additional beds would be required from other age ranges in order to meet the demand. This was due to YAB's capacity being insufficient in the $\text{UB}_h^{\text{max, bed}}$ constraint. This would suggest the average LOS or the demand for the year 2019-2020, was larger for this hospital compared to previous years. Equation (6.3) displays the formulation of the demands, with the overall demands listed in Tables D.4 and D.5.

	Year	Total Beds		Total Staff		Objective Function Value (£)
		x^{bed}	u^{bed}	x^{staff}	u^{staff}	
Deterministic	2017-2018	1,021	-	386	-	891,125.20 = EV_{17-18}
	2018-2019	1,006	-	392	-	896,754.40 = EV_{18-19}
	2019-2020	997	-	392	-	883,629.40 = EV_{19-20}
Stochastic	2017-2018	852	368	326	140	935,847.92 = RP_{17-18}
	2018-2019	844	362	330	136	937,628.88 = RP_{18-19}
	2019-2020	851	340	334	128	926,433.28 = RP_{19-20}
Test A	2017-2018	1,021	192	386	104	965,491.36 = EEV_{17-18}
	2018-2019	1,006	191	392	104	969,817.36 = EEV_{18-19}
	2019-2020	997	189	392	100	957,284.20 = EEV_{19-20}

Table 6.3: The EV, RP and EEV values for the x^{bed}, u^{bed}, x^{staff}, u^{staff} decision variables and objective function using the regression tree and the yearly average LOS.

$$D_{s,r,year} = \sum_{h \in r} D_{s,h,year} = \frac{\text{Number of Patients}_{s,h} \times \text{Node Average LOS}_{s,h,year}}{\text{Number of Days in Year}} \tag{6.3}$$

The results listed in Table 6.4 have objective function values which were comparable to those in the initial experiment. The difference in the deterministic results differs by a range of 0.72% to 0.94%. This highlights the possibility that by utilising the yearly node average LOS, there is little difference between the results.

	Year	Total Beds		Total Staff		Objective Function Value (£)
		x^{bed}	u^{bed}	x^{staff}	u^{staff}	
Deterministic	2017-2018	986	-	380	-	897,414.00 = EV_{17-18}
	2018-2019	1,015	-	392	-	882,601.40 = EV_{18-19}
	2019-2020	1,024	-	400	-	900,162.00 = EV_{19-20}
Stochastic	2017-2018	826	356	316	138	893,038.64 = RP_{17-18}
	2018-2019	858	358	336	136	926,218.92 = RP_{18-19}
	2019-2020	871	358	342	132	945,905.68 = RP_{19-20}
Test A	2017-2018	986	193	380	106	920,396.24 = EEV_{17-18}
	2018-2019	1,015	194	392	106	955,761.64 = EEV_{18-19}
	2019-2020	1,024	198	400	106	976,584.64 = EEV_{19-20}

Table 6.4: The EV, RP and EEV values for the x^{bed}, u^{bed}, x^{staff}, u^{staff} decision variables and objective function using the regression tree and the yearly average LOS for each hospital and specialty.

6.2.2 Regression Tree and Specific LOS

The second method also utilised the regression tree shown in Figure 5.6. Instead of utilising the average LOS for each of the 30 end nodes, the specific LOS for each hospital and specialty inside that node was calculated. Each of the demands' generation processes are shown in Equation (6.4).

$$D_{s,r} = \sum_{h \in r} D_{s,h} = \frac{\text{Number of Patients}_{s,h} \times \text{Specific LOS}_{s,h}}{1096} \tag{6.4}$$

Table 6.5 presents the second node within the regression tree and determines how each of the demands was produced within this node. These findings demonstrated that employing particular hospital and specialty LOS, increased the demand for beds overall in RGH by one bed when compared to Table 6.1. The generated demands can be viewed in Table D.6.

Hospital	**Specialty**	**Count**	**Specific LOS**	**Average Daily Demand**
Nevill Hall Hospital	Care Of The Elderly	2696	11.412	28.0729927
Royal Gwent Hospital	Care Of The Elderly	6076	11.554	64.0510949
Ysbyty Aneurin Bevan	Care Of The Elderly	4	6.250	0.0228102

Table 6.5: The count of admissions and the associated specific LOS for each hospital and specialty within the second node of the regression tree. The average daily bed demand has additionally been calculated.

The findings for the deterministic and two-stage stochastic models are shown in Table 6.6, with the EEV also being calculated to determine the VSS.

	Total Beds		**Total Staff**		**Objective Function Value (£)**
	x^{bed}	u^{bed}	x^{staff}	u^{staff}	
Deterministic	1,011	-	388	-	889,242.60 = EV
Stochastic	842	361	320	134	925,599.36 = RP
Test A	1,011	186	579	94	958,630.76 = EEV

Table 6.6: The EV, RP and EEV values for the x^{bed}, u^{bed}, x^{staff}, u^{staff} decision variables and objective function using the regression tree and the specific LOS across all three years.

The specific LOS model had lower deterministic and two-stage stochastic objective values as compared to the regression tree with average LOS findings. This shows that if the exact LOS was used, rather than node averages, additional cost savings would be produced. The VSS produced a saving of £33,031.40 per day (3.57%). The specific LOS generated from each regression tree node could be used to analyse these results on an annual basis (Tables D.7 and D.8). To calculate the demands for each specialty and region for each year, Equation (6.5) could be applied.

$$D_{s,r,year} = \sum_{h \in r} D_{s,h,year} = \frac{\text{Number of Patients}_{s,h} \times \text{Specific LOS}_{s,h,year}}{\text{Number of Days in Year}} \tag{6.5}$$

Table 6.14 presents the results after the tool had optimised the bed and staffing numbers based on the demand values. The number of staff deployed in the first two years remained constant before reducing in the third year. Each year saw a reduction in the EV overall. The VSS ranged from 3.28% to 3.45%, once more demonstrating the advantages of employing the stochastic approach.

	Year	**Total Beds**		**Total Staff**		**Objective Function Value (£)**
		x^{bed}	u^{bed}	x^{staff}	u^{staff}	
Deterministic	2017-2018	1,009	-	370	-	874,693.00 = EV_{17-18}
	2018-2019	998	-	370	-	873,709.00 = EV_{18-19}
	2019-2020	988	-	364	-	869,959.80 = EV_{19-20}
Stochastic	2017-2018	839	369	310	132	917,423.68 = RP_{17-18}
	2018-2019	832	363	308	132	914,315.56 = RP_{18-19}
	2019-2020	838	344	310	130	911,420.56 = RP_{19-20}
Test A	2017-2018	1,009	194	370	100	947,517.00 = EEV_{17-18}
	2018-2019	998	189	370	100	945,583.80 = EEV_{18-19}
	2019-2020	988	189	364	100	942,856.60 = EEV_{19-20}

Table 6.7: The EV, RP and EEV values for the x^{bed}, u^{bed}, x^{staff}, u^{staff} decision variables and objective function using the regression tree and the yearly specific LOS.

6.2.3 Classification Tree and Average LOS

The third method utilised the classification tree displayed by Figure 5.7. The classification tree yielded 30 patient groupings, with patients falling into one of two categories. Recall Equation (6.1) to create demands and to calculate the $\text{D}_{s,r}$ variable:

$$D_{s,r} = \sum_{h \in r} D_{s,h} = \frac{\text{Number of Patients}_{s,h} \times \text{Node Average LOS}}{1096} \tag{6.1 revisited}$$

For patients who fell into a '<1' node, meant the majority of patients were discharged on the same day they were admitted. In spite of this, there were certain patients in every case who were put in this category despite not meeting the criteria. As a result, to take these individuals into consideration, the average LOS would not be zero days.

If a patient fulfilled all seven of the following criteria, they were grouped into the ninth node of the CART tree.

1. Admission method = Elective - waiting list
2. Specialty ≠ Trauma & Orthopaedic
3. Specialty ≠ General Surgery
4. Specialty ≠ Urology
5. Specialty ≠ Ear, Nose & Throat
6. Specialty ≠ Gynaecology
7. Specialty = Respiratory

For this node, the majority of patients were grouped into the '<1' category, and the average LOS was less than one (0.913918 days). The average daily demand required for each specialty and hospital inside this node may subsequently be determined using this value (Table 6.8). The associated demands are presented in Table D.9, within the Appendix.

Hospital	Specialty	Count	Average LOS	Average Daily Demand
Nevill Hall Hospital	Respiratory	690	0.913918	0.575368
Royal Gwent Hospital	Respiratory	553	0.913918	0.461128

Table 6.8: The count of admissions and the associated average LOS for each hospital and specialty within the ninth node of the classification tree. The average daily bed demand has additionally been calculated.

The results using these demands are shown in Table 6.9. By deploying 1,015 beds and 428 nurses, an EV value of £826,712.60 was produced. Similar to prior findings, fewer beds were used than the averages produced in Chapter 5. A significant reduction in the total objective function resulted from the deployment of beds to different hospital locations. The location of these beds can be seen in Figure D.37 in the Appendix.

	Total Beds		**Total Staff**		**Objective Function Value (£)**
	x^{bed}	u^{bed}	x^{staff}	u^{staff}	
Deterministic	1,015	-	428	-	826,712.60 = EV
Stochastic	862	352	360	138	866,576.52 = RP
Test A	1,015	190	428	102	892,880.68 = EEV

Table 6.9: The EV, RP and EEV values for the x^{bed}, u^{bed}, x^{staff}, u^{staff} decision variables and objective function using the classification tree and the average LOS across all three years.

The VSS was calculated to be 3.04%, demonstrating that there is a difference between the two models using this strategy even when classification trees are used to generate the demand. In comparison to the earlier models, the results used a similar number of beds, but the objective values were much lower. One explanation for this would be that fewer beds were deployed in the more expensive units because their LOS's were shorter and their daily demands are consequently lower.

Equation (6.2) described how each of the demands was calculated after further analysis on a year-to-year basis. The formulated demand was inputted into the deterministic and two-stage stochastic models, (Tables D.10 and D.11) after being summed up across each node.

$$D_{s,r,year} = \sum_{h \in r} D_{s,h,year} = \frac{\text{Number of Patients}_{s,h} \times \text{Node Average LOS}_{year}}{\text{Number of Days in Year}} \tag{6.2 revisited}$$

The EV, RP, and EEV values are shown in Table 6.10, demonstrating how the objective value fluctuates from year-to-year. Although using a comparable number of beds as in the prior experiment, the objective values obtained are lower. Figures D.25 - D.30 show where these beds are located. The two-stage stochastic model's advantage is demonstrated by the VSS, which varies from 3.11% to 3.35%. These findings demonstrate the advantages of yearly planning as opposed to preparing in three year increments.

	Year	**Total Beds**		**Total Staff**		**Objective Function Value (£)**
		x^{bed}	u^{bed}	x^{staff}	u^{staff}	
Deterministic	2017-2018	979	-	390	-	805,452.00 = EV_{17-18}
	2018-2019	1,002	-	398	-	834,264.60 = EV_{18-19}
	2019-2020	1,005	-	404	-	842,867.80 = EV_{19-20}
Stochastic	2017-2018	819	350	324	136	843,654.48 = RP_{17-18}
	2018-2019	844	359	338	140	875,908.56 = RP_{18-19}
	2019-2020	842	363	346	134	882,667.16 = RP_{19-20}
Test A	2017-2018	979	184	390	92	869,933.28 = EEV_{17-18}
	2018-2019	1,002	193	398	106	904,084.84 = EEV_{18-19}
	2019-2020	1,005	194	404	96	912,268.44 = EEV_{19-20}

Table 6.10: The EV, RP and EEV values for the x^{bed}, u^{bed}, x^{staff}, u^{staff} decision variables and objective function using the classification tree and the yearly average LOS.

Instead of utilising the node average for the model, these results were expanded to include the average LOS per year. This applied Equation (6.3) to calculate the demands for the model and the results can be seen in Tables D.12 and D.13. By employing this technique, it ensured that specialty LOS was taken into consideration, particularly for specialties which have longer LOS's.

$$D_{s,r,year} = \sum_{h \in r} D_{s,h,year} = \frac{\text{Number of Patients}_{s,h} \times \text{Node Average LOS}_{s,h,year}}{\text{Number of Days in Year}} \tag{6.3 revisited}$$

The results are presented within Table 6.11, and graphically illustrated in Figures D.31 - D.36 in the Appendix. The results show that more beds and nurses were deployed in 2017–2018 when compared against using the average node LOS. The objective value was lower from 2018, as fewer beds and nurses were being deployed.

	Year	**Total Beds**		**Total Staff**		**Objective Function Value (£)**
		x^{bed}	u^{bed}	x^{staff}	u^{staff}	
Deterministic	2017-2018	993	-	398	-	$815{,}342.60 = \text{EV}_{17-18}$
	2018-2019	993	-	398	-	$829{,}485.60 = \text{EV}_{18-19}$
	2019-2020	995	-	400	-	$837{,}880.00 = \text{EV}_{19-20}$
Stochastic	2017-2018	832	359	330	140	$856{,}341.72 = \text{RP}_{17-18}$
	2018-2019	838	352	338	140	$870{,}815.16 = \text{RP}_{18-19}$
	2019-2020	830	363	340	140	$878{,}140.24 = \text{RP}_{19-20}$
Test A	2017-2018	993	192	398	100	$883{,}441 = \text{EEV}_{17-18}$
	2018-2019	993	191	398	104	$899{,}117.12 = \text{EEV}_{17-18}$
	2019-2020	995	192	400	102	$907{,}630.08 = \text{EEV}_{19-20}$

Table 6.11: The EV, RP and EEV values for the x^{bed}, u^{bed}, x^{staff}, u^{staff} decision variables and objective function using the classification tree and the yearly average LOS for each hospital and specialty.

6.2.4 Classification Tree and Specific LOS

The classification tree depicted in Figure 5.7 was used by the fourth and final method. Instead of utilising the average LOS for each of the 30 end nodes, the LOS was calculated for each hospital and specialty inside that node. The generation of each demand is shown in Equation (6.4).

$$D_{s,r} = \sum_{h \in r} D_{s,h} = \frac{\text{Number of Patients}_{s,h} \times \text{Specific LOS}_{s,h}}{1096} \tag{6.4 revisited}$$

Table 6.12 presents the ninth node within the classification tree and determines how each of the demands was generated within this node. These results in comparison with Table 6.8, have shown that using specific hospital and specialty, LOS overall did not increase the number of beds required in this node. The generated demands can be viewed in Table D.14. At the deterministic stage, the required number of beds decreased across all nodes from 1,015 beds to 1,011 beds (Table 6.13).

Hospital	Specialty	Count	Average LOS	Average Daily Demand
Nevill Hall Hospital	Respiratory	690	0.531884	0.3348540
Royal Gwent Hospital	Respiratory	553	1.39057	0.7016423

Table 6.12: The count of admissions and the associated specific LOS for each hospital and specialty within the ninth node of the classification tree. The average daily bed demand has additionally been calculated.

Table 6.13 presents the results for the deterministic and two-stage stochastic model, with the EEV also being calculated to determine the VSS.

	Total Beds		**Total Staff**		**Objective Function Value (£)**
	x^{bed}	u^{bed}	x^{staff}	u^{staff}	
Deterministic	1,011	-	388	-	889,242.60 = EV
Stochastic	842	361	320	134	925,599.36 = RP
Test A	1,011	186	388	94	958,624.36 = EEV

Table 6.13: The EV, RP and EEV values for the x^{bed}, u^{bed}, x^{staff}, u^{staff} decision variables and objective function using the classification tree and the specific LOS across all three years.

Comparing to the regression tree with average LOS results, the deterministic and two-stage stochastic objective values were higher in the specific LOS model. This suggested that using node averages might not produce sufficient capacity for beds and staff. The VSS produced a saving of £21,344.72 per day (2.09%).

The individual LOS generated from each regression tree node could be used to analyse these results on an annual basis (Tables D.15 and D.16). Equation (6.5) was used to determine the demands for each specialty and region for each year.

$$D_{s,r,year} = \sum_{h\in r} D_{s,h,year} = \frac{\text{Number of Patients}_{s,h} \times \text{Specific LOS}_{s,h,year}}{\text{Number of Days in Year}} \tag{6.5 revisited}$$

The results have shown that the model optimised the bed and staff numbers based on the demand data, as shown in Table 6.14. Over time the number of beds and nurses deployed has declined. Although more beds are deployed in the first year, the second and third years have larger objective values due to different specialty beds being deployed.

	Year	Total Beds x^{bed}	Total Beds u^{bed}	Total Staff x^{staff}	Total Staff u^{staff}	Objective Function Value (£)
Deterministic	2017-2018	1,007	-	388	-	869,740.60 = EV_{17-18}
	2018-2019	1,004	-	386	-	875,147.20 = EV_{18-19}
	2019-2020	1,000	-	380	-	877,707.00 = EV_{19-20}
Stochastic	2017-2018	836	367	318	136	911,555.20 = RP_{17-18}
	2018-2019	839	367	320	138	917,521.08 = RP_{18-19}
	2019-2020	848	351	328	136	922,334.40 = RP_{19-20}
Test A	2017-2018	1,007	193	388	100	941,426.60 = EEV_{17-18}
	2018-2019	1,004	192	386	106	948,915.04 = EEV_{18-19}
	2019-2020	1,000	193	380	104	952,017.56 = EEV_{19-20}

Table 6.14: The EV, RP and EEV values for the x^{bed}, u^{bed}, x^{staff}, u^{staff} decision variables and objective function using the classification tree and the yearly specific LOS.

6.3 Scenario Analysis

Scenario analysis is a powerful tool used in strategic planning and decision making. It involves developing and examining a variety of hypothetical future scenarios in order to understand their potential effects on a given situation or system. This approach allows decision-makers to investigate a range of possible outcomes and uncertainties, which can be used to develop their strategies and plans. Scenario analysis typically involves identifying the key drivers of change and uncertainty in a particular situation or system. These drivers can include economic trends, technological developments, political factors, and social changes. Following the identification of these drivers, various scenarios are created by taking into account how they could interact and change over time. The resulting scenarios provide decision-makers with a range of plausible futures to consider, each with its own set of opportunities and challenges. By exploring these various scenarios, decision-makers can better understand the risks and opportunities associated with different strategies and plans, and make more informed decisions about how to proceed.

This section will utilise the CART models to feed into the deterministic and two-stage stochastic models. The underlying assumption with predictive models is that the patterns observed in the historical data will continue to hold in the future under similar circumstances. This might be particularly challenging in healthcare, particularly for predicting LOS in hospitals. Moreover, historical data used for training the predictive models may include instances of poor system performance or inefficiencies that could impact the generalisability of the model's predictions to future scenarios.

When using a point estimate (i.e. average) for predicting hospital LOS, typically the mean LOS is calculated across the entire data set. This single average value is then used as the prediction for all patients, regardless of their individual characteristics. Consequently, this approach does not account for the diversity of patients, medical conditions, and other factors that contribute to variations. By utilising CART, patient subgroups based on characteristics are created and subgroup specific average LOS values for predictions are created. This approach recognises the diversity in patient profiles and medical conditions, providing more variation in the demand values.

This section aims to utilise the methods of linking the predictive and prescriptive models and apply a variety of scenarios to them. Specific scenarios in ABUHB have been identified through collaboration with senior staff within the health board. The health board raised four main concerns regarding future changes and how this may impact bed and staffing requirements. These were as follows:

1. Addition of a new hospital
2. What if demand cannot be met?
3. Re-evaluating the current setup
4. Long-term predictions

The remainder of this section will discuss each of the above points.

6.3.1 Addition of a New Hospital

The Grange University Hospital (GUH), a new hospital with a focus on critical care, opened within the health board in November 2020 with the goal of treating the most seriously ill patients or those with significant injuries. It also serves as the designated trauma centre for the area [288]. The hospital opened with 560 beds and features a 24 hours acute assessment unit, A&E unit and provides 24/7 emergency care for patients that need specialist and critical care. To help alleviate the strains brought on by the second wave of Covid-19 and winter seasonal stresses, GUH opened earlier than its planned date of March 2021. The hospital was designed to treat patients who cannot be safely managed at one of the local general hospitals, and as such required specialties to be relocated throughout the health board. As of 2022, GUH catered for 18 specialties and as a result the number of specialties offered by other hospitals was reduced. In total, the number of specialties locations offered by the hospitals reduced from 98 to 92. An updated visual of specialty and hospital locations can be seen within Figure 6.1.

In order to incorporate GUH into the model with the existing data, the assumption has been made that patients will be admitted to any hospital within the health board

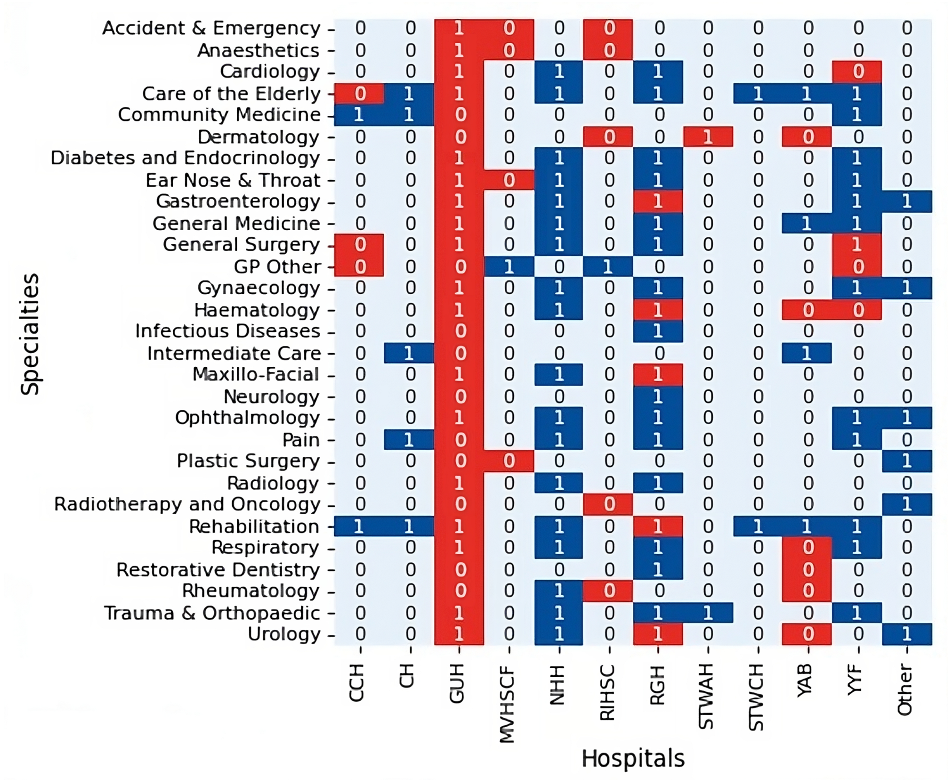


Figure 6.1: An updated version of the hospitals and specialties in ABUHB, with a '1' indicating a specialty is present in a given hospital. For cells with a red background, displays where specialties have opened or closed since the opening of the Grange University Hospital (GUH).

and the regional restrictions are lifted. This results in the following constraints, where the $D_{s,r}$ parameter is changed to D_s:

Deterministic Model

$$\sum_{h \in \mathcal{H}} x^{\text{bed}}_{s,h} \geq D_s \quad \forall s \in \mathcal{S} \tag{6.6}$$

Two-Stage Stochastic Model

$$\sum_{h \in \mathcal{H}} x^{\text{bed}}_{s,h} + \sum_{h \in \mathcal{H}} u^{\text{bed}}_{s,h,k} \geq D_{s,k} \quad \forall s \in \mathcal{S}, k \in \mathcal{K} \tag{6.7}$$

The cost for each Welsh specialty bed was calculated using open source data from Public Health Scotland. The specialty average for the Welsh-generated data was the same as the specialty average for the Scottish-generated data, and the Welsh data also fell within the specified range of the Scottish-generated data. In order to match the average of the Scottish data, new values had to be constructed because the numbers were originally created using the prior hospital/specialty location possibilities without taking GUH into account. If a specialty remains in the same hospital, the previous cost will remain the same. Nonetheless, in some circumstances, hospitals no longer offer as many specialties as they once did, falling short of the Scottish average. As a result, these values are altered to conform to the average. Similarly, since the opening of GUH, the number of available beds within existing hospitals has decreased. GUH has 560 beds within the hospital available to all patients, however, the number has been scaled down to 300 to take the age group into consideration. Because fewer beds are available, this has an impact on the $\mathrm{K}_{s,h}$, $\mathrm{UB}_h^{\text{max, bed, 1st}}$ and $\mathrm{UB}_h^{\text{max, bed, 2nd}}$ variables.

To demonstrate how the model would perform with the addition of GUH and the relaxation of the regional demand constraint, the regression tree with the specific LOS over three years' worth of data will be utilised (Table D.6 within the Appendix). The model utilised the three scenarios previously discussed, where demand remains constant, increases by 20% and decreases by 20%, all with equal probabilities of a third. The deterministic model yielded an objective value of £669,699.20 with the deployment of 982 beds and 326 nursing staff (Table 6.15). In the case of the two-stage stochastic model, this amount rises to £686,198.04. These findings demonstrate that establishing GUH and redesigning the specialties within hospitals resulted in a difference of approximately £200,000.00. This shows the benefit to decision-makers of opening this hospital, and the potential savings if the ABUHB costings resembled those of the NHS Scotland data. The VSS is calculated to be 5.20% demonstrating the benefits of utilising the two-stage stochastic model over the traditional deterministic model.

	Total Beds		**Total Staff**		**Objective Function Value (£)**
	x^{bed}	u^{bed}	x^{staff}	u^{staff}	
Deterministic	982	-	326	-	£669,699.20 = EV
Stochastic	817	399	268	120	£686,198.04 = RP
Test A	982	188	326	76	£721,851.04 = EEV

Table 6.15: The EV, RP and EEV values for the x^{bed}, u^{bed}, x^{staff}, u^{staff} decision variables and objective function for Scenario 1 where the hospital GUH is added.

6.3.2 M-Penalty

The models so far have only considered hard constraints. Hard constraints are where constraints must be satisfied by any feasible solution to produce an optimal solution. However, in reality, if there is not sufficient capacity, patients cannot be admitted into hospitals and are either treated at home or transferred to a neighbouring health board for treatment. This situation would then result in an additional cost. The penalty can be incorporated into the existing constraints by the addition of the decision variable z, where $z \in \mathbb{N}$. In order to account for this within the objective function, a cost term of M is added, where M is a fixed cost regardless of hospital or specialty. If this was to be made more specific to the specialty or hospital, the subscripts $_{s,h}$ could be added and a two-dimensional variable generated. The objective function value and first constraint are thus formulated as follows.

Deterministic Model

$$\min \sum_{h\in\mathcal{H}} \sum_{s\in\mathcal{S}} (x^{\text{bed}}_{s,h} c^{\text{bed}}_{s,h} + \sum_{b\in\mathcal{B}} x^{\text{staff}}_{s,b,h} c^{\text{staff}}_{b}) + Mz \tag{6.8}$$

$$\sum_{h\in\mathcal{H}} x^{\text{bed}}_{s,h} \geq D_s + z \quad \forall s \in \mathcal{S} \tag{6.9}$$

Two-Stage Stochastic Model

$$\begin{aligned} \min \sum_{h\in\mathcal{H}} \sum_{s\in\mathcal{S}} (x^{\text{bed}}_{s,h} c^{\text{bed, 1st}}_{s,h} + \sum_{b\in\mathcal{B}} x^{\text{staff}}_{s,b,h} c^{\text{staff, 1st}}_{b})+ \\ \sum_{k\in\mathcal{K}} \sum_{h\in\mathcal{H}} \sum_{s\in\mathcal{S}} p_k (u^{\text{bed}}_{s,h,k} c^{\text{bed, 2nd}}_{s,h} + \sum_{b\in\mathcal{B}} u^{\text{staff}}_{s,b,k,h} c^{\text{staff, 2nd}}_{b}) + Mz \end{aligned} \tag{6.10}$$

$$\sum_{h\in\mathcal{H}} x^{\text{bed}}_{s,h} + \sum_{h\in\mathcal{H}} u^{\text{bed}}_{s,h,k} \geq D_{s,k} + z \quad \forall s \in \mathcal{S}, k \in \mathcal{K} \tag{6.11}$$

These new objective functions and constraints can be inputted into the OpenSolver model and solved with the regression tree nodes as demands. Since hard constraints have been used previously, the model will produce the same results as those in Section 6.3.1 and the objective function values shown in Table 6.15. However, if hospital beds are reallocated to other patient age groups within the hospital, or decision-makers decide to close hospitals, or specialties within certain hospitals, testing will be necessary to ascertain the effects on overall costs and resource requirements.

If decision-makers decided to reduce the number of available beds within STWAH to frail and elderly patients, then this would cause the services of dermatology and T&O to close within this hospital. Although T&O services can be transferred to RGH, GUH, NHH or YYF, no other hospital offers dermatology treatments. Therefore if we make the assumption that dermatology patients would have to be treated at home or at a different health board, this would be with an additional cost of M.

If we define M as having a value of £2,500.00, this value exceeds all second stage hospital costs throughout the health board. It was assumed that M is not scenario dependant meaning that if in one scenario, a penalty occurred, it occurred for all scenarios. Table 6.16 presents the results of the deterministic and two-stage stochastic models. The health board received a deterministic outcome that allocated 982 beds and 326 nurses, yielding an EV of £669,699.20, which is equal to the previous example in Table 6.15. In this case, demand could always be met, and therefore $z = 0$. The two-stage stochastic model involves the deployment of 816 beds in the first stage and a maximum of 407 beds in the second stage. This increases the objective value to £706,437.68, an increase of 2.91% without using the M-penalty method. The VSS can be calculated to be 2.19% with an additional saving of £15,499.76 per day by using the stochastic solution over the deterministic.

	Total Beds		**Total Staff**		**Objective Function Value (£)**
	x^{bed}	u^{bed}	x^{staff}	u^{staff}	
Deterministic	982	-	326	-	£669,699.20 = EV
Stochastic	816	407	268	120	£706,437.68 = RP
Test A	982	188	326	76	£721,937.44 = EEV

Table 6.16: The EV, RP and EEV values for the x^{bed}, u^{bed}, x^{staff}, u^{staff} decision variables and objective function for Scenario 2 where the M-penalty method is added.

The greatest effects of the M-penalty can be seen within Section 6.3.4, where the demands have been increased significantly to simulate the effects of a pandemic similar to those of Covid-19 to determine the robustness of the healthcare system.

6.3.3 Re-evaluating the Current Setup

This section aims to re-evaluate the current arrangement of specialties bed sites in the health board and to ascertain the most effective method of specialty reorganisation. This operates under the assumption that a patient can be admitted to any hospital within the health board and that every hospital has the capability of having any specialty. Since the opening of GUH in 2020, specialties have been rearranged around the health board. This work will determine the most efficient way to organise beds and nursing staff in order to meet the current demand.

The models will operate under the premise that patients can be admitted into any local hospital (Section 6.3.1) and that there is a penalty (Section 6.3.2) if demand is not satisfied throughout the health board.

Previously, specialty bed costs from Public Health Scotland have been utilised for the models (Table 5.29). As this scenario allows all hospitals to have all specialties, the cost matrix for first and second stage beds requires modifying to incorporate this. The prior costings will be recalculated using the new values for each hospital and specialty that fall within the specified ranges, with the average being set at the average of the Scottish data. If a hospital already has a specialty, the cost was applied as in the preceding instance. Although the financial figures are not directly correlated to ABUHB, this scenario test is still beneficial in terms of theory as to how the health board could go and implement this method. The results can then be compared to Table 6.16 as to the difference in costings.

Table 6.17 displays the final figures once the model has been executed. Within this model, 928 beds are deployed with 382 nursing staff required. In the majority of cases, specialties are localised to one or two hospitals (Figure D.51). To deploy the specialty of T&O, three hospitals are required and for the rehabilitation specialty, four hospitals are required. There is no overlap between these two specialties, therefore if a hospital has a T&O ward, it would not have a rehabilitation ward. The VSS is calculated to be 7%, showing once again, the benefit of utilising the stochastic model over the deterministic model. Although the financial figures may not be accurate, it can be recommended to ABUHB that further cost savings could be made if they are able to consolidate their specialties into one or two hospitals rather than providing a large number of specialties per hospital increasing resource costs.

	Total Beds		**Total Staff**		**Objective Function Value (£)**
	x^{bed}	u^{bed}	x^{staff}	u^{staff}	
Deterministic	982	-	328	-	£552,898.60 = EV
Stochastic	809	403	270	118	£555,746.48 = RP
Test A	982	188	328	76	£594,651.64 = EEV

Table 6.17: The EV, RP and EEV values for the x^{bed}, u^{bed}, x^{staff}, u^{staff} decision variables and objective function for Scenario 3 where the hospital setup is re-evaluated.

6.3.4 Long-term Planning

Long-term planning decisions in healthcare are critical in determining how demand will fluctuate and change in the future. There are four main reasons as to why long-term planning is essential:

1. Anticipation of future needs: Long-term planning helps healthcare organisations to anticipate future needs and plan accordingly. Healthcare providers can plan to extend certain services that are suited to the requirements of the elderly, such as COTE if an area is experiencing an ageing population.

2. Financial stability: The NHS has a limited spending budget per year assigned by the Government. By planning for future demand and capacity, decision-makers can ensure they have the resources required, without overspending, to continue to provide quality care.

3. Improved patient outcomes: Long-term planning can allow healthcare organisations to focus on preventative measures and early intervention. By antici-

pating future healthcare needs, organisations can develop strategies to address them proactively, leading to better health outcomes for patients.

4. Resource allocation: Long-term planning allows healthcare organisations to allocate resources effectively. With this planning, decision-makers can make informed decisions about where to invest resources, such as building new healthcare facilities, hiring additional staff or purchasing new equipment.

These reasons highlight how critical it is for the NHS to be able to adapt to change. The demand and pressures on the NHS are expected to increase over future years due to rising populations [289], Covid-19 recovery backlog [290] and lifestyle factors that lead to increases in hospital admissions [291].

6.3.4.1 Number of Available Nursing Staff

The impact of a reduction in the number of nursing staff available is examined in the following scenario. There are several reasons why this might happen, including nurses leaving the profession after the Covid-19 pandemic [292], nurses taking industrial action for fairer pay and working conditions [293] or due to sickness [294].

If we make the assumption the number of nursing staff available is reduced to 160 for each band in the first stage, and an additional 40 available within the second stage, the EV produced a value of £673,093.00 (Table 6.18). Due to the reduced availability of staff, it caused the z decision variable to be equal to three, as wards could not open as they did not meet the safe staffing levels.

The VSS was calculated to be 3.50% with daily savings of £24,756.00. The location of where beds should be deployed in this scenario can be seen in Figures D.51 and D.52.

	Total Beds		**Total Staff**		**Objective Function Value (£)**
	x^{bed}	u^{bed}	x^{staff}	u^{staff}	
Deterministic	979	-	320	-	£673,093.00 = EV
Stochastic	891	290	304	80	£698,680.60 = RP
Test A	979	191	320	80	£723,436.60 = EEV

Table 6.18: The EV, RP and EEV values for the x^{bed}, u^{bed}, x^{staff}, u^{staff} decision variables and objective function for Scenario 4 where the nursing capacity is reduced.

The scenario analysis can use more complex scenarios by modifying the demand through individual nodes by linking the predictive and prescriptive paradigms. This is more realistic than the current practice of increasing and decreasing the demands by a fixed percentage. The complete three years' worth of data will be utilised. Although more savings could be achieved by planning on a smaller time frame, such as annually, it would be impracticable for the health board to implement such adjustments frequently. There will be two long-term prediction scenarios examined. The first will investigate how the introduction of virtual wards could reduce demand, while the second will analyse a sudden increase in demand.

6.3.4.2 Introduction of Virtual Wards

It is well-known within the medical community that patients who receive care at home recover quicker [295]. This can be due to more support from family members and care-givers or less risk of infection. A new development in healthcare are virtual hospital wards, which came forth in response to the Covid-19 pandemic. With the support of these wards, patients can obtain treatment and monitoring in the convenience of their own homes, lessening the strain on hospitals and lowering the risk of contracting Covid-19. With the use of various digital technologies, virtual hospital wards can offer patients remote monitoring, doctor consultations, and access to medical supplies and medications. This approach to healthcare delivery has the potential to revolutionise the way practitioners provide care to patients, particularly those with chronic conditions, and enhancing patient outcomes while reducing healthcare costs. Virtual wards can be used to discharge patients more quickly or prevent them from being admitted at all [296].

If ABUHB were to implement similar virtual wards as adopted in other regions of the UK, this could provide numerous benefits. Cardiology and respiratory care are two of the disciplines where the Croydon Health Services NHS Trust has implemented virtual wards [297]. The trust found cost savings of approximately £1,080.00 per patient, and only 20% of patients were required to be admitted to hospital. This can be applied to ABUHB by decreasing the demand for cardiology and respiratory services in one scenario by 10% and in another scenario by 30%.

The deterministic demand remains unchanged assuming there is no implementation by decision-makers to add virtual wards to the hospitals. This results in the deterministic model producing an objective value of £623,597.20 (Table 6.19). The two scenarios within the stochastic model reduce the number of beds and staff required to be deployed to a maximum of 973 beds and 324 staff, generating an objective value of £653,397.64. The VSS was calculated to be 2.60% if the demand on cardiology and respiratory admissions were to decline. This highlights the advantages of virtual wards and enables decision-makers to assess its viability from a financial and logistical standpoint.

	Total Beds		**Total Staff**		**Objective Function Value (£)**
	x^{bed}	u^{bed}	x^{staff}	u^{staff}	
Deterministic	942	-	316	-	£623,597.20 = EV
Stochastic	877	96	286	38	£653,397.64 = RP
Test A	942	36	316	14	£670,407.28 = EEV

Table 6.19: The EV, RP and EEV values for the x^{bed}, u^{bed}, x^{staff}, u^{staff} decision variables and objective function for Scenario 5 with the introduction of virtual wards.

6.3.4.3 Sudden Increase in Demand

In January 2020, Covid-19 was declared a Public Health Emergency of International Concern, with this being characterised as a pandemic on the 11^{th} March 2020 [119]. This caused sudden and extreme pressure on the NHS which was already under previous stress from inadequate planning and under-resourcing [298]. Within Chapter 2, it was discussed how there had been little planning within elderly and frail healthcare literature for sudden increases in demand within their modelling scenarios. The next scenario will consider how the model will cope with another similar Covid-19 pandemic situation. The deterministic model will utilise the normal regression demand, with the scenarios considering if demand across all specialties and hospitals increased by 20% and 40%.

Table 6.20 presents the results if demand were to suddenly increase and appropriate planning had not taken place. The objective value increases by 35.75% in the stochastic model compared to the deterministic model. This is a large increase of unexpected demand with the total number of beds increasing by up to 425. The VSS produces a value of 4.96%, with the objective function almost a third higher than the value of the deterministic model.

	Total Beds		**Total Staff**		**Objective Function Value (£)**
	x^{bed}	u^{bed}	x^{staff}	u^{staff}	
Deterministic	942	-	316	-	£623,597.20 = EV
Stochastic	1089	278	350	100	£895,035.64 = RP
Test A	942	421	316	150	£939,386.84 = EEV

Table 6.20: The EV, RP and EEV values for the x^{bed}, u^{bed}, x^{staff}, u^{staff} decision variables and objective function for Scenario 6 with the sudden increase in demand.

6.3.4.4 Applying CART to Target Nodes

The CART tree presents a more sophisticated alternative to averaging. Since Covid-19 there is now a backlog of patients waiting for inpatient treatment in hospital, and hospital managers are under increasing pressure to provide more availability of appointments. One of these specialities under increasing pressure is the trauma and orthopaedic service [299]. These patients' health deteriorate over time whilst waiting for appointments and therefore it is necessary for them to be seen quickly [300]. Instead of projecting a straightforward 10% increase in the expected demand for the trauma and orthopaedic (T&O) specialty, obtained by multiplying the average LOS by the count, the CART nodes can be skillfully employed. By selecting specific nodes tailored to the T&O specialty, we can precisely adjust the count within those nodes. This approach takes into account the diverse LOS values, resulting in a demand node that goes beyond a simple average increase. When determining the average, there are multiple options to consider. One option is to focus solely on nodes containing the T&O specialty, ensuring a more targeted approach. Alternatively, one could include all nodes offering T&O services, broadening the scope for calculation. The benefit of using CART lies in its flexibility, as users have the freedom to handpick nodes that align with their unique requirements, thus creating a more personalised and refined demand projection.

The first of the following examples will analyse simply increasing the overall demand by 10% of T&O services using the average demands from Table 5.31. The second example will use the regression tree to target those which are specific T&O nodes (Nodes four, 15 and 16 from Figure 5.6). The count will be increased by 10%.

The demand for the first example totals 972.4808 daily demand for beds compared to the second where the sum is 979.2914. Even though within the second example, not all the T&O nodes are increased (since only the nodes where T&O is the only specialty are included), the daily bed demand is larger than in the first example. This shows that by using the regression tree to generate the demand, more variation has been included.

Table 6.21 displays the results for the two examples. The number of beds and nursing staff deployed remains similar with the largest difference of four beds and four nurses. The VSS solution was calculated to be 4.91% and 5.00% for the first and second examples, respectively. This highlights the benefit of using the CART model to deploy the demand, as higher VSS values can be generated, and more variation within the demands is created. This enables a more realistic and representative of the real-world problem. Through the utilisation of CART, the user can enhance the model's predictive capabilities, enabling it to cater to more specific and precise future demands. This is achieved by fine-tuning each of the end nodes within the tree structure, rather than just examining one specific specialty, effectively integrating reliable future forecasts into the model's decision-making process. As a result, the model becomes better equipped to provide tailored and accurate predictions for upcoming scenarios.

		Total Beds		Total Staff		Objective Value (£)
		x^{bed}	u^{bed}	x^{staff}	u^{staff}	
Example 1	Deterministic	992	-	330	-	£679,546.00 = EV
	Stochastic	828	410	274	122	£697,705.96 = RP
	Test A	992	190	330	78	£732,625.12 = EEV
Example 2	Deterministic	996	-	328	-	£668,553.60 = EV
	Stochastic	841	400	272	122	£689,015.92 = RP
	Test A	996	196	328	84	£722,848.16 = EEV

Table 6.21: The EV, RP and EEV values for the x^{bed}, u^{bed}, x^{staff}, u^{staff} decision variables and objective function for Scenario 7 with a 10% increase in demand for T&O services. Example 1 is the case where the overall average demand is increased by 10% and Example 2 is targeting T&O only nodes within the regression tree and increasing the demand by 10%.

This section has provided an overview of a variety of scenarios that the models are able to plan for. This can aid decision-makers when planning services by determining how beds and staff would need to be deployed for future demands. Whilst tailored to specific questions determined by ABUHB, the flexibility within the model allows the user to apply this to other scenarios they may wish to investigate.

6.4 Generalisability of Results

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 study will be beneficial and instructive for other academics, professionals, and policymakers who could be working in other locations or with various populations. Also, generalising findings contributes to a study's external validity, which is crucial for creating a solid body of scientific knowledge.

Both Microsoft Excel and Python implementations have been supplied, in Chapter 7 and are available on GitHub [287], in order to make the models flexible for use by other researchers and healthcare specialists. Excel's OpenSolver was adopted since it can be utilised by staff at all levels within the ABUHB and does not require any prior programming skills. The health board's data is conveniently saved in Microsoft Excel files, making it easy for users to enter their data into the model. The Python model is also provided because it is flexible, allowing users to make changes quickly and simply, in response to evolving data or requirements. Because of its versatility, the model can still make precise predictions as more data becomes available. Furthermore, an adaptable Python model allows for more efficient experimentation and testing, as it can be quickly adjusted and re-run with different parameters. To ensure users are able to apply this to their own work, Chapter 7 provides a tutorial on how to utilise both models.

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 health board, the equations can be applied 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 health board. 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 into the model. The models are adaptable and reliable to suit a variety of healthcare situations.

6.5 Summary

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.

The analysis conducted has proven to be incredibly helpful for the healthboard on various fronts. One key takeaway from the models is the clear demonstration of the drawbacks of planning solely based on averages. 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. For a healthboard accustomed to simpler average models, this project has showcased the true potential of mathematical modelling, revealing its power in unravelling complexities, optimising operations, and delivering data-driven insights into healthcare planning. Moreover, the scenario analysis has provided the healthboard with valuable revelations regarding potential structural changes within the organisation. By delving into the intricate factors that influence overall bed demand, they now have a comprehensive understanding of how different variables can impact resource requirements. Armed with this knowledge, the health board is better equipped to make informed and strategic choices in terms of resource allocation and capacity planning.

This chapter 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 South East Wales. 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. The addition of GUH, adding a soft constraint penalty and determining future scenarios were the three avenues explored. These results showed the potential and robustness of the models, which enables them to be applied to future scenarios that the health board may wish to investigate. The models are also generalisable so can be applied to any age demographic or hospital region and therefore can be used in other aspects of ABUHB and applied to worldwide healthcare organisations.

In the following chapter, Chapter 7, a tutorial is provided on how to use the Microsoft Excel OpenSolver and the Python PuLP tools.

Chapter 7

Decision Support Tool for Multi-Hospital Planning of Frail and Elderly Resource Capacities

7.1 Introduction

This chapter will provide a tutorial guide on how to use the two tools discussed within Chapters 5 and 6. These tools are adaptable so can be applied to other health boards and scenarios or other patient groupings. Section 7.2 will discuss the Microsoft Excel OpenSolver Tool which is specific to ABUHB. This model and approach can be replicated and applied to any other health board. Section 7.3 discusses the Python PuLP implementation of the model, which can be generalised to any health board situation.

7.2 Excel Implementation

Microsoft Excel is a widely used tool for data analysis and management, and OpenSolver is a powerful optimisation engine that can be used to solve complex problems within Excel. In this guide, we will provide step by step instructions on how to run the OpenSolver model, from setting up your data and formulating your problem to running the optimisation and analysing the results.

The OpenSolver model without the ABUHB data has been provided on GitHub [287] to allow users to enter their own data and hospital specialties. Each type of parameter has its own individual sheet and is clearly named to ensure the planners can easily access data. Due to the limitations of the software, all decision variables in the optimisation model must be on the same sheet. For visualisation purposes, the decision variables are automatically transferred into the model sheets after the experiment has run.

Data

Figure 7.1 displays the data which is used by the model. This is stored within the 'Data' tab and allows users to enter and change the parameters to suit their model. The user is required to enter the hospital and specialty into which a patient is admitted. The 'Short LOS' determines the number of nights spent in hospital, whilst the 'LOS_hours' determines the continuous time spent in hospital. Additionally, the date column can be added if the user wishes to split their model by year, season, month or days of the week. Finally, there is the NHS Patient Identifier which is unique to the patient.

	A	B	C	D	E	F
1	**Specialty**	**Hospital**	**Short LOS**	**LOS_hours**	**Date**	**NHS Patient Identifier**
2	General Medicine	Nevill Hall Hospital	4	97.03333333	01/01/2019	1001
3	Gastroenterology	Royal Gwent Hospital	4	108.4	01/01/2019	1002
4	General Surgery	Ysbyty Ystrad Fawr	1	26.5	02/01/2019	1003
5	Trauma & Orthopaedic	Royal Gwent Hospital	9	210.8333333	02/01/2019	1004
6	Respiratory	Royal Gwent Hospital	8	209.05	02/01/2019	1005
7	Care Of The Elderly	Royal Gwent Hospital	2	58.2	03/01/2019	1006
8	Rehabilitation	Ysbyty Ystrad Fawr	10	238.5	03/01/2019	1007
9	Urology	Royal Gwent Hospital	1	30.08333333	04/01/2019	1008
10	Diabetes And Endocrinology	Royal Gwent Hospital	2	46.23333333	05/01/2019	1009
11	GP Other	Rhymney Integrated Health & Social Care Centre	177	4257.983333	05/01/2019	1010

Figure 7.1: The data requirements for the Excel OpenSolver plugin where the user is required to enter, as a minimum, the 'Specialty', 'Hospital' and either 'Short LOS' or 'LOS_hours' for each patient.

Demand

The demand for each specialty is automatically generated from the data inputted by the user and is stored in the 'Demands' tab within the Excel spreadsheet. The average demand is calculated by determining if a specialty and hospital combination is present, and if so, calculating the total. Similarly, if patients do fall within these combinations then the average LOS, using 'LOS_hours', is calculated for each specialty and hospital. These values are then multiplied together and subsequently divided by 24, as the LOS is given in hours, and then by the total number of days in the data, to give a daily demand. This means if a user wants to determine a monthly demand, the 'Number of Days in Data' can be changed to the number of months within the data. This is then stored as a table shown in Figure 7.2.

Possible Hospital Locations

The tab entitled 'Hospital-Specialty' contains data regarding whether a hospital is able to open a specialty. In order to restrict the number of beds that can be

	A	B	C	D	E	F	G
1	AVERAGE Daily Beds Req.	Region 1	Region 2	Region 3	Region 4	Region 5	Other
2	Accident & Emergency	2.1081	0.0000	0.0000	0.0000	9.1846	0.0000
3	Anaesthetics	4.6079	0.0000	0.0000	0.0000	0.7734	0.0000
4	Cardiology	16.0947	0.0000	0.0000	0.0000	9.8809	0.0002
5	Care of the Elderly	94.5387	57.7380	0.7489	8.7416	46.4786	0.0000
6	Community Medicine	0.0000	6.9952	0.0000	0.3121	12.9756	0.0000
7	Dermatology	2.4192	0.0000	0.0000	0.0000	0.0000	0.0000
8	Diabetes and Endocrinology	14.5635	21.2838	0.0000	0.0000	17.2387	0.0000
9	Ear Nose & Throat	3.2480	0.0041	0.0000	0.0000	0.0000	0.0000
10	Gastroenterology	13.0208	0.6065	0.0000	0.0000	20.1331	0.0725
11	General Medicine	84.9712	0.9846	0.0115	0.0000	14.1695	0.0000
12	General Surgery	46.5808	0.5390	0.0000	0.0000	21.8943	0.0006
13	GP Other	0.0000	9.8734	0.0000	0.0000	15.2880	0.0000
14	Gynaecology	2.1194	0.1222	0.0000	0.0000	1.1716	0.0002
15	Haematology	3.0718	0.0320	0.0000	0.0000	1.8339	0.0002
16	Infectious Diseases	7.1817	0.0000	0.0000	0.0000	0.0000	0.0000
17	Intermediate Care	0.0000	0.0000	0.3426	0.3248	0.0000	0.0000
18	Maxillo-Facial	1.1831	0.0000	0.0000	0.0000	0.0243	0.0000
19	Neurology	1.5828	0.0000	0.0000	0.0000	0.0000	0.0000
20	Ophthalmology	2.5538	0.0028	0.0000	0.0000	0.1968	0.5232
21	Pain	0.0586	0.0055	0.0000	0.0037	0.0131	0.0000
22	Plastic Surgery	0.0000	0.0000	0.0000	0.0000	0.0336	0.0002
23	Radiology	0.0146	0.0000	0.0000	0.0000	0.0026	0.0000
24	Radiotherapy and Oncology	0.2265	0.0000	0.0000	0.0000	0.0000	0.0000
25	Rehabilitation	62.8610	32.9653	68.9710	31.9917	24.4659	0.0000
26	Respiratory	29.8010	0.0000	0.0000	0.0000	27.8372	0.0000
27	Restorative Dentistry	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000
28	Rheumatology	0.0004	0.0015	0.0000	0.0000	0.0128	0.0000
29	Trauma & Orthopaedic	60.3126	0.6665	0.0000	0.0000	41.5186	0.0000
30	Urology	12.3397	0.0521	0.0000	0.0000	0.2034	0.0702
31							

I	J
Number of Days in Data:	1096
Number of Weeks:	156.57
Number of Months:	36.53
Number of Years:	3.00

Figure 7.2: The average daily bed demand matrix automatically generated by Excel, segmented by region and specialty. The user is required to enter the number of days within the data.

deployed in a hospital's location, the total bed capacity for the 1^{st} and 2^{nd} stages of the model can be seen within Figure 7.3. As some hospitals may not be able to open full capacity to one specialty, due to resource or space limitations, the user is able to reduce these values whilst still being able to open up to the full capacity across other specialties. If a value of zero is present, this means the hospital is unable to open that specialty.

Hospital Costs

Within the 'Hospital Costs' tab in the spreadsheet, the user is able to enter the average daily cost for each specialty. Similar to demands, if the user wants to work in a different time frame; monthly, seasonally, or yearly, the cost figures can also be adjusted. Figure 7.4 displays the 1^{st} stage hospital costs, with an identical matrix grid being found below in the spreadsheet for the 2^{nd} stage costs.

Staffing

The 'Staffing' tab in the Excel worksheet contains all the necessary information for the staffing requirement. Firstly, the staff to patient ratio can be altered depending on the band level and the specialty (Figure 7.5). The hourly and daily cost per member of staff can also be changed. This flexibility allows pay rises to be included and flexibility in costings across other countries. Similarly, the cost of NHS bank

	A	B	C	D	E	F	G	H	I	J	K	L	M
1		Possible Hospital Locations											
2		RGH	STWAH	STWCH	YYF	RIHSC	YAB	GUH	CH	NHH	CCH	MVHSF	Other
3	**Accident and Emergency**	588	0	0	0	0	0	0	0	309	0	0	0
4	**Anaesthetics**	588	0	0	0	0	0	0	0	309	0	0	0
5	**Cardiology**	588	0	0	0	0	0	0	0	309	0	0	80
6	**Care of the Elderly**	588	0	75	163	0	80	0	73	309	26	0	0
7	**Community Medicine**	0	0	0	163	0	0	0	73	0	26	0	0
8	**Dermatology**	588	28	75	163	0	0	0	0	0	0	0	0
9	**Diabetes and Endocrinology**	588	0	0	163	0	0	0	0	309	0	0	0
10	**Ear, Nose and Throat**	588	0	0	163	0	0	0	0	0	0	0	0
11	**Gastroenterology**	588	28	0	163	0	0	0	0	309	0	0	80
12	**General Medicine**	588	0	0	163	0	80	0	0	309	0	0	0
13	**General Surgery**	588	0	0	163	0	0	0	0	309	26	0	80
14	**GP Other**	0	0	0	0	12	0	0	0	0	26	16	80
15	**Gynaecology**	588	0	0	163	0	0	0	0	309	0	0	80
16	**Haematology**	588	28	0	163	0	0	0	0	309	0	0	80
17	**Infectious Diseases**	588	0	0	0	0	0	0	0	0	0	0	0
18	**Intermediate Care**	0	0	0	0	0	80	0	73	0	0	0	0
19	**Maxillo-Facial**	588	28	0	0	0	0	0	0	309	0	0	0
20	**Neurology**	588	0	0	0	0	0	0	0	0	0	0	0
21	**Opthamology**	588	0	0	163	0	0	0	0	309	0	0	80
22	**Pain**	588	0	0	163	0	0	0	73	309	0	0	0
23	**Plastic Surgery**	0	0	0	0	0	0	0	0	309	0	0	80
24	**Radiology**	588	0	0	0	0	0	0	0	309	0	0	0
25	**Radiotherapy and Oncology**	588	0	0	0	0	0	0	0	0	0	0	80
26	**Rehabiltation**	588	28	75	163	0	80	0	73	309	26	0	0
27	**Respiratory**	588	0	0	0	0	0	0	0	309	0	0	0
28	**Restorative Dentistry**	588	0	0	163	0	0	0	0	0	0	0	0
29	**Rheumatology**	588	0	0	163	0	0	0	0	309	0	0	0
30	**Trauma & Orthopaedic**	588	28	0	163	0	0	0	0	309	0	0	0
31	**Urology**	588	28	0	163	0	0	0	0	309	0	0	80
32													
33	**Total capacity xbed**	588	28	75	163	12	80	0	73	309	26	16	80
34	**Additional capacity ubed**	177	9	28	49	4	24	0	22	93	8	5	24
35													

Figure 7.3: The maximum number of beds that can be deployed to each hospital location for each specialty. Additionally, the user is required to enter the total capacity for the hospitals in the first and second stages.

	A	B	C	D	E	F	G	H	I	J	K	L	M
1		**1st Stage Hospital Costs**											
2		RGH	STWAH	STWCH	YYF	RIHSC	YAB	GUH	CH	NHH	CCH	MVHSCF	Other
3	Accident & Emergency	0	0	0	0	0	0	247	0	0	0	0	0
4	Anaesthetics	0	0	0	0	0	0	1021	0	0	0	0	0
5	Cardiology	396	0	0	0	0	0	551	0	895	0	0	0
6	Care of the Elderly	457	0	755	493	0	542	577	472	743	0	0	0
7	Community Medicine	0	0	0	942	0	0	0	1100	0	1021	0	0
8	Dermatology	0	1381	0	0	0	0	0	0	0	0	0	0
9	Diabetes and Endocrinology	296	0	0	1021	0	0	1270	0	1497	0	0	0
10	Ear Nose & Throat	481	0	0	501	0	0	481	0	501	0	0	0
11	Gastroenterology	448	0	0	639	0	0	402	0	959	0	0	832
12	General Medicine	390	0	0	94	0	65	290	0	611	0	0	0
13	General Surgery	472	0	0	304	0	0	849	0	539	0	0	0
14	GP Other	0	0	0	0	190	0	0	0	0	0	460	0
15	Gynaecology	1007	0	0	528	0	0	517	0	292	0	0	241
16	Haematology	1299	0	0	0	0	0	1255	0	1070	0	0	0
17	Infectious Diseases	711	0	0	0	0	0	0	0	0	0	0	0
18	Intermediate Care	0	0	0	0	0	39	0	197	0	0	0	0
19	Maxillo-Facial	2026	0	0	0	0	0	1940	0	264	0	0	0
20	Neurology	1273	0	0	0	0	0	0	0	0	0	0	0
21	Ophthalmology	957	0	0	369	0	0	729	0	1416	0	0	174
22	Pain	124	0	0	111	0	0	0	134	143	0	0	0
23	Plastic Surgery	0	0	0	0	0	0	0	0	0	0	0	902
24	Radiology	945	0	0	0	0	0	1021	0	1097	0	0	0
25	Radiotherapy and Oncology	0	0	0	0	0	0	0	0	0	0	0	1089
26	Rehabilitation	1973	0	972	975	0	1021	1274	1983	1987	1455	0	0
27	Respiratory	330	0	0	448	0	0	448	0	566	0	0	0
28	Restorative Dentistry	140	0	0	0	0	0	0	0	0	0	0	0
29	Rheumatology	0	0	0	0	0	0	0	0	596	0	0	0
30	Trauma & Orthopaedic	678	651	0	610	0	0	703	0	873	0	0	0
31	Urology	102	0	0	0	0	0	140	0	374	0	0	900

Figure 7.4: The user is required to enter the first stage cost for each hospital and specialty combination.

and agency staff was also included. Finally, the maximum number of staff that can be deployed both in the first and second stages is detailed.

One limitation of the Excel model is higher levels of staff cannot perform the role of the lower band staff members. This is due to the non-linearity of the constraint which the COIN-OR CBC (Linear solver) cannot handle.

	Ratio of Staff	
	Band 5	Band 6
Accident & Emergency	0.250	0.250
Anaesthetics	0.250	0.250
Cardiology	0.125	0.125
Care of the Elderly	0.1	0.1
Community Medicine	0.1	0.1
Dermatology	0.1	0.1
Diabetes and Endocrinology	0.125	0.125
Ear Nose & Throat	0.125	0.125
Gastroenterology	0.125	0.125
General Medicine	0.25	0.25
General Surgery	0.25	0.25
GP Other	0.1	0.1
Gynaecology	0.125	0.125
Haematology	0.125	0.125
Infectious Diseases	0.125	0.125
Intermediate Care	0.1	0.1
Maxillo-Facial	0.125	0.125
Neurology	0.125	0.125
Ophthalmology	0.125	0.125
Pain	0.125	0.125
Plastic Surgery	0.125	0.125
Radiology	0.125	0.125
Radiotherapy and Oncology	0.125	0.125
Rehabilitation	0.1	0.1
Respiratory	0.125	0.125
Restorative Dentistry	0.125	0.125
Rheumatology	0.125	0.125
Trauma & Orthopaedic	0.25	0.25
Urology	0.125	0.125
Staffing cost xstaff per day	338.88	419.52
Cost of Agency/Bank staff per day	454.8	560.64
Staffing cost xstaff per hour	14.12	17.48
Cost of Agency/Bank staff per hour	18.95	23.36
Total capacity xstaff	400	400
Total Agency/Bank staff capacity	400	400

Figure 7.5: The user is required to enter the ratio of nursing band staff to each specialty. Additionally, the user is required to enter the cost per hour of the first and second stage nurses staff, and the total capacity.

7.2.1 Deterministic Model

The deterministic model is stored within the 'Deterministic' tab, where the optimisation model can be run and the results analysed. With the OpenSolver add-in installed, the model can be easily accessed through the Data ribbon and then selecting the Model on the OpenSolver toolbar. This brings up Figure 7.6, which depicts the objective function cell, the type of problem (maximisation or minimisation), the decision variables, the model's constraints and the type of solver engine. Using the options button; the maximum solution time, branch and bound tolerance and the maximum number of iterations can also be changed. To solve the model, the 'Solve' button can be selected on the OpenSolver toolbar.

Once executed, the total cost of the model is shown within the sheet, along with the total number of beds and staff to be deployed (Figure 7.7). The model shows where each of these beds should be deployed across the specialties and hospitals, and the overall number of beds within each hospital. Similarly, the number of staff deployed for each band can be visualised within the same worksheet, as shown in Figure 7.8.

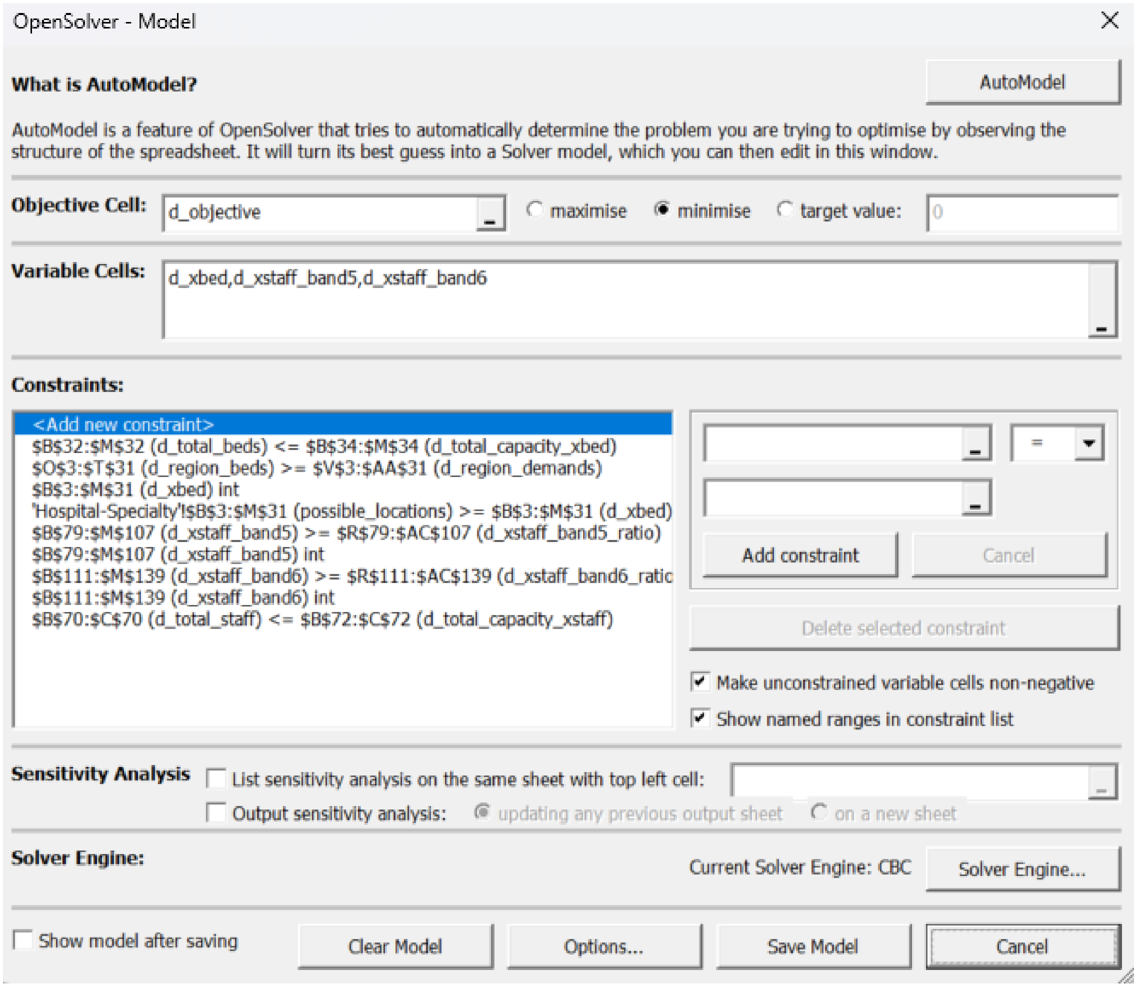


Figure 7.6: The OpenSolver objective cell, decision variable cells and constraints for the deterministic implementation. The 'Current Solver Engine' can be changed to Gurobi if a license is available.

		Xbed											
		RGH	STWAH	STWCH	YYF	RIHSC	YAB	GUH	CH	NHH	CCH	MVHSF	Other
3	Accident & Emergency	3	0	0	0	0	0	0	0	10	0	0	0
4	Anaesthetics	5	0	0	0	0	0	0	0	1	0	0	0
5	Cardiology	17	0	0	0	0	0	0	0	10	0	0	1
6	Care of the Elderly	95	0	0	58	0	1	0	9	47	0	0	0
7	Community Medicine	0	0	0	7	0	0	0	1	0	13	0	0
8	Dermatology	0	3	0	0	0	0	0	0	0	0	0	0
9	Diabetes and Endocrinology	15	0	0	22	0	0	0	0	18	0	0	0
10	Ear Nose & Throat	4	0	0	1	0	0	0	0	0	0	0	0
11	Gastroenterology	0	14	0	1	0	0	0	0	21	0	0	1
12	General Medicine	85	0	0	1	0	1	0	0	15	0	0	0
13	General Surgery	47	0	0	1	0	0	0	0	22	0	0	1
14	GP Other	0	0	0	0	10	0	0	0	0	0	16	0
15	Gynaecology	3	0	0	1	0	0	0	0	2	0	0	1
16	Haematology	4	0	0	1	0	0	0	0	2	0	0	1
17	Infectious Diseases	8	0	0	0	0	0	0	0	0	0	0	0
18	Intermediate Care	0	0	0	0	0	1	0	1	0	0	0	0
19	Maxillo-Facial	0	2	0	0	0	0	0	0	1	0	0	0
20	Neurology	2	0	0	0	0	0	0	0	0	0	0	0
21	Ophthalmology	3	0	0	1	0	0	0	0	1	0	0	1
22	Pain	1	0	0	1	0	0	0	1	1	0	0	0
23	Plastic Surgery	0	0	0	0	0	0	0	0	1	0	0	1
24	Radiology	1	0	0	0	0	0	0	0	1	0	0	0
25	Radiotherapy and Oncology	1	0	0	0	0	0	0	0	0	0	0	0
26	Rehabilitation	0	0	63	33	0	69	0	32	12	13	0	0
27	Respiratory	30	0	0	0	0	0	0	0	28	0	0	0
28	Restorative Dentistry	1	0	0	0	0	0	0	0	0	0	0	0
29	Rheumatology	1	0	0	1	0	0	0	0	1	0	0	0
30	Trauma & Orthopaedic	52	9	0	1	0	0	0	0	42	0	0	0
31	Urology	13	0	0	1	0	0	0	0	1	0	0	1
32	TOTAL BEDS	391	28	63	131	10	72	0	44	237	26	16	8
33		<=	<=	<=	<=	<=	<=	<=	<=	<=	<=	<=	<=
34	Total Capacity xbed	588	28	75	163	12	80	0	73	309	26	16	80

		Region beds							Region Demands					
		Region 1	Region 2	Region 3	Region 4	Region 5	Other		Region 1	Region 2	Region 3	Region 4	Region 5	Other
3	Accident & Emergency	3	0	0	0	10	0	>=	2.108091	0	0	0	9.184566	0
4	Anaesthetics	5	0	0	0	1	0	>=	4.607919	0	0	0	0.773419	0
5	Cardiology	17	0	0	0	10	1	>=	16.09469	0	0	0	9.88089	0.000192
6	Care of the Elderly	95	58	1	9	47	0	>=	94.5387	57.73797	0.748914	8.741643	46.47864	0
7	Community Medicine	0	7	0	1	13	0	>=	0	6.995192	0	0.312057	12.97563	0
8	Dermatology	3	0	0	0	0	0	>=	2.419185	0	0	0	0	0
9	Diabetes and Endocrinology	15	22	0	0	18	0	>=	14.56348	21.28381	0	0	17.2387	0
10	Ear Nose & Throat	4	1	0	0	0	0	>=	3.247995	0.004111	0	0	0	0
11	Gastroenterology	14	1	0	0	21	1	>=	13.02079	0.60654	0	0	20.13312	0.072491
12	General Medicine	85	1	1	0	15	0	>=	84.97115	0.984621	0.011481	0	14.16952	0
13	General Surgery	47	1	0	0	22	1	>=	46.5808	0.538987	0	0	21.89427	0.000576
14	GP Other	0	10	0	0	16	0	>=	0	9.873353	0	0	15.28797	0
15	Gynaecology	3	1	0	0	2	1	>=	2.11941	0.122213	0	0	1.171593	0.000192
16	Haematology	4	1	0	0	2	1	>=	3.071771	0.032025	0	0	1.833921	0.000192
17	Infectious Diseases	8	0	0	0	0	0	>=	7.181749	0	0	0	0	0
18	Intermediate Care	0	0	1	1	0	0	>=	0	0	0.342639	0.324786	0	0
19	Maxillo-Facial	2	0	0	0	1	0	>=	1.183052	0	0	0	0.024335	0
20	Neurology	2	0	0	0	0	0	>=	1.58284	0	0	0	0	0
21	Ophthalmology	3	1	0	0	1	1	>=	2.553845	0.002842	0	0	0.196848	0.523204
22	Pain	1	1	0	1	1	0	>=	0.058633	0.005474	0	0.003656	0.01313	0
23	Plastic Surgery	0	0	0	0	1	1	>=	0	0	0	0	0.033643	0.000192
24	Radiology	1	0	0	0	1	0	>=	0.014583	0	0	0	0.002597	0
25	Radiotherapy and Oncology	1	0	0	0	0	0	>=	0.226512	0	0	0	0	0
26	Rehabilitation	63	33	69	32	25	0	>=	62.861	32.96529	68.97105	31.9917	24.46588	0
27	Respiratory	30	0	0	0	28	0	>=	29.80097	0	0	0	27.8372	0
28	Restorative Dentistry	1	0	0	0	0	0	>=	8.3E-05	0	0	0	0	0
29	Rheumatology	1	1	0	0	1	0	>=	0.000361	0.001464	0	0	0.012836	0
30	Trauma & Orthopaedic	61	1	0	0	42	0	>=	60.31262	0.666489	0	0	41.51862	0
31	Urology	13	1	0	0	1	1	>=	12.33965	0.052112	0	0	0.203415	0.070174

Objective Function - Daily Cost: £ 904280.80

Total Beds Deployed: 1026

Total Staff Deployed: 414

Figure 7.7: The output from the deterministic model once solved, displaying the number of beds to deploy to each hospital and specialty. The total daily cost, beds and staff are also summarised.

		xstaff	
40		Band 5	Band 6
41	Accident & Emergency	4	4
42	Anaesthetics	3	3
43	Cardiology	6	6
44	Care of the Elderly	23	23
45	Community Medicine	4	4
46	Dermatology	1	1
47	Diabetes and Endocrinology	8	8
48	Ear Nose & Throat	2	2
49	Gastroenterology	7	7
50	General Medicine	28	28
51	General Surgery	20	20
52	GP Other	3	3
53	Gynaecology	4	4
54	Haematology	4	4
55	Infectious Diseases	1	1
56	Intermediate Care	2	2
57	Maxillo-Facial	2	2
58	Neurology	1	1
59	Ophthalmology	4	4
60	Pain	4	4
61	Plastic Surgery	2	2
62	Radiology	2	2
63	Radiotherapy and Oncology	1	1
64	Rehabilitation	26	26
65	Respiratory	8	8
66	Restorative Dentistry	1	1
67	Rheumatology	3	3
68	Trauma & Orthopaedic	28	28
69	Urology	5	5
70	TOTAL STAFF	207	207
71		<=	<=
72	Total Capactiy xstaff	400	400

Figure 7.8: The staffing output from the deterministic model once solved, displaying the number of nursing staff to deploy to each specialty and of which band level.

7.2.2 Two-Stage Stochastic Model

The two-stage stochastic model is stored within the 'Stochastic' tab. Due to the large number of decision variables within the stochastic model (6,960 variables), the model is stored within 'SVariables' and the results are automatically transferred into the 'Stochastic' tab.

In addition to the deterministic model parameters, the scenarios and probabilities for each scenario are required. The Excel tool can use up to four scenarios by changing the values as demonstrated in Figure 7.9.

Scenario 1: Demand Scalar:	Probability
0.95	0.25

Figure 7.9: The scenario selector for the two-stage stochastic implementation. Users are prompted to input the demand scalar and the corresponding probability of occurrence.

Similar to the deterministic model, the stochastic model's constraints can be seen within the 'Data' ribbon and selecting the 'Model' option. Figure 7.10 shows the objective function cell, as well as the decision variable locations. The stochastic model contains 31 constraints, as well as the option to make unconstrained variable cells non-negative. Once again, the COIN-OR CBC linear solver or the Gurobi can be used.

Once the model has run, the output produced is similar to the one shown in Figure 7.11. The model determines how many beds and staff to deploy in the first and

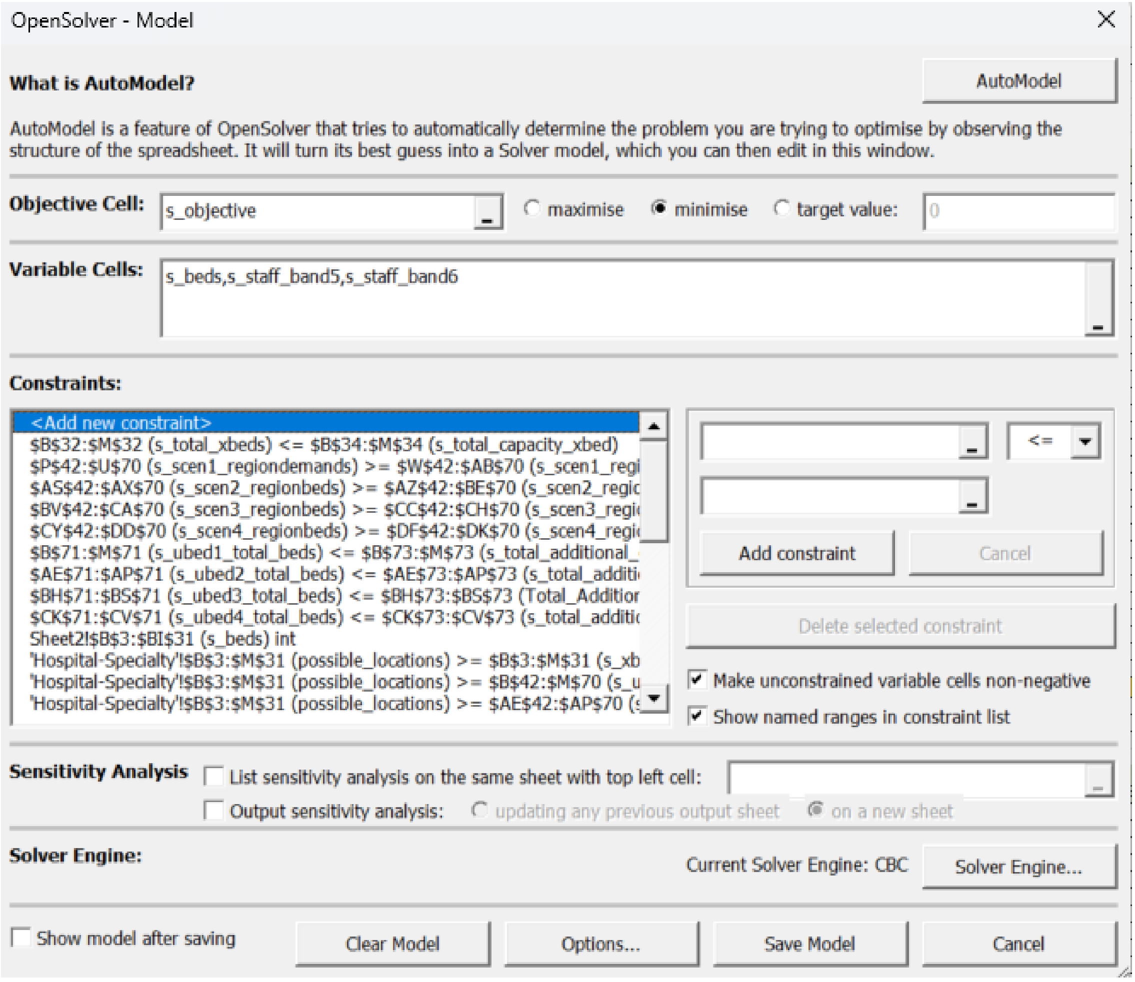


Figure 7.10: The OpenSolver objective cell, decision variables and constraints for the two-stage stochastic implementation. The ‘Current Solver Engine’ can be changed to Gurobi if a license is available.

second stages to ensure the minimum demand is met. A summary is provided within rows 37 and 38 on the worksheet, but the full breakdown of the results can be seen within the remainder of the worksheet.

7.2.3 Test A

Test A is stored in the ‘TestA’ tab, where the optimisation model can be run and the results analysed. Similar to the stochastic model, due to the large number of decision variables within the optimisation model, the model is stored within ‘OVariables’ and the results are automatically transferred into the ‘TestA’ tab.

For Test A, the first stage of the model is required to be fixed to the results of the deterministic model. The Excel spreadsheet has been setup in a way that requires no additional user input is required for this Test. This is achieved by linking the cells together, i.e., `=Deterministic!B3`, which would copy the first hospital and specialty into the first stage of the model.

	A	B	C	D	E	F	G	H	I	J	K	L	M
1		Xbed											
2		RGH	STWAH	STWCH	YYF	RIHSC	YAB	GUH	CH	NHH	CCH	MVHSF	Other
3	Accident and Emergency	3	0	0	0	0	0	0	0	8	0	0	0
4	Anaesthetics	4	0	0	0	0	0	0	0	1	0	0	0
5	Cardiology	13	0	0	0	0	0	0	0	8	0	0	1
6	Care of the Elderly	76	0	0	50	0	1	0	9	38	0	0	0
7	Community Medicine	0	0	0	7	0	0	0	1	0	10	0	0
8	Dermatology	0	3	0	0	0	0	0	0	0	0	0	0
9	Diabetes and Endocrinology	15	0	0	18	0	0	0	0	14	0	0	0
10	Ear, Nose and Throat	4	0	0	1	0	0	0	0	0	0	0	0
11	Gastroenterology	0	8	0	1	0	0	0	0	17	0	0	1
12	General Medicine	68	0	0	2	0	1	0	0	12	0	0	0
13	General Surgery	40	0	0	1	0	0	0	0	19	0	0	1
14	GP Other	0	0	0	0	10	0	0	0	0	0	16	0
15	Gynaecology	3	0	0	1	0	0	0	0	2	0	0	1
16	Haematology	4	0	0	1	0	0	0	0	2	0	0	1
17	Infectious Diseases	8	0	0	0	0	0	0	0	0	0	0	0
18	Intermediate Care	0	0	0	0	0	1	0	1	0	0	0	0
19	Maxillo-Facial	0	2	0	0	0	0	0	0	1	0	0	0
20	Neurology	2	0	0	0	0	0	0	0	0	0	0	0
21	Opthamology	3	0	0	1	0	0	0	0	1	0	0	1
22	Pain	1	0	0	1	0	0	0	1	1	0	0	0
23	Plastic Surgery	0	0	0	0	0	0	0	0	1	0	0	1
24	Radiology	1	0	0	0	0	0	0	0	1	0	0	0
25	Radiotherapy and Oncology	1	0	0	0	0	0	0	0	0	0	0	0
26	Rehabiltation	0	0	50	27	0	59	0	26	4	16	0	0
27	Respiratory	24	0	0	0	0	0	0	0	23	0	0	0
28	Restorative Dentistry	1	0	0	0	0	0	0	0	0	0	0	0
29	Rheumatology	1	0	0	1	0	0	0	0	1	0	0	0
30	Trauma & Orthopaedic	34	15	0	1	0	0	0	0	34	0	0	0
31	Urology	15	0	0	1	0	0	0	0	1	0	0	1
32	TOTAL BEDS	321	28	50	114	10	62	0	38	189	26	16	8
33		<=	<=	<=	<=	<=	<=	<=	<=	<=	<=	<=	<=
34	Total Capacity xbed	588	28	75	163	12	80	0	73	309	26	16	80
35													
36													
37	Objective Function - Daily Cost: £ 945500.48				Total Beds First Stage: 862					Total Staff First Stage: 348			
38					Max beds Second Stage: 38					Max beds Second Stage: 12			

Figure 7.11: The output from the two-stage stochastic model once solved, displayed the number of beds to deploy to each hospital and specialty. The total daily cost, first and second stage beds and nursing staff are also summarised.

Similar to the previous two models, the constraints can be seen within the 'Data' ribbon and selecting the 'Model' option. Figure 7.12 displays the output of the model, showing a summary of the daily costing figures as well as the numbers of beds and staff deployed in the first and second stages.

	A	B	C	D	E	F	G	H	I	J	K	L	M
1		Xbed											
2		RGH	STWAH	STWCH	YYF	RIHSC	YAB	GUH	CH	NHH	CCH	MVHSF	Other
3	Accident and Emergency	3	0	0	0	0	0	0	0	10	0	0	0
4	Anaesthetics	5	0	0	0	0	0	0	0	1	0	0	0
5	Cardiology	17	0	0	0	0	0	0	0	10	0	0	1
6	Care of the Elderly	95	0	0	58	0	1	0	9	47	0	0	0
7	Community Medicine	0	0	0	7	0	0	0	1	0	13	0	0
8	Dermatology	0	3	0	0	0	0	0	0	0	0	0	0
9	Diabetes and Endocrinology	15	0	0	22	0	0	0	0	18	0	0	0
10	Ear, Nose and Throat	4	0	0	1	0	0	0	0	0	0	0	0
11	Gastroenterology	0	14	0	1	0	0	0	0	21	0	0	1
12	General Medicine	85	0	0	1	0	1	0	0	15	0	0	0
13	General Surgery	47	0	0	1	0	0	0	0	22	0	0	1
14	GP Other	0	0	0	0	10	0	0	0	0	0	16	0
15	Gynaecology	3	0	0	1	0	0	0	0	2	0	0	1
16	Haematology	4	0	0	1	0	0	0	0	2	0	0	1
17	Infectious Diseases	8	0	0	0	0	0	0	0	0	0	0	0
18	Intermediate Care	0	0	0	0	0	1	0	1	0	0	0	0
19	Maxillo-Facial	0	2	0	0	0	0	0	0	1	0	0	0
20	Neurology	2	0	0	0	0	0	0	0	0	0	0	0
21	Opthamology	3	0	0	1	0	0	0	0	1	0	0	1
22	Pain	1	0	0	1	0	0	0	1	1	0	0	0
23	Plastic Surgery	0	0	0	0	0	0	0	0	1	0	0	1
24	Radiology	1	0	0	0	0	0	0	0	1	0	0	0
25	Radiotherapy and Oncology	1	0	0	0	0	0	0	0	0	0	0	0
26	Rehabiltation	0	0	63	33	0	69	0	32	12	13	0	0
27	Respiratory	30	0	0	0	0	0	0	0	28	0	0	0
28	Restorative Dentistry	1	0	0	0	0	0	0	0	0	0	0	0
29	Rheumatology	1	0	0	1	0	0	0	0	1	0	0	0
30	Trauma & Orthopaedic	52	9	0	1	0	0	0	0	42	0	0	0
31	Urology	13	0	0	1	0	0	0	0	1	0	0	1
32	TOTAL BEDS	391	28	63	131	10	72	0	44	237	26	16	8
33		<=	<=	<=	<=	<=	<=	<=	<=	<=	<=	<=	<=
34	Total Capacity xbed	588	28	75	163	12	80	0	73	309	26	16	80
35													
36													
37	Objective Function - Daily Cost: £ 974078.16				Total Beds First Stage: 1026					Total Staff First Stage: 414			
38					Max beds Second Stage: 19					Max beds Second Stage: 7			
39													

Figure 7.12: The output from Test A model implementation once solved, displayed the number of beds to deploy to each hospital and specialty. The total daily cost, first and second stage beds and nursing staff are also summarised.

7.2.4 Test B

Test B is stored in the 'TestB' tab of the workbook, and its purpose is to determine if the correct first stage variables have been chosen.

This test requires the zero or lowest value stage variables of the deterministic model to be set to zero or the lower bound. As some specialties cannot open within certain hospitals of ABUHB, this means there will always be zero variables, and therefore the model is setup in a way to determine if zero beds are deployed in the deterministic model, then it will return zero. Otherwise, it will return the maximum number of beds that can be opened within the hospital. An example of how this is generated is as follows: `IF(Deterministic!B3=0, 0, 'Hospital-Specialty'!B3)`, where cell B3 is the first hospital/specialty combination. The formula checks the 'Deterministic' tab to determine whether or not the B3 cell is 0, if so, then a zero is placed into the cell. If not, the value from the corresponding cell in the 'Hospital-Specialty' tab is taken.

If there was a scenario in which, all specialties in all hospitals had been opened, the user would have to manually enter the lower bound.

The results of the 'IF statement' are then outputted into cells `AB3:AM31` (Figure 7.14). As OpenSolver works with the linear programming add-in, the 'IF statement' causes this to become non-linear. To overcome this, the user is required to copy and paste cells `AB3:AM31` into `O3:Z31` ensuring only the values and not the formulae are copied over.

Figure 7.13 illustrates the additional constraint necessary for the model to run successfully. To access this constraint, along with the other constraints, the user can select the 'Data' ribbon and the 'Model' option. The output from the model, showing the objective function and bed and staff totals can be seen in Figure 7.15.

$B$3:$M$31 (a_xbed) <= $O$3:$Z$31 (Test_B)

Figure 7.13: The additional constraint required for Test B, where the zero variables within the deterministic model are set as the lower bound for the first stage within Test B.

Test B – Copy Cells AB3:AM31 into O3:Z31											
588	0	0	0	0	0	0	0	309	0	0	0
588	0	0	0	0	0	0	0	309	0	0	0
588	0	0	0	0	0	0	0	309	0	0	80
588	0	0	163	0	80	0	73	309	0	0	0
0	0	0	163	0	0	0	73	0	26	0	0
0	28	0	0	0	0	0	0	0	0	0	0
588	0	0	163	0	0	0	0	309	0	0	0
588	0	0	163	0	0	0	0	0	0	0	0
0	28	0	163	0	0	0	0	309	0	0	80
588	0	0	163	0	80	0	0	309	0	0	0
588	0	0	163	0	0	0	0	309	0	0	80
0	0	0	0	12	0	0	0	0	0	16	0
588	0	0	163	0	0	0	0	309	0	0	80
588	0	0	163	0	0	0	0	309	0	0	80
588	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	80	0	73	0	0	0	0
0	28	0	0	0	0	0	0	309	0	0	0
588	0	0	0	0	0	0	0	0	0	0	0
588	0	0	163	0	0	0	0	309	0	0	80
588	0	0	163	0	0	0	73	309	0	0	0
0	0	0	0	0	0	0	0	309	0	0	80
588	0	0	0	0	0	0	0	309	0	0	0
588	0	0	0	0	0	0	0	0	0	0	0
0	0	75	163	0	80	0	73	309	26	0	0
588	0	0	0	0	0	0	0	309	0	0	0
588	0	0	0	0	0	0	0	0	0	0	0
588	0	0	163	0	0	0	0	309	0	0	0
588	28	0	163	0	0	0	0	309	0	0	0
588	0	0	163	0	0	0	0	309	0	0	80

Test B – Copy Cells AB3:AM31 into O3:Z31											
588	0	0	0	0	0	0	0	309	0	0	0
588	0	0	0	0	0	0	0	309	0	0	0
588	0	0	0	0	0	0	0	309	0	0	80
588	0	0	163	0	80	0	73	309	0	0	0
0	0	0	163	0	0	0	73	0	26	0	0
0	28	0	0	0	0	0	0	0	0	0	0
588	0	0	163	0	0	0	0	309	0	0	0
588	0	0	163	0	0	0	0	0	0	0	0
0	28	0	163	0	0	0	0	309	0	0	80
588	0	0	163	0	80	0	0	309	0	0	0
588	0	0	163	0	0	0	0	309	0	0	80
0	0	0	0	12	0	0	0	0	0	16	0
588	0	0	163	0	0	0	0	309	0	0	80
588	0	0	163	0	0	0	0	309	0	0	80
588	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	80	0	73	0	0	0	0
0	28	0	0	0	0	0	0	309	0	0	0
588	0	0	0	0	0	0	0	0	0	0	0
588	0	0	163	0	0	0	0	309	0	0	80
588	0	0	163	0	0	0	73	309	0	0	0
0	0	0	0	0	0	0	0	309	0	0	80
588	0	0	0	0	0	0	0	309	0	0	0
588	0	0	0	0	0	0	0	0	0	0	0
0	0	75	163	0	80	0	73	309	26	0	0
588	0	0	0	0	0	0	0	309	0	0	0
588	0	0	0	0	0	0	0	0	0	0	0
588	0	0	163	0	0	0	0	309	0	0	0
588	28	0	163	0	0	0	0	309	0	0	0
588	0	0	163	0	0	0	0	309	0	0	80

Figure 7.14: For Test B, an 'if statement' is required to be used to determine the zero bound of the decision variables from the Deterministic output. As OpenSolver is a linear programming solver, the user is required to copy and paste the values only from cells `AB3:AM31` to `O3:Z31`.

	A	B	C	D	E	F	G	H	I	J	K	L	M
1		Xbed											
2		RGH	STWAH	STWCH	YYF	RIHSC	YAB	GUH	CH	NHH	CCH	MVHSF	Other
3	Accident and Emergency	3	0	0	0	0	0	0	0	12	0	0	0
4	Anaesthetics	6	0	0	0	0	0	0	0	1	0	0	0
5	Cardiology	17	0	0	0	0	0	0	0	10	0	0	1
6	Care of the Elderly	95	0	0	58	0	1	0	9	47	0	0	0
7	Community Medicine	0	0	0	7	0	0	0	1	0	13	0	0
8	Dermatology	0	3	0	0	0	0	0	0	0	0	0	0
9	Diabetes and Endocrinology	15	0	0	22	0	0	0	0	18	0	0	0
10	Ear, Nose and Throat	4	0	0	1	0	0	0	0	0	0	0	0
11	Gastroenterology	0	14	0	1	0	0	0	0	21	0	0	1
12	General Medicine	86	0	0	2	0	1	0	0	15	0	0	0
13	General Surgery	48	0	0	1	0	0	0	0	23	0	0	1
14	GP Other	0	0	0	0	10	0	0	0	0	0	16	0
15	Gynaecology	3	0	0	1	0	0	0	0	2	0	0	1
16	Haematology	4	0	0	1	0	0	0	0	2	0	0	1
17	Infectious Diseases	8	0	0	0	0	0	0	0	0	0	0	0
18	Intermediate Care	0	0	0	0	0	1	0	1	0	0	0	0
19	Maxillo-Facial	0	2	0	0	0	0	0	0	1	0	0	0
20	Neurology	2	0	0	0	0	0	0	0	0	0	0	0
21	Opthamology	3	0	0	1	0	0	0	0	1	0	0	1
22	Pain	1	0	0	1	0	0	0	1	1	0	0	0
23	Plastic Surgery	0	0	0	0	0	0	0	0	1	0	0	1
24	Radiology	1	0	0	0	0	0	0	0	1	0	0	0
25	Radiotherapy and Oncology	1	0	0	0	0	0	0	0	0	0	0	0
26	Rehabiltation	0	0	63	33	0	69	0	32	12	13	0	0
27	Respiratory	30	0	0	0	0	0	0	0	28	0	0	0
28	Restorative Dentistry	1	0	0	0	0	0	0	0	0	0	0	0
29	Rheumatology	1	0	0	1	0	0	0	0	1	0	0	0
30	Trauma & Orthopaedic	52	9	0	1	0	0	0	0	42	0	0	0
31	Urology	15	0	0	1	0	0	0	0	1	0	0	1
32	TOTAL BEDS	396	28	63	132	10	72	0	44	240	26	16	8
33		<=	<=	<=	<=	<=	<=	<=	<=	<=	<=	<=	<=
34	Total Capacity xbed	588	28	75	163	12	80	0	73	309	26	16	80
35													
36													
37	Objective Function - Daily Cost: £ 976601.16				Total Beds First Stage: 1035					Total Staff First Stage: 414			
38					Max beds Second Stage: 19					Max beds Second Stage: 7			

Figure 7.15: The output from Test B model implementation once solved, displayed the number of beds to deploy to each hospital and specialty. The total daily cost, first and second stage beds and nursing staff are also summarised.

7.2.5 Test C

Test C is stored in the 'TestC' tab of the workbook and its purpose is to determine the upgradeability of the model. Similar to Test A, the Excel spreadsheet has been set up in a way that requires no additional user input. An example of how this is achieved is as follows: `=Deterministic!B3` which will copy the first hospital and specialty.

This results in the following matrix to be inputted into cells `O3:Z31`, as shown in Figure 7.16.

Therefore an additional constraint to the original two-stage stochastic model is added into the constraint list to ensure the first stage variables meet the minimum deterministic values. The constraint is shown in Figure 7.17.

The remainder of the model remains the same as the two-stage stochastic deployment and the output of the results can be seen as exampled in Figure 7.18.

Deterministic Minimum Values											
3	0	0	0	0	0	0	0	10	0	0	0
5	0	0	0	0	0	0	0	1	0	0	0
17	0	0	0	0	0	0	0	10	0	0	1
95	0	0	58	0	1	0	9	47	0	0	0
0	0	0	7	0	0	0	1	0	13	0	0
0	3	0	0	0	0	0	0	0	0	0	0
15	0	0	22	0	0	0	0	18	0	0	0
4	0	0	1	0	0	0	0	0	0	0	0
0	14	0	1	0	0	0	0	21	0	0	1
85	0	0	1	0	1	0	0	15	0	0	0
47	0	0	1	0	0	0	0	22	0	0	1
0	0	0	0	10	0	0	0	0	0	16	0
3	0	0	1	0	0	0	0	2	0	0	1
4	0	0	1	0	0	0	0	2	0	0	1
8	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	1	0	1	0	0	0	0
0	2	0	0	0	0	0	0	1	0	0	0
2	0	0	0	0	0	0	0	0	0	0	0
3	0	0	1	0	0	0	0	1	0	0	1
1	0	0	1	0	0	0	1	1	0	0	0
0	0	0	0	0	0	0	0	1	0	0	1
1	0	0	0	0	0	0	0	1	0	0	0
1	0	0	0	0	0	0	0	0	0	0	0
0	0	63	33	0	69	0	32	12	13	0	0
30	0	0	0	0	0	0	0	28	0	0	0
1	0	0	0	0	0	0	0	0	0	0	0
1	0	0	1	0	0	0	0	1	0	0	0
52	9	0	1	0	0	0	0	42	0	0	0
13	0	0	1	0	0	0	0	1	0	0	1

Figure 7.16: The minimum values which are required to be met within the first stage of Test C. This will determine the upgradeability of the model.

$B$3:$M$31 (c_xbed) >= $O$3:$Z$31 (Test_C)

Figure 7.17: The additional constraint required for Test C, where the deterministic values are the minimum bound for the first stage of Test C.

	A	B	C	D	E	F	G	H	I	J	K	L	M
1		Xbed											
2		RGH	STWAH	STWCH	YYF	RIHSC	YAB	GUH	CH	NHH	CCH	MVHSF	Other
3	**Accident and Emergency**	3	0	0	0	0	0	0	0	12	0	0	0
4	**Anaesthetics**	6	0	0	0	0	0	0	0	1	0	0	0
5	**Cardiology**	17	0	0	0	0	0	0	0	10	0	0	1
6	**Care of the Elderly**	95	0	0	58	0	1	0	9	47	0	0	0
7	**Community Medicine**	0	0	0	7	0	0	0	1	0	13	0	0
8	**Dermatology**	0	3	0	0	0	0	0	0	0	0	0	0
9	**Diabetes and Endocrinology**	15	0	0	22	0	0	0	0	18	0	0	0
10	**Ear, Nose and Throat**	4	0	0	1	0	0	0	0	0	0	0	0
11	**Gastroenterology**	0	14	0	1	0	0	0	0	21	0	0	1
12	**General Medicine**	86	0	0	2	0	1	0	0	15	0	0	0
13	**General Surgery**	48	0	0	1	0	0	0	0	23	0	0	1
14	**GP Other**	0	0	0	0	10	0	0	0	0	0	16	0
15	**Gynaecology**	3	0	0	1	0	0	0	0	2	0	0	1
16	**Haematology**	4	0	0	1	0	0	0	0	2	0	0	1
17	**Infectious Diseases**	8	0	0	0	0	0	0	0	0	0	0	0
18	**Intermediate Care**	0	0	0	0	0	1	0	1	0	0	0	0
19	**Maxillo-Facial**	0	2	0	0	0	0	0	0	1	0	0	0
20	**Neurology**	2	0	0	0	0	0	0	0	0	0	0	0
21	**Opthamology**	3	0	0	1	0	0	0	0	1	0	0	1
22	**Pain**	1	0	0	1	0	0	0	1	1	0	0	0
23	**Plastic Surgery**	0	0	0	0	0	0	0	0	1	0	0	1
24	**Radiology**	1	0	0	0	0	0	0	0	1	0	0	0
25	**Radiotherapy and Oncology**	1	0	0	0	0	0	0	0	0	0	0	0
26	**Rehabiltation**	0	0	63	33	0	69	0	32	12	13	0	0
27	**Respiratory**	30	0	0	0	0	0	0	0	28	0	0	0
28	**Restorative Dentistry**	1	0	0	0	0	0	0	0	0	0	0	0
29	**Rheumatology**	1	0	0	1	0	0	0	0	1	0	0	0
30	**Trauma & Orthopaedic**	52	9	0	1	0	0	0	0	42	0	0	0
31	**Urology**	15	0	0	1	0	0	0	0	1	0	0	1
32	**TOTAL BEDS**	396	28	63	132	10	72	0	44	240	26	16	8
33													
34	**Total Capacity xbed**	588	28	75	163	12	80	0	73	309	26	16	80
35													
36													
37	**Objective Function - Daily Cost: £ 976601.16**				**Total Beds First Stage: 1035**					**Total Staff First Stage: 414**			
38					**Max beds Second Stage: 19**					**Max beds Second Stage: 7**			
39													

Figure 7.18: The output from the deterministic model once solved, displaying the number of beds to deploy to each hospital and specialty. The total daily cost, beds and staff are also summarised.

7.3 Python Implementation

Python is a high-level programming language that can be used for a wide range of applications such as web development, data analysis, scientific computing, artificial intelligence, and automation. In this guide, a step by step tutorial is provided and the input requirements of the user will be discussed.

The deterministic and two-stage stochastic optimisation models have been provided on GitHub [287] to enable users to apply this to their own scenario. The model requires users to install the PuLP package [301], and these models were developed using PuLP version 2.3. Python contains a library named ‘itertools’, which is also required. If the user has a Gurobi license, then this also requires importing into Python. For the models to run, these two libraries are required to be imported, as follows:

```
import pulp
import itertools
import gurobipy
```

7.3.1 Deterministic Model

The Python models use functions to pass data through and stores variables until later required. The first function initialises the problem and sets the decision variables. The ‘pulp.LpProblem’ class creates a new linear programming problem with the name used within the output .lp file and the sense of the objective, whether this be a minimisation or maximisation. The ‘Lp.Variable’ term creates the decision variables and stores them within a dictionary. The lower bound of the variables is set to zero to ensure there are no non-negative constraints, and the values are set to integer. In this case, the ‘xbed’ variable is a two variable dictionary and the ‘xstaff’ is a three variable dictionary.

```
def initialise_deterministic_minimisation_problem(
    specialties, hospitals, bands
):
    """
    Initialise the mininmisation problem.
    Set decision variables for the models.
    """
    sh = [(s,h) for s in specialties for h in hospitals]
    shb = [(s,h,b) for s in specialties for h in hospitals for b in
    bands]

    prob = pulp.LpProblem("Deterministic", pulp.LpMinimize)

    xbed = pulp.LpVariable.dicts(
```

```
        "Xbed", (specialties, hospitals), lowBound=0, cat='Integer'
    )
    xstaff = pulp.LpVariable.dicts(
        "Xstaff", (specialties, hospitals, bands), lowBound=0,cat='Integer'
    )
    return prob, sh, shb, xbed, xstaff
```

Once the model is initialised, the deterministic constraints as listed in Section 4.3.4, can be inputted into the model. The deterministic model in total contains 10 constraints. PuLP uses the assignment operator '+=', to store the results of an expression to the 'prob' term. The class 'lp.Sum' is used to sum the list of linear expressions. For loops are used to cycle through all the elements in the list. This can be seen as follows in the 'add_deterministic_constraints' function:

```
def add_deterministic_constraints(xbed, xstaff, UBbed, UBstaff, D, K, R, sh, shb, prob):
    """
    Add the constraints that are required for the deterministic model

    - Constraints 1-6: Ensures demand is met across all specialties and all regions
    - Constraint 7: Ensures beds are only able to open in a ward if the facilities are able to be opened
    - Constraint 8: Ensures staffing ratios are met
    - Constraint 9: Ensures beds deployed does not exceed maximum capacity of hospital
    - Constraint 10: Ensures staff deployed does not exceed maximum staffing resources
    """

    for s in specialties:
        prob += pulp.lpSum(xbed[s][h] for h in region1) >= pulp.lpSum(D[s][0]) #Constraint 1
        prob += pulp.lpSum(xbed[s][h] for h in region2) >= pulp.lpSum(D[s][1]) #Constraint 2
        prob += pulp.lpSum(xbed[s][h] for h in region3) >= pulp.lpSum(D[s][2]) #Constraint 3
        prob += pulp.lpSum(xbed[s][h] for h in region4) >= pulp.lpSum(D[s][3]) #Constraint 4
        prob += pulp.lpSum(xbed[s][h] for h in region5) >= pulp.lpSum(D[s][4]) #Constraint 5
        prob += pulp.lpSum(xbed[s][h] for h in region6) >= pulp.lpSum(D[s][5]) #Constraint 6

        for h in hospitals:
```

```
            prob += pulp.lpSum(xbed[s][h])<= pulp.lpSum(K[s][h]) # Constraint 7

            for b in bands:
                prob += pulp.lpSum(xstaff[s][h][b]) >= pulp.lpSum(R[s][b]*(xbed[s][h])) #Constraint 8

    for h in hospitals:
        prob += pulp.lpSum(xbed[s][h] for s in specialties) <= UBbed[h]  #Constraint 9

    for b in bands:
        prob += pulp.lpSum(xstaff[s][h][b] for (s,h) in sh) <= UBstaff[b] # #Constraint 10

    return prob
```

The final function solves the deterministic model by calling all the previous functions. This is where the objective function of the model is defined and has to satisfy the constraints, which are called from the previous function. The model is solved using the 'prob.solve()' term, where the solver; COIN-OR CBC linear solver, is selected. The 'maxSeconds' determines the maximum time for the solver in seconds, whilst the 'fracgap' sets the tolerance for the solver to stop.

```
def solve_deterministic_minimisation_problem(
    specialties,
    bands,
    hospitals,
    regions,
    D,
    K,
    R,
    cbed,
    cstaff,
    UBbed,
    UBstaff,
):

    """
    Solves the deterministic problem, with the objective function being minimised.
    """
    prob, sh, shb, xbed, xstaff = initialise_deterministic_minimisation_problem(
        specialties=specialties,
        hospitals=hospitals,
        bands=bands
```

```
    )

    prob += (
        pulp.lpSum((xbed[s][h] * cbed[s][h]) for (s,h) in sh) +
        pulp.lpSum((xstaff[s][h][b]*cstaff[b]) for (s,h,b) in shb)
    )

    prob = add_deterministic_constraints(
        xbed=xbed,
        xstaff=xstaff,
        UBbed=UBbed,
        UBstaff=UBstaff,
        D=D,
        K=K,
        R=R,
        sh=sh,
        shb=shb,
        prob=prob,
    )
    # The user can select one of the two optimisers:
    # prob.solve(pulp.GUROBI())
    # prob.solve(pulp.PULP_CBC_CMD())
    return prob
```

The values for the parameters are then required to be entered by the user. The total number of specialties, bands and regions can be entered manually and using the itertools package, lists are created. For each region, the hospitals are inputted into each array. For example, if region one had the first three hospitals the line 'region1 = [0,1,2]' would be displayed. All regions are then summed together to determine the total number of hospitals within the health board. A two-dimensional demand array, D, is required by the user. Each row represents a specialty and each column represents the region. Similarly, the K value depicting how many beds can be deployed within each specialty and each hospital can also be inputted. Again, each row represents a specialty, with the column representing the hospital. Next, the ratios of each band of staff to specialty is required. The row of the array is representing specialty and the columns represent each band of nurse. The variable 'cbed' relates to the cost of a bed per specialty in each hospital. The structure of this is identical to the K array. The 'cstaff' parameter is a one-dimensional matrix where each entry relates to the cost of each band of nurse. Similarly, the 'UBstaff' parameter is the maximum number of nurses for each band that are able to be deployed. Finally, the 'UBubed' is the maximum number of hospital beds that can be deployed and each entry represents each hospital. The user can select one of the two solvers, either the CBC_CMD solver: `prob.solve(pulp.PULP_CBC_CMD())` or the Gurobi solver: `prob.solve(pulp.GUROBI())`.

```
    """
    These values can be altered to the specific requirements of the
    user
    """
    specialties = list(itertools.chain(range(0, ))) #Creates list
   of specialties
    bands = list(itertools.chain(range(0, ))) #Creates list of
   nursing bands
    regions = list(itertools.chain(range(0, ))) #Creates List of
   regions

    region1 = []
    region2 = []
    region3 = []
    region4 = []
    region5 = []
    region6 = []
    hospitals = region1 + region2 +region3 + region4 +region5 +
   region6
    D = [
    [],
    ]
    K = [
    [],
    ]
    R = [
    [],
    ]
    cbed = [
    [],
    ]
    cstaff = []
    UBstaff = [
    [],
    ]
    UBbed =[
    [],
    ]
```

In order for the optimisation to be solved, the user entered parameters are required to be fed into the model. This is computed by the following code:

```
    """
    Feeds the parameters into the deterministic optimisation model
    """
    prob = solve_deterministic_minimisation_problem(
    specialties,
    bands,
```

```
    hospitals,
    regions,
    D,
    K,
    R,
    cbed,
    cstaff,
    UBbed,
    UBstaff,
)
```

In order to ensure the model has been solved, LpStatus, returns the status of the problem, either suggesting an optimal solution has been found or the model is infeasible. The total objective function can also be displayed using the 'value' parameter. To display all non-zero decision variables, a 'for' loop has been created which prints out the decision variable name along with the number of beds or staff to deploy.

```
    print("Solution Status = ", pulp.LpStatus[prob.status])
    print("Total price = ", pulp.value(prob.objective))
    for v in prob.variables():
        if v.varValue >= 0:
            print(v.name, "=", v.varValue)
```

7.3.2 Two-Stage Stochastic Model

The two-stage stochastic model follows a similar structure to the deterministic model with similar functions used to develop the model. Within this section, the differences between the two models will be discussed. The first function is named the 'initialise_stochastic_minimisation_problem' where the first and second stage decision variables are set and the problem is initialised. The second stage variable 'ubed' is a three variable dictionary which includes specialties, hospitals and bands. The 'ustaff' decision variable is a four variable dictionary with parameters specialties, hospitals, bands and scenarios.

```
def initialise_stochastic_minimisation_problem(
    specialties, hospitals, bands, regions, scenarios
):
    """
    Initialise the minimisation problem.
    Set decision variables for the models.
    """
    sh = [(s,h) for s in specialties for h in hospitals]
    shb = [(s,h,b) for s in specialties for h in hospitals for b in
    bands]
```

```
    shk = [(s,h,k) for s in specialties for h in hospitals for k in scenarios]
    srhk = [(s,r,h,k) for s in specialties for r in regions for h in hospitals for k in scenarios]
    sbhk = [(s,b,h,k) for s in specialties for b in bands for h in hospitals for k in scenarios]

    prob = pulp.LpProblem("Stochastic", pulp.LpMinimize)

    xbed = pulp.LpVariable.dicts(
        "Xbed", (specialties,hospitals), lowBound=0, cat = 'Integer'
    )
    xstaff = pulp.LpVariable.dicts(
        "Xstaff", (specialties,hospitals,bands), lowBound=0,cat = 'Integer'
    )
    ubed = pulp.LpVariable.dicts(
        "Ubed",(specialties,hospitals,scenarios), lowBound=0, cat='Integer'
    )
    ustaff = pulp.LpVariable.dicts(
        "Ustaff",(specialties,hospitals,bands,scenarios), lowBound=0, cat='Integer'
    )
    return prob, sh, shb, shk, srhk, sbhk, xbed, xstaff, ubed, ustaff
```

The two-stage stochastic modelling constraints as generated in Section 4.4.4, can be initialised into the model. The two-stage stochastic model has a total of 14 constraints, an additional four compared to the deterministic model. Constraint 8 ensures the 'ubed' deployment is under capacity for each specialty ward in each hospital. Constraint 10 enables the patient to nurse ratio to be met. Constraints 12 and 14 ensure the deployment of beds and staff are below the maximum capacity for each.

```
def add_stochastic_constraints(
    xbed, xstaff, ubed, ustaff, UBbed, UBstaff, UBubed, UBustaff, D, R, K, prob, sh, shb, shk, srhk, sbhk
):

    """
    Add the constraints that are required for the stochastic model

    - Constraints 1-6: Ensures demand is met across all specialties and all regions
    - Constraint 7: Ensures beds are only able to open in a ward if
```

```
     the facilities are able to be opened - 1st stage
    - Constraint 8: Ensures beds are only able to open in a ward if the facilities are able to be opened - 2nd stage
    - Constraint 9: Ensures staffing ratios are met in the first stage
    - Constraint 10: Ensures staffing ratios are met in the first stage
    - Constraint 11: Ensures beds deployed does not exceed maximum capacity of hospital - 1st stage
    - Constraint 12: Ensures beds deployed does not exceed maximum capacity of hospital - 2nd stage
    - Constraint 13: Ensures staff deployed does not exceed maximum staffing resources - 1st stage
    - Constraint 14: Ensures staff deployed does not exceed maximum staffing resources - 2nd stage
    """

    for k in scenarios:
        for s in specialties:
            prob += pulp.lpSum(xbed[s][h] + ubed[s][h][k] for h in region1) >= pulp.lpSum(D[s][0][k]) #Constraint 1
            prob += pulp.lpSum(xbed[s][h] + ubed[s][h][k] for h in region2) >= pulp.lpSum(D[s][1][k]) #Constraint 2
            prob += pulp.lpSum(xbed[s][h] + ubed[s][h][k] for h in region3) >= pulp.lpSum(D[s][2][k]) #Constraint 3
            prob += pulp.lpSum(xbed[s][h] + ubed[s][h][k] for h in region4) >= pulp.lpSum(D[s][3][k]) #Constraint 4
            prob += pulp.lpSum(xbed[s][h] + ubed[s][h][k] for h in region5) >= pulp.lpSum(D[s][4][k]) #Constraint 5
            prob += pulp.lpSum(xbed[s][h] + ubed[s][h][k] for h in region6) >= pulp.lpSum(D[s][5][k]) #Constraint 6

    for s in specialties:
        for h in hospitals:
            prob += pulp.lpSum(xbed[s][h]) <= pulp.lpSum(K[s][h]) #Constraint 7

    for s in specialties:
        for h in hospitals:
            prob += pulp.lpSum(ubed[s][h][k] for k in scenarios)<= pulp.lpSum(K[s][h]) #Constraint 8

            for b in bands:
                prob += pulp.lpSum(xstaff[s][h][b]) >= pulp.lpSum(R[s][b]*(xbed[s][h])) #Constraint 9

                for k in scenarios:
                    prob += pulp.lpSum(ustaff[s][h][b][k])>= pulp.
```

```
   lpSum(R[s][b]*(ubed[s][h][k])) #Constraint 10

    for h in hospitals:
        prob += pulp.lpSum(xbed[s][h] for s in specialties) <= UBbed[h] #Constraint 11

    for k in scenarios:
        for h in hospitals:
            prob += pulp.lpSum(ubed[s][h][k] for s in specialties) <=UBubed[h][k] #Constraint 12

    for b in bands:
        prob += pulp.lpSum(xstaff[s][h][b] for (s,h) in sh) <= UBstaff[b] #Constraint 13

        for k in scenarios:
            prob += pulp.lpSum(ustaff[s][h][b][k] for (s,h) in sh) <= UBustaff[b][k] #Constraint 14

    return prob
```

The final stochastic function solves the two-stage stochastic model by calling the two previous functions. The objective function is defined and stored within the prob variable. The solver COIN-OR CBC linear solver or the Gurobi solver can once again be selected using the `prob.solve(pulp.PULP_CBC_CMD())` or `prob.solve(pulp.GUROBI())` command, respectively.

```
def solve_stochastic_minimisation_problem(
    specialties,
    bands,
    hospitals,
    regions,
    scenarios,
    pscenarios,
    D,
    R,
    K,
    c1bed,
    c2bed,
    c1staff,
    c2staff,
    UBbed,
    UBstaff,
):
    """
    Solves the stochastic problem, with the objective function being minimised.
    """
```

```
    prob, sh, shb, shk, srhk, sbhk, xbed, xstaff, ubed, ustaff = initialise_stochastic_minimisation_problem(
        specialties=specialties,
        hospitals=hospitals,
        bands=bands,
        regions=regions,
        scenarios=scenarios
    )
    prob +=(
        pulp.lpSum((xbed[s][h]*c1bed[s][h]) for (s,h) in sh) +
        pulp.lpSum((xstaff[s][h][b]*c1staff[b]) for (s,h,b) in shb) +
        pulp.lpSum(pscenarios[k]*(ubed[s][h][k]*c2bed[s][h]) for (s,h,k) in shk)+
        pulp.lpSum(pscenarios[k]*(ustaff[s][h][b][k]*c2staff[b]) for (s,b,h,k) in sbhk)
    )

    prob = add_stochastic_constraints(
        xbed=xbed,
        xstaff=xstaff,
        ubed=ubed,
        ustaff=ustaff,
        UBbed=UBbed,
        UBstaff=UBstaff,
        UBubed=UBubed,
        UBustaff=UBustaff,
        D=D,
        R=R,
        K=K,
        sh=sh,
        shb=shb,
        shk=shk,
        srhk=srhk,
        sbhk=sbhk,
        prob=prob,
    )
    # The user can select one of the two optimisers:
    # prob.solve(pulp.GUROBI())
    # prob.solve(pulp.PULP_CBC_CMD())
    return prob
```

In addition to the deterministic model, an additional six parameters are required. The second stage costs for beds (c2bed) and staff (c2staff) are generated with two-dimensional and one-dimensional arrays, respectively. The upper bounds for beds and staff are generated using two-dimensional arrays where a column represents each scenario and the row represents either the hospital or nursing bands depending on the parameter. The number of scenarios can also be generated into an array and then the probability of each of these scenarios occurring. The demand array requires converting to a three-dimensional array since the demand is scenario dependent. For this, each row represents a specialty, and each column represents a region. The scenario is determined by the column inside each of the arrays.

```
"""
These values can be altered to the specific requirements of the user
"""
specialties = list(itertools.chain(range(0, ))) #Creates list of specialties
bands = list(itertools.chain(range(0, ))) #Creates list of nursing bands
regions = list(itertools.chain(range(0, ))) #Creates List of regions

region1 = []
region2 = []
region3 = []
region4 = []
region5 = []
region6 = []
hospitals = region1 + region2 +region3 + region4 +region5 + region6
D = [
[[],[]],
]
K = [
[],
]
R = [
[],
]
c1bed = [
[],
]
c2bed = [
[],
]
c1staff = []
c2staff = []
UBstaff = [
[],
]
UBustaff = [
[],
]
```

```
    UBbed = [
    [],
    ]
    UBubed =[
    [],
    ]
    scenarios = []
    pscenarios = []
```

The optimisation model is then solved by feeding the parameters entered by the user into the following function:

```
    """
    Feeds the parameters into the two-stage stochastic optimisation
    model
    """
    prob = solve_stochastic_minimisation_problem(
    specialties,
    bands,
    hospitals,
    regions,
    scenarios,
    pscenarios,
    D,
    R,
    K,
    c1bed,
    c2bed,
    c1staff,
    c2staff,
    UBbed,
    UBstaff
)
```

As with the deterministic model, the results can then be outputted to the user, where the status and overall objective functions are printed for the user. Additionally, the non-zero decision variables are printed for the user.

```
    print("Solution Status = ", pulp.LpStatus[prob.status])
    print("Total price = ", pulp.value(prob.objective))
    for v in prob.variables():
        if v.varValue >= 0:
            print(v.name, "=", v.varValue)
```

7.3.3 Test A

The Test A model follows an almost identical structure to the two-stage stochastic model discussed in Section 7.3.2. This section will provide an updated code.

```python
def initialise_testa_minimisation_problem(
    specialties, hospitals, bands, regions, scenarios
):
    """
    Initialise the minimisation problem.
    Set decision variables for the models.
    """
    sh = [(s,h) for s in specialties for h in hospitals]
    shb = [(s,h,b) for s in specialties for h in hospitals for b in bands]
    shk = [(s,h,k) for s in specialties for h in hospitals for k in scenarios]
    srhk = [(s,r,h,k) for s in specialties for r in regions for h in hospitals for k in scenarios]
    sbhk = [(s,b,h,k) for s in specialties for b in bands for h in hospitals for k in scenarios]

    prob = pulp.LpProblem("Test A", pulp.LpMinimize)

    xstaff = pulp.LpVariable.dicts(
        "Xstaff", (specialties,hospitals,bands), lowBound=0,cat = 'Integer'
    )
    ubed = pulp.LpVariable.dicts(
        "Ubed",(specialties,hospitals,scenarios), lowBound=0, cat='Integer'
    )
    ustaff = pulp.LpVariable.dicts(
        "Ustaff",(specialties,hospitals,bands,scenarios), lowBound=0, cat='Integer'
    )
    return prob, sh, shb, shk, srhk, sbhk, xbed, xstaff, ubed, ustaff
```

Test A has a total of 12 constraints since the xbed dependent constraints have been removed. The user could also remove the xstaff constraints and add this as a separate variable, since these have already been predetermined. Due to the nature of the code, the xstaff numbers will remain consistent regardless of the method chosen. The remainder of the constraints remain the same.

```python
def add_testa_constraints(
    xbed, xstaff, ubed, ustaff, UBbed, UBstaff, UBubed, UBustaff, D, R, K, prob, sh, shb, shk, srhk, sbhk
):

    """
    Add the constraints that are required for the Test A model

```

```
    - Constraints 1-6: Ensures demand is met across all specialties and all regions
    - Constraint 7: Ensures beds are only able to open in a ward if the facilities are able to be opened - 2nd stage
    - Constraint 8: Ensures staffing ratios are met in the first stage
    - Constraint 9: Ensures staffing ratios are met in the second stage
    - Constraint 10 Ensures beds deployed does not exceed maximum capacity of hospital - 2nd stage
    - Constraint 11 Ensures staff deployed does not exceed maximum staffing resources - 1st stage
    - Constraint 12 Ensures staff deployed does not exceed maximum staffing resources - 2nd stage
      """

    for k in scenarios:
        for s in specialties:
            prob += pulp.lpSum(xbed[s][h] + ubed[s][h][k] for h in region1) >= pulp.lpSum(D[s][0][k]) #Constraint 1
            prob += pulp.lpSum(xbed[s][h] + ubed[s][h][k] for h in region2) >= pulp.lpSum(D[s][1][k]) #Constraint 2
            prob += pulp.lpSum(xbed[s][h] + ubed[s][h][k] for h in region3) >= pulp.lpSum(D[s][2][k]) #Constraint 3
            prob += pulp.lpSum(xbed[s][h] + ubed[s][h][k] for h in region4) >= pulp.lpSum(D[s][3][k]) #Constraint 4
            prob += pulp.lpSum(xbed[s][h] + ubed[s][h][k] for h in region5) >= pulp.lpSum(D[s][4][k]) #Constraint 5
            prob += pulp.lpSum(xbed[s][h] + ubed[s][h][k] for h in region6) >= pulp.lpSum(D[s][5][k]) #Constraint 6

    for s in specialties:
        for h in hospitals:
            prob += pulp.lpSum(ubed[s][h][k] for k in scenarios)<= pulp.lpSum(K[s][h]) #Constraint 7

            for b in bands:
                prob += pulp.lpSum(xstaff[s][h][b]) >= pulp.lpSum(R[s][b]*(xbed[s][h])) #Constraint 8

                for k in scenarios:
                    prob += pulp.lpSum(ustaff[s][h][b][k])>= pulp.lpSum(R[s][b]*(ubed[s][h][k])) #Constraint 9

    for k in scenarios:
        for h in hospitals:
            prob += pulp.lpSum(ubed[s][h][k] for s in specialties) <=UBubed[h][k] #Constraint 10
```

```

    for b in bands:
        prob += pulp.lpSum(xstaff[s][h][b] for (s,h) in sh) <= UBstaff[b] #Constraint 11

        for k in scenarios:
            prob += pulp.lpSum(ustaff[s][h][b][k] for (s,h) in sh) <= UBustaff[b][k] #Constraint 12

    return prob
```

Similar to the prior examples, the final function solves the optimisation problem by calling the two previous functions. The COIN-OR CBC linear solver or the Gurobi solver can once again be selected using the `prob.solve(pulp.PULP_CBC_CMD())` or `prob.solve(pulp.GUROBI())` command, respectively.

```
    def solve_testa_minimisation_problem(
    specialties,
    bands,
    hospitals,
    regions,
    scenarios,
    pscenarios,
    D,
    R,
    K,
    c1bed,
    c2bed,
    c1staff,
    c2staff,
    UBbed,
    UBstaff,
):
    """
    Solves the Test A problem, with the objective function being minimised.
    """
    prob, sh, shb, shk, srhk, sbhk, xbed, xstaff, ubed, ustaff = initialise_testa_minimisation_problem(
        specialties=specialties,
        hospitals=hospitals,
        bands=bands,
        regions=regions,
        scenarios=scenarios
    )
    prob +=(
        pulp.lpSum((xbed[s][h]*c1bed[s][h]) for (s,h) in sh) +
        pulp.lpSum((xstaff[s][h][b]*c1staff[b]) for (s,h,b) in shb)
```

```
    +
        pulp.lpSum(pscenarios[k]*(ubed[s][h][k]*c2bed[s][h]) for (s
   ,h,k) in shk)+
        pulp.lpSum(pscenarios[k]*(ustaff[s][h][b][k]*c2staff[b])
   for (s,b,h,k) in sbhk)
    )

    prob = add_testa_constraints(
        xbed=xbed,
        xstaff=xstaff,
        ubed=ubed,
        ustaff=ustaff,
        UBbed=UBbed,
        UBstaff=UBstaff,
        UBubed=UBubed,
        UBustaff=UBustaff,
        D=D,
        R=R,
        K=K,
        sh=sh,
        shb=shb,
        shk=shk,
        srhk=srhk,
        sbhk=sbhk,
        prob=prob,
    )
    # The user can select one of the two optimisers:
    # prob.solve(pulp.GUROBI())
    # prob.solve(pulp.PULP_CBC_CMD())
    return prob
```

In addition to the two-stage stochastic model, the xbed values are required to be entered in the form of a three-dimensional array, where the specialties are the rows and the hospitals are the columns. These can either be manually entered or the following code can be used to manipulate the results from the deterministic model:

```
    import pandas as pd
    import numpy as np
    b = [] #Create an empty array
    for v in prob.variables():
        if v.name[0:4] == "Xbed": #Filter xbed variables only
            b.append(v.varValue) #Assign values to b
    df = pd.DataFrame(np.zeros((29,12))) #Create a 3d array with
   zeros
    for j in range(0,29): #Iterate over each row
        df.iloc[j] = b[j*12:(j+1)*12] #Assign columns
    df = df[[0,1,4,5,6,7,8,9,10,11,2,3]] #Reorder the columns
    df = df.reindex([0,11,21,22,23,24,25,26,27,28,1,2
```

```
,3,4,5,6,7,8,9,10,12,13,14,15,16,17,18,19,20]) #Reorder the
rows
df.replace(-0,0, inplace=True)
xbed = df.values
```

The xbed array can then be used within the following data requirements:

```
"""
These values can be altered to the specific requirements of the
user
"""
specialties = list(itertools.chain(range(0, ))) #Creates list
of specialties
bands = list(itertools.chain(range(0, ))) #Creates list of
nursing bands
regions = list(itertools.chain(range(0, ))) #Creates List of
regions

region1 = []
region2 = []
region3 = []
region4 = []
region5 = []
region6 = []
hospitals = region1 + region2 +region3 + region4 +region5 +
region6
D = [
[[],[]],
]
K = [
[],
]
R = [
[],
]
c1bed = [
[],
]
c2bed = [
[],
]
c1staff = []
c2staff = []
UBstaff = [
[],
]
UBustaff = [
[],
```

```
    ]
    UBbed = [
    [],
    ]
    UBubed =[
    [],
    ]
    scenarios = []
    pscenarios = []
    xbed=[
    [],
    ]
```

The optimisation model is then solved by feeding the parameters entered by the user into the following function:

```
    prob = solve_testa_minimisation_problem(
    specialties,
    bands,
    hospitals,
    regions,
    scenarios,
    pscenarios,
    D,
    R,
    K,
    c1bed,
    c2bed,
    c1staff,
    c2staff,
    UBbed,
    UBstaff
)
```

As with the previous two implementations, the results can be outputted to the user displaying the values for each of the decision variables and the overall objective function.

```
    print("Solution Status = ", pulp.LpStatus[prob.status])
    print("Total price = ", pulp.value(prob.objective))
    for v in prob.variables():
        if v.varValue >=0:
            print(v.name, "=", v.varValue)
```

7.3.4 Test B

Similar to Test A, Test B also follows a nearly identical structure to that of the two-stage stochastic implementation. The function `initialise_stochastic_minimisation_problem` is employed, as the same set of four decision variables are utilised within the model.

The function `add_stochastic_constraints`, is changed since Test B requires an additional constraint of the lowest stage variables are set to zero or their lower bound. Therefore a new function called `add_testb_constraints` is generated, with the additional constraint added:

```python
def add_testb_constraints(
    xbed, xstaff, ubed, ustaff, UBbed, UBstaff, UBubed, UBustaff, D, R, K, prob, sh, shb, shk, srhk, sbhk, TESTB
):

    """
    Add the constraints that are required for the Test B model

    - Constraints 1-6: Ensures demand is met across all specialties and all regions
    - Constraint 7: Ensures beds are only able to open in a ward if the facilities are able to be opened - 1st stage
    - Constraint 8: Ensures beds are only able to open in a ward if the facilities are able to be opened - 2nd stage
    - Constraint 9: Ensures staffing ratios are met in the first stage
    - Constraint 10: Ensures staffing ratios are met in the first stage
    - Constraint 11: Ensures beds deployed does not exceed maximum capacity of hospital - 1st stage
    - Constraint 12: Ensures beds deployed does not exceed maximum capacity of hospital - 2nd stage
    - Constraint 13: Ensures staff deployed does not exceed maximum staffing resources - 1st stage
    - Constraint 14: Ensures staff deployed does not exceed maximum staffing resources - 2nd stage
    - Constraint 15: Ensures the xbed does not exceed the the lower bound of the deterministic model
      """

    for k in scenarios:
        for s in specialties:
            prob += pulp.lpSum(xbed[s][h] + ubed[s][h][k] for h in region1) >= pulp.lpSum(D[s][0][k]) #Constraint 1
            prob += pulp.lpSum(xbed[s][h] + ubed[s][h][k] for h in region2) >= pulp.lpSum(D[s][1][k]) #Constraint 2
            prob += pulp.lpSum(xbed[s][h] + ubed[s][h][k] for h in region3) >= pulp.lpSum(D[s][2][k]) #Constraint 3
            prob += pulp.lpSum(xbed[s][h] + ubed[s][h][k] for h in region4) >= pulp.lpSum(D[s][3][k]) #Constraint 4
```

```
            prob += pulp.lpSum(xbed[s][h] + ubed[s][h][k] for h in region5) >= pulp.lpSum(D[s][4][k]) #Constraint 5
            prob += pulp.lpSum(xbed[s][h] + ubed[s][h][k] for h in region6) >= pulp.lpSum(D[s][5][k]) #Constraint 6

    for s in specialties:
        for h in hospitals:
            prob += pulp.lpSum(xbed[s][h]) <= pulp.lpSum(K[s][h]) #Constraint 7

    for s in specialties:
        for h in hospitals:
            prob += pulp.lpSum(ubed[s][h][k] for k in scenarios)<= pulp.lpSum(K[s][h]) #Constraint 8

            for b in bands:
                prob += pulp.lpSum(xstaff[s][h][b]) >= pulp.lpSum(R[s][b]*(xbed[s][h])) #Constraint 9

                for k in scenarios:
                    prob += pulp.lpSum(ustaff[s][h][b][k])>= pulp.lpSum(R[s][b]*(ubed[s][h][k])) #Constraint 10

    for h in hospitals:
        prob += pulp.lpSum(xbed[s][h] for s in specialties) <= UBbed[h] #Constraint 11

    for k in scenarios:
        for h in hospitals:
            prob += pulp.lpSum(ubed[s][h][k] for s in specialties) <=UBubed[h][k] #Constraint 12

    for b in bands:
        prob += pulp.lpSum(xstaff[s][h][b] for (s,h) in sh) <= UBstaff[b] #Constraint 13

        for k in scenarios:
            prob += pulp.lpSum(ustaff[s][h][b][k] for (s,h) in sh) <= UBustaff[b][k] #Constraint 14
    for s in specialties:
        for h in hospitals:
            prob += pulp.lpSum(xbed[s][h]) <= TESTB[s][h] #Constraint 15
    return prob
```

The final function solves the optimisation programme by calling the two previous functions. The user can select the optimiser they wish to use for the calculations.

```
def solve_testb_minimisation_problem(
    specialties,
    bands,
    hospitals,
    regions,
    scenarios,
    pscenarios,
    D,
    R,
    K,
    c1bed,
    c2bed,
    c1staff,
    c2staff,
    UBbed,
    UBstaff,
    TESTB
):
    """
    Solves the deterministic problem, with the objective function being minimised.
    """
    prob, sh, shb, shk, srhk, sbhk, xbed, xstaff, ubed, ustaff = initialise_stochastic_minimisation_problem(
        specialties=specialties,
        hospitals=hospitals,
        bands=bands,
        regions=regions,
        scenarios=scenarios
    )
    prob +=(
        pulp.lpSum((xbed[s][h]*c1bed[s][h]) for (s,h) in sh) +
        pulp.lpSum((xstaff[s][h][b]*c1staff[b]) for (s,h,b) in shb) +
        pulp.lpSum(pscenarios[k]*(ubed[s][h][k]*c2bed[s][h]) for (s,h,k) in shk)+
        pulp.lpSum(pscenarios[k]*(ustaff[s][h][b][k]*c2staff[b]) for (s,b,h,k) in sbhk)
    )

    prob = add_testb_constraints(
        xbed=xbed,
        xstaff=xstaff,
        ubed=ubed,
        ustaff=ustaff,
        UBbed=UBbed,
        UBstaff=UBstaff,
        UBubed=UBubed,
```

```
        UBustaff=UBustaff,
        D=D,
        R=R,
        K=K,
        sh=sh,
        shb=shb,
        shk=shk,
        srhk=srhk,
        sbhk=sbhk,
        prob=prob,
        TESTB=TESTB
    )
    # The user can select one of the two optimisers:
    # prob.solve(pulp.GUROBI())
    # prob.solve(pulp.PULP_CBC_CMD())
    return prob
```

A new three-dimensional array is required to be created, either manually or via manipulation of the deterministic results to produce TESTB.

The following code produced a new array, TESTB from the deterministic model, based on the fact that there will be at least one zero in the lower bound of the model. If this is not the case, the user is required to manually enter the values into the TESTB array instead:

```
import pandas as pd
import numpy as np
b = []
for v in prob.variables():
    if v.name[0:4] == "Xbed":
        b.append(v.varValue)
df =pd.DataFrame(np.zeros((29,12)))
K_array = pd.DataFrame(K, columns=df.columns, index=df.index) # Turns the array K into a dataframe
for i, row in df.iterrows():
    for col in df.columns:
        if df.at[i, col] > 0:
        # Replace the value with the corresponding value from dataset K
            df.at[i, col] = K_array.at[i, col]
TESTB = df
```

The new three-dimensional array can then be implemented into the following data requirements for the model.

```
"""
These values can be altered to the specific requirements of the user
```

```
 """
 specialties = list(itertools.chain(range(0, ))) #Creates list
of specialties
 bands = list(itertools.chain(range(0, ))) #Creates list of
nursing bands
 regions = list(itertools.chain(range(0, ))) #Creates List of
regions

 region1 = []
 region2 = []
 region3 = []
 region4 = []
 region5 = []
 region6 = []
 hospitals = region1 + region2 +region3 + region4 +region5 +
region6
 D = [
 [[],[]],
 ]
 K = [
 [],
 ]
 R = [
 [],
 ]
 c1bed = [
 [],
 ]
 c2bed = [
 [],
 ]
 c1staff = []
 c2staff = []
 UBstaff = [
 [],
 ]
 UBustaff = [
 [],
 ]
 UBbed = [
 [],
 ]
 UBubed =[
 [],
 ]
 scenarios = []
 pscenarios = []
 TESTB=[
```

```
    [],
    ]
```

The following code will then solve the model by feeding in the created parameters into the
`solve_testa_minimisation_problem` function.

```
    prob = solve_testc_minimisation_problem(
    specialties,
    bands,
    hospitals,
    regions,
    scenarios,
    pscenarios,
    D,
    R,
    K,
    c1bed,
    c2bed,
    c1staff,
    c2staff,
    UBbed,
    UBstaff,
    TESTB
)
```

As previously, the results can then be outputted for analysis and comparisons.

```
    print("Solution Status = ", pulp.LpStatus[prob.status])
    print("Total price = ", pulp.value(prob.objective))
    for v in prob.variables():
        if v.varValue >=0:
            print(v.name, "=", v.varValue)
```

7.3.5 Test C

Similar to Test A and B, Test B also follows a nearly identical structure to that of the two-stage stochastic implementation. The function
`initialise_stochastic_minimisation_problem` is employed, as the same set of four decision variables are utilised within the model.

The function `add_stochastic_constraints`, is changed since Test C requires an additional constraint of the deterministic solution to act as a minimum value for the first stage. A new function called `add_testc_constraints` is generated, with an additional constraint added:

```
def add_testc_constraints(
```

```
    xbed, xstaff, ubed, ustaff, UBbed, UBstaff, UBubed, UBustaff, D, R, K, prob, sh, shb, shk, srhk, sbhk, First_stage
):

    """
    Add the constraints that are required for the Test C model

    - Constraints 1-6: Ensures demand is met across all specialties and all regions
    - Constraint 7: Ensures beds are only able to open in a ward if the facilities are able to be opened - 1st stage
    - Constraint 8: Ensures beds are only able to open in a ward if the facilities are able to be opened - 2nd stage
    - Constraint 9: Ensures staffing ratios are met in the first stage
    - Constraint 10: Ensures staffing ratios are met in the second stage
    - Constraint 11: Ensures beds deployed does not exceed maximum capacity of hospital - 1st stage
    - Constraint 12: Ensures beds deployed does not exceed maximum capacity of hospital - 2nd stage
    - Constraint 13: Ensures staff deployed does not exceed maximum staffing resources - 1st stage
    - Constraint 14: Ensures staff deployed does not exceed maximum staffing resources - 2nd stage
    - Constraint 15: Deterministic values must be met as a minimum for the first stage
      """

    for k in scenarios:
        for s in specialties:
            prob += pulp.lpSum(xbed[s][h] + ubed[s][h][k] for h in region1) >= pulp.lpSum(D[s][0][k]) #Constraint 1
            prob += pulp.lpSum(xbed[s][h] + ubed[s][h][k] for h in region2) >= pulp.lpSum(D[s][1][k]) #Constraint 2
            prob += pulp.lpSum(xbed[s][h] + ubed[s][h][k] for h in region3) >= pulp.lpSum(D[s][2][k]) #Constraint 3
            prob += pulp.lpSum(xbed[s][h] + ubed[s][h][k] for h in region4) >= pulp.lpSum(D[s][3][k]) #Constraint 4
            prob += pulp.lpSum(xbed[s][h] + ubed[s][h][k] for h in region5) >= pulp.lpSum(D[s][4][k]) #Constraint 5
            prob += pulp.lpSum(xbed[s][h] + ubed[s][h][k] for h in region6) >= pulp.lpSum(D[s][5][k]) #Constraint 6

    for s in specialties:
        for h in hospitals:
            prob += pulp.lpSum(xbed[s][h]) <= pulp.lpSum(K[s][h]) #Constraint 7
```

```

    for s in specialties:
        for h in hospitals:
            prob += pulp.lpSum(ubed[s][h][k] for k in scenarios)<= pulp.lpSum(K[s][h]) #Constraint 8

            for b in bands:
                prob += pulp.lpSum(xstaff[s][h][b]) >= pulp.lpSum(R[s][b]*(xbed[s][h])) #Constraint 9

                for k in scenarios:
                    prob += pulp.lpSum(ustaff[s][h][b][k])>= pulp.lpSum(R[s][b]*(ubed[s][h][k])) #Constraint 10

    for h in hospitals:
        prob += pulp.lpSum(xbed[s][h] for s in specialties) <= UBbed[h] #Constraint 11

    for k in scenarios:
        for h in hospitals:
            prob += pulp.lpSum(ubed[s][h][k] for s in specialties) <=UBubed[h][k] #Constraint 12

    for b in bands:
        prob += pulp.lpSum(xstaff[s][h][b] for (s,h) in sh) <= UBstaff[b] #Constraint 13

        for k in scenarios:
            prob += pulp.lpSum(ustaff[s][h][b][k] for (s,h) in sh) <= UBustaff[b][k] #Constraint 14
    for s in specialties:
        for h in hospitals:
            prob += pulp.lpSum(xbed[s][h]) >= First_stage[s][h]
    return prob
```

The next function takes the previous two functions and optimises them based on the objective function given.

```

def solve_testc_minimisation_problem(
    specialties,
    bands,
    hospitals,
    regions,
    scenarios,
    pscenarios,
    D,
    R,
```

```
    K,
    c1bed,
    c2bed,
    c1staff,
    c2staff,
    UBbed,
    UBstaff,
    First_stage
):
    """
    Solves the Test C problem, with the objective function being minimised
    """
    prob, sh, shb, shk, srhk, sbhk, xbed, xstaff, ubed, ustaff = initialise_stochastic_minimisation_problem(
        specialties=specialties,
        hospitals=hospitals,
        bands=bands,
        regions=regions,
        scenarios=scenarios
    )
    prob +=(
        pulp.lpSum((xbed[s][h]*c1bed[s][h]) for (s,h) in sh) +
        pulp.lpSum((xstaff[s][h][b]*c1staff[b]) for (s,h,b) in shb) +
        pulp.lpSum(pscenarios[k]*(ubed[s][h][k]*c2bed[s][h]) for (s,h,k) in shk)+
        pulp.lpSum(pscenarios[k]*(ustaff[s][h][b][k]*c2staff[b]) for (s,b,h,k) in sbhk)
    )

    prob = add_testc_constraints(
        xbed=xbed,
        xstaff=xstaff,
        ubed=ubed,
        ustaff=ustaff,
        UBbed=UBbed,
        UBstaff=UBstaff,
        UBubed=UBubed,
        UBustaff=UBustaff,
        D=D,
        R=R,
        K=K,
        sh=sh,
        shb=shb,
        shk=shk,
        srhk=srhk,
        sbhk=sbhk,
```

```
        prob=prob,
        First_stage=First_stage
    )
    # The user can select one of the two optimisers
    # prob.solve(pulp.GUROBI())
    # prob.solve(pulp.PULP_CBC_CMD())
    return prob
```

The user is required to enter the deterministic results into the model as a three-dimensional array. As previously discussed, this can be implemented either via manipulation of the deterministic output or manually entering. The following provides the code to use alongside the deterministic model to generate the `First_stage` array.

```
\begin{lstlisting}[language=python]
    import pandas as pd
    import numpy as np
    b = [] #Create an empty array
    for v in prob.variables():
        if v.name[0:4] == "Xbed": #Filter xbed variables only
            b.append(v.varValue) #Assign values to b
    df = pd.DataFrame(np.zeros((29,12))) #Create a 3d array with
   zeros
    for j in range(0,29): #Iterate over each row
        df.iloc[j] = b[j*12:(j+1)*12] #Assign columns
    df = df[[0,1,4,5,6,7,8,9,10,11,2,3]] #Reorder the columns
    df = df.reindex([0,11,21,22,23,24,25,26,27,28,1,2
    ,3,4,5,6,7,8,9,10,12,13,14,15,16,17,18,19,20]) #Reorder the
   rows
    df.replace(-0,0, inplace=True)
    First_stage = df.values
```

The `First_stage` along with the other variables can be inputted into the model in the following forms:

```
    """
    These values can be altered to the specific requirements of the
    user
    """
    specialties = list(itertools.chain(range(0, ))) #Creates list
   of specialties
    bands = list(itertools.chain(range(0, ))) #Creates list of
   nursing bands
    regions = list(itertools.chain(range(0, ))) #Creates List of
   regions

    region1 = []
    region2 = []
```

```
    region3 = []
    region4 = []
    region5 = []
    region6 = []
    hospitals = region1 + region2 +region3 + region4 +region5 +
   region6
    D = [
    [[],[]],
    ]
    K = [
    [],
    ]
    R = [
    [],
    ]
    c1bed = [
    [],
    ]
    c2bed = [
    [],
    ]
    c1staff = []
    c2staff = []
    UBstaff = [
    [],
    ]
    UBustaff = [
    [],
    ]
    UBbed = [
    [],
    ]
    UBubed =[
    [],
    ]
    scenarios = []
    pscenarios = []
    First_stage=[
    [],
    ]
```

The variables are then called through the model into the model.

```
    prob = solve_testc_minimisation_problem(
    specialties,
    bands,
    hospitals,
    regions,
```

```
    scenarios,
    pscenarios,
    D,
    R,
    K,
    c1bed,
    c2bed,
    c1staff,
    c2staff,
    UBbed,
    UBstaff,
    First_stage
)
```

The final stage is for the model to output the results of the objective function along with the values of each of the decision variables.

```
    print("Solution Status = ", pulp.LpStatus[prob.status])
    print("Total price = ", pulp.value(prob.objective))
    for v in prob.variables():
        if v.varValue >0:
            print(v.name, "=", v.varValue)
```

Chapter 8

Conclusions

KESS2 funded this research [15] in collaboration with the Clinical Futures [16], within the Aneurin Bevan University Health Board (ABUHB). The aim of the project was to produce a decision support tool, supporting the mathematical modelling unit in bed and staffing resource requirements. The chapter serves as a summary of the research undertaken in this thesis. It provides a brief overview of the research questions listed in Section 1 and the methods used to answer them. The chapter also presents contributions of the thesis, limitations, impact in practice and recommendations for future work.

8.1 Research Summary

Chapter 1 provided an introduction to the frail and elderly population within Wales. The chapter discussed the demographic changes that are occurring within the population and the impact that these changes will have on the healthcare system. The types of hospitals and specialties which the health board currently have were also discussed. Current bed and staff planning methods were analysed with the following four research questions being introduced:

1. How do the clinical and demographical attributes of frail and elderly patients effect their length of stay within hospital?
2. How best can specialties be organised among a network of hospitals to ensure staffing and bed costs are minimised, whilst still meeting the demand for frail and elderly patients?
3. Can linking predictive and prescriptive analytics provide improvements for decision making for frail and elderly services?
4. How can deterministic and two-stage stochastic models be used to plan hos-

pital services for frail and elderly patients within Aneurin Bevan University Health Board?

Chapter 2 provided two literature reviews of the practice of Operational Research and Management Science (OR/MS) approaches in the planning of care for the frail and elderly. The underutilisation of OR/MS techniques, the absence of comprehensive holistic care planning, and the implications of increases in demand on healthcare systems have all been noted as gaps in current literature. Within this thesis, these gaps were addressed.

Chapter 3 addressed the theory underlying the most popular predictive analytical techniques presently used in healthcare. The results demonstrated the benefit of utilising more complex models, such as classification and regression trees (CART), instead of simpler models, such as linear regression to predict the length of stay (LOS) of frail and elderly patients. These results yielded a more accurate prediction of LOS, which is important for planning purposes. A step by step practical example was also included so that healthcare professionals could quickly apply these strategies to their own departments and data. To enable model adaptation and parameter optimisation, detailed executable Python code was provided.

Chapter 4 provided an introduction to two prescriptive methodologies, deterministic and two-stage stochastic modelling. Expanding on the two-stage stochastic programming paradigm and building on the tests introduced by Maggioni and Wallace [257], this chapter went further by creating two-dimensional decision variables which are dependent on each other along with the application to a different field of research, namely frail and elderly patient planning. The equations generated allow for the optimisation of the number of beds and staff required to meet demand. The models created were robust in terms of working ability. Furthermore, the modelling was shared with ABUHB, especially how the models work. The tests discussed in [257], have also been employed, applied and evaluated to each of the examples.

Chapter 5 presented the findings of the predictive and prescriptive analytical models. Section 5.2 provided an overview of the current data and trends within ABUHB and within the frail and elderly community. Section 5.3 aimed to answer the first research question by generating CART models to predict LOS of frail and elderly patients. The models also compared the impact of frailty on LOS. The results highlighted the improved R^2 and accuracy scores when using CART models over traditional linear and logistic regression methods. These CART models also enabled patient groupings of similar attributes to be determined. Section 5.4 aimed to answer the second research question by applying the deterministic and two-stage stochastic models generated in Chapter 4 to ABUHB data. The models determined how beds should be planned and staff deployed based on figures from Public Health Scotland

and NHS Jobs, to ensure costings are minimised. Results showed the benefits of utilising the two-stage stochastic model to plan their beds and staff over traditional deterministic models. Any savings made by the NHS could be reinvested into other areas of healthcare.

Chapter 6 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 South East Wales. Research questions three and four were answered by comparing the CART results to the traditional averages. The primary aim of employing CART models was to explore the potential benefits of using prescriptive methods for resource capacity planning in the healthcare context. By using CART models, we aimed to gain a more comprehensive understanding of the factors influencing LOS and its variations among patients. The predictive capabilities of CART allowed us to identify non-linear relationships and interactions among various patient characteristics and medical conditions, leading to more accurate and individualised LOS predictions. The derived daily bed demand, informed by the CART predicted LOS, provided a more realistic representation of the variation within hospital LOS. Unlike traditional averages, which might overlook patient-specific factors affecting LOS, the CART-based approach captured a wider range of LOS variations, reflecting the diversity and complexity of patient care requirements.

Finally, Chapter 7 provided a tutorial on how to use the deterministic and two-stage stochastic models generated. The models were implemented in Microsoft Excel using the OpenSolver add-in, and in Python using the PuLP package. Both versions of the models were included to reach a wider audience and subsequently uploaded to GitHub for future use. These models are available from [287]. These tutorials aim to provide a step by step guide on how to use the models and be applied to other healthcare organisations. As future patient demographics change, the models can be rerun with updated data to determine the most efficient way to plan beds and staff.

8.2 Research Contributions

The findings presented within this thesis have provided a number of novel contributions to the literature on OR and healthcare applications. These contributions are as follows:

- The literature reviews presented in Chapter 2 provided a comprehensive overview of the current literature on frail and elderly care planning with OR/MS methods and hierarchical prediction models to predict LOS. This allowed themes and methods to be identified and enabled gaps within the literature to be determined. The reviews focused specifically on frail and elderly patients, showing the limited research published within this area.

- The development of the predictive models (Chapter 3) provided a novel method to predict LOS of frail and elderly patients, instead of considering these patients within the adult population. This allowed for the impact of frailty on LOS to be determined. This chapter has used sophisticated techniques which are underutilised within the context of healthcare.
- Prescriptive models were developed to plan beds and staff for frail and elderly patients (Chapter 4). These models expanded upon the work of Maggioni and Wallace [257], by applying to the area of healthcare OR and analysing bed and staff requirements. Instead of planning on a ward by ward basis, these models enabled holistic planning to take place across the health board.
- By linking predictive and prescriptive analytics, decision-makers can achieve a more comprehensive view of their data and use it to make more informed decisions. Chapter 6 demonstrated how these methods could be linked, providing a number of examples of different methods. This allowed for scenario analysis to be performed, using a combination of techniques to provide unique insights into the ABUHB healthcare system.

8.3 Limitations of the Study

There are several limitations to this research. Firstly, the reliance on historical activity data to predict future demands poses a significant constraint. One of the primary concerns is the potential omission of unmet demand from the dataset. Activity data typically capture the services that have been provided and recorded in the system, but they may not fully represent the actual demand for healthcare services. Unmet demand, or the demand that goes unaddressed due to capacity constraints or other factors, is critical to consider in resource allocation to ensure the system can meet the true needs of the population. Another challenge is the inclusion of LOS which reflect poor historic system performance. When historical LOS's are incorporated into the model, they may inadvertently perpetuate inefficiencies or suboptimal practices from the past.

Another limitation of the study is its reliance on pre-Covid-19 data for capacity allocation modelling. As the Covid-19 pandemic has had a profound and unprecedented impact on healthcare systems worldwide, using data prior to the pandemic might not fully reflect the current and future resource allocation needs. The pandemic has introduced unique challenges, such as surges in patient volumes, changes in patient acuity, and shifts in healthcare priorities. The demand for resources, including beds, nursing staff, and other medical supplies, has been substantially affected during this period.

A notable limitation of the model is its omission of ward sizes and the number of beds or staff per ward in the resource allocation process. Instead, the model adopts a more holistic approach, considering the healthcare facility as a whole entity. While this simplification may offer practicality and ease of implementation, it overlooks crucial ward-level variations in patient capacity and staffing requirements.

This research relied on the use of open source data from StatsWales [4] and Public Health Scotland [11]. Therefore, there was potentially inaccurate or imprecise data to populate the model. The accuracy of the model's predictions heavily relies on the quality and reliability of the data used as input.

Finally, both the Excel and Python tools, utilise various constraints and objective functions that were presented in Sections 4.3 and 4.4 respectively. Whilst other limitations can be alleviated by changing the data within the Excel worksheets and Python scripts, if new constraints were needed to be added, this would require the users to have knowledge and understanding of the formulation of the mathematical constraints.

8.4 Impact in Practice

This research collaboration with the Clinical Futures team at ABUHB has significantly influenced the development and direction of the project. ABUHB's substantial time and financial investment in the research reflect their interest in deriving benefits from the outcomes. At the time when the thesis was finished, the planning team were aiming to share the results with the executive board using an SBAR. This may lead to further support the Aneurin Bevan Continuous Improvement team (ABCi), as the project continues with the support of the interim director of planning and the lead of the mathematical modelling unit. They have shown interest in the model's potential and are expected to utilise it to craft a compelling case study for senior decision-makers within the health board.

Moreover, the team at ABCi offers an engaging analytics program, providing interactive training to front-line staff in data analysis and mathematical modelling techniques [302, 303]. This presents an exciting opportunity to integrate these models into their program, ensuring broader dissemination and impactful utilisation of the research findings. By being part of their curriculum, these models have the potential to empower healthcare professionals with valuable insights, ultimately driving informed decision-making and resource optimisation within the healthcare domain.

8.5 Future Work

The work presented in this research has provided a number of insights into the ABUHB healthcare system, however, there are a number of areas that could be further explored. The following areas for future study were identified:

- Chapter 3 and Chapter 5: The predictive models presented within this thesis could be further developed to include other attributes detailing a patient's medical history. These may include the number of previous admissions, the number of previous admissions to the same ward and the number of previous admissions to the same specialty.
- Chapter 4 and Chapter 5: The prescriptive models presented within this thesis could be further developed to include additional variables. Further work could include planning specialties by specific wards rather than generalising across specialties. Additionally, the demand for resources within the hospital such as phlebotomists, radiographers and physiotherapists could be included.
- Chapter 5 and Chapter 6: The models were developed using either three years' worth of data or splitting the data by year. This could be further developed by using a time constraint to plan on specific time periods rather than on a longer-term time scale. This would create a more dynamic model where the health board would be able to adapt to seasonal demand changes or determine how beds and nursing resources would change on a smaller time scale.
- Chapter 6: The linked predictive and prescriptive models presented within this thesis could use sampling from the end nodes rather than using the average. This would provide a randomised solution to the problem, which could be used for prediction purposes. Further investigation into population predictions could also be used within the models as a separate input. This chapter also investigated a range of various scenarios, including the addition of the Grange University Hospital (GUH). Due to the limitation of the data received, i.e., pre-2020, the impact of the new hospital was unable to be determined, as this opened in 2021, however, using the data prior to its opening, GUH could still be investigated. Therefore, the model still provided useful results and recommendations for bed planning and nursing staff. To determine how beds should be planned, more recent data should be used and the effects on demand following the Covid-19 pandemic can be visualised. The model could be developed further to have real time updates of the demand entering the system so planning can be conducted on a more operational scale.

Finally, further research could consider analysing other areas of the ABUHB healthcare system, as well as other age groups.

Appendix A

List of Hospitals and Specialties Within ABUHB

This Appendix contains a list of each of the specialties offered by ABUHB and the hospitals in which they can be found within. These hospitals and specialties are first discussed in Chapter 1, and are utilised within Chapters 5 and 6

Specialties	**Hospitals**
Accident & Emergency	Nevill Hall Hospital
	Royal Gwent Hospital
Anaesthetics	Nevill Hall Hospital
	Royal Gwent Hospital
Cardiology	Nevill Hall Hospital
	Offsite
	Royal Gwent Hospital
Care Of The Elderly	Chepstow Community Hospital
	County Hospital
	Nevill Hall Hospital
	Royal Gwent Hospital
	St Woolos Community Hospital
	Ysbyty Aneurin Bevan
	Ysbyty Ystrad Fawr
Community Medicine	Chepstow Community Hospital
	County Hospital
	Ysbyty Ystrad Fawr
Dermatology	Royal Gwent Hospital
	St Woolos Acute Hospital
	St Woolos Community Hospital
	Ysbyty Ystrad Fawr
Diabetes And Endocrinology	Nevill Hall Hospital

	Royal Gwent Hospital
	Ysbyty Ystrad Fawr
Ear Nose & Throat	Royal Gwent Hospital
	Ysbyty Ystrad Fawr
GP Other	Chepstow Community Hospital
	Monnow Vale Health and Social Care Facility
	Rhymney Integrated Health & Social Care Centre
Gastroenterology	Nevill Hall Hospital
	Offsite
	Outsource
	Royal Gwent Hospital
	St Woolos Acute Hospital
	Ysbyty Ystrad Fawr
General Medicine	Nevill Hall Hospital
	Royal Gwent Hospital
	Ysbyty Aneurin Bevan
	Ysbyty Ystrad Fawr
General Surgery	Chepstow Community Hospital
	Nevill Hall Hospital
	Offsite
	Royal Gwent Hospital
	Ysbyty Ystrad Fawr
Gynaecology	Nevill Hall Hospital
	Offsite
	Royal Gwent Hospital
	Ysbyty Ystrad Fawr
Haematology	Nevill Hall Hospital
	Offsite
	Royal Gwent Hospital
	St Woolos Acute Hospital
	Ysbyty Ystrad Fawr
Infectious Diseases	Royal Gwent Hospital
Intermediate Care	County Hospital
	Ysbyty Aneurin Bevan
Maxillo-Facial	Nevill Hall Hospital
	Royal Gwent Hospital
	St Woolos Acute Hospital
Neurology	Royal Gwent Hospital
Ophthalmology	Nevill Hall Hospital
	Offsite
	Outsource
	Outsource - CareUK

	Royal Gwent Hospital
	Ysbyty Ystrad Fawr
Pain	County Hospital
	Nevill Hall Hospital
	Royal Gwent Hospital
	Ysbyty Ystrad Fawr
Plastic Surgery	Nevill Hall Hospital
	Offsite
Radiology	Nevill Hall Hospital
	Royal Gwent Hospital
Radiotherapy And Oncology	Royal Gwent Hospital
Rehabilitation	Chepstow Community Hospital
	County Hospital
	Nevill Hall Hospital
	Royal Gwent Hospital
	St Woolos Acute Hospital
	St Woolos Community Hospital
	Ysbyty Aneurin Bevan
	Ysbyty Ystrad Fawr
Respiratory	Nevill Hall Hospital
	Royal Gwent Hospital
Restorative Dentistry	Royal Gwent Hospital
Rheumatology	Nevill Hall Hospital
	Royal Gwent Hospital
	Ysbyty Ystrad Fawr
Trauma & Orthopaedic	Nevill Hall Hospital
	Royal Gwent Hospital
	St Woolos Acute Hospital
	Ysbyty Ystrad Fawr
Urology	Nevill Hall Hospital
	Offsite
	Royal Gwent Hospital
	St Woolos Acute Hospital
	University Hospital Of Wales
	Ysbyty Ystrad Fawr

Table A.1: Hospital and Specialty Locations in ABUHB

Appendix B

Literature Review Supplementary Material

This appendix has been divided into two sections since two literature reviews were completed within Chapter 2. Section B.1 refers to the literature review contained in Section 2.2, whereas Section B.2 refers to the literature review contained in Section 2.3.

B.1 Application of OR/MS Methods to Frail and Elderly Healthcare - Figures and Tables (Chapter 2.2)

Ref	Authors	JCR	Country	Research Aim	Condition	Method	Planning Decision	Setting
[70]	Abe et al.	GG	Japan	Examining and Forecasting	Acute	Machine Learning	Strategic	Single Hospital
[97]	Ambagtsheer et al.	GG	Australia	Examining	Chronic	Machine Learning	Strategic	Community Care
[94]	Arling et al.	HPS	USA	Forecasting	Chronic	Optimisation	Strategic	Community Care
[98]	Arvelo et al.	Other	Spain	Examining	Chronic	Optimisation	Strategic	Community Care
[71]	Azad et al.	GG	Canada	Examining	Acute	Statistical Analysis	None	Single Hospital
[83]	Bae et al.	OR/MS	USA	Forecasting	Chronic	Simulation	Strategic	Community Care
[74]	Beaupre et al.	Other	Canada	Examining	Chronic	Statistical Analysis	None	Single Hospital
[95]	Borowiak et al.	GG	Poland	Forecasting	Chronic	Statistical Analysis	Tactical	Community Care
[68]	Cepoiu-Martin and Bischak	MI	Canada	Examining	Chronic	Simulation	Strategic	Community Care
[112]	Chaussalet et al.	MI	UK	Improving	Chronic	Queuing Models	Strategic	Single Hospital
[79]	Christodoulou and Taylor	HPS	UK	Forecasting	Chronic	Markov	Strategic	Single Hospital
[52]	Davari and Van Woensel	OR/MS	UK	Forecasting	Chronic	Optimisation	Strategic	Multiple Hospitals and Community
[99]	Desai et al.	HPS	UK	Forecasting	Chronic	Simulation	Strategic	Community Care
[100]	Eggink et al.	HPS	The Netherlands	Forecasting	Chronic	Simulation	Tactical	Community Care
[90]	Eveborn et al.	OR/MS	Sweden	Improving	Chronic	Heuristics	Operational	Community Care
[61]	Faddy and McClean	MI	UK	Examining	Chronic	Markov	Strategic	Single Hospital and Community
[80]	Franck et al.	OR/MS	France	Examining	Chronic	Simulation	Strategic	Multiple Hospitals
[69]	Franklin and Hunter	GG	UK	Examining	Acute	Markov	Strategic	Single Hospital
[62]	Garg et al.	Other	UK	Examining	Chronic	Markov	None	Single Hospital and Community
[63]	Garg et al.	HPS	UK	Improving and Forecasting	Chronic	Markov	Tactical	Single Hospital and Community
[84]	Gassoumis et al.	Other	USA	Examining	Chronic	Machine Learning	Strategic	Community Care
[64]	Gordon et al.	MI	Italy	Examining	Chronic	Markov	Operational	Single Hospital and Community
[65]	Gordon et al.	HPS	Italy	Forecasting	Chronic	Markov	Strategic	Single Hospital and Community
[81]	Gorunescu et al.	HPS	UK	Forecasting	Chronic	Queuing Models	Strategic	Single Hospital
[91]	Grenouilleau et al.	IE	Canada	Forecasting	Chronic	Metaheuristic	Operational	Community Care
[92]	Guo et al.	Other	USA	Examining	Chronic	Statistical Analysis	Strategic	Community Care
[113]	Hamdani et al.	IE	France	Improving	Chronic	Markov	Tactical	Single Hospital
[66]	Hare et al.	HPS	Canada	Forecasting	Chronic	Markov	Strategic	Single Hospital and Community
[114]	Harrison	HPS	USA	Examining	Chronic	Statistical Analysis	Tactical	Multiple Hospitals
[109]	Heggestad	HPS	Norway	Forecasting	Chronic	Statistical Analysis	Strategic	Multiple Hospitals
[53]	Intrevado et al.	HPS	Canada	Examining	Chronic	Optimisation	Strategic	Multiple Hospitals and Community

Table B.1a: Summary of papers identified through the Scopus search in the first literature review, with a total of 62 papers.

Ref	Authors	JCR	Country	Research Aim	Condition	Method	Planning Decision	Setting
[54]	Johnson et al.	OR/MS	USA	Forecasting	Chronic	Optimisation	Strategic	Multiple Hospitals and Community
[101]	Katsaliaki et al	HPS	UK	Forecasting	Chronic	Simulation	Strategic	Community Care
[51]	Kerpershoek et al.	GG	European wide	Examining	Chronic	Anderson Model	Strategic	Community Care
[72]	Kul et al.	GG	Italy	Examining	Acute	Machine Leaning	Operational	Single Hospital
[85]	Li et al	IE	USA	Forecasting	Chronic	Newsvendor Model	Strategic	Community Care
[55]	Lim et al.	OR/MS	Hong Kong	Improving	Chronic	Routing	Tactical	Multiple Hospital and Community
[89]	Lin et al.	IE	Hong Kong	Improving	Chronic	Metaheuristic	Strategic	Community Care
[110]	Marshall and McClean	HPS	UK	Forecasting	Chronic	Markov	Strategic	Single Hospital
[111]	Marshall and McClean	OR/MS	UK	Forecasting	Chronic	Markov	Tactical	Single Hospital
[82]	Marshall et al.	Other	Italy	Forecasting	Chronic	Markov	Strategic	Multiple Hospitals
[56]	McClean and Millard	OR/MS	UK	Examining	Chronic	Markov	Strategic	Multiple Hospitals and Community
[86]	Mohammadi Bidhandi et al.	OR/MS	Canada	Forecasting	Chronic	Queuing models and Simulation	strategic	Community Care
[96]	Muramatsu et al.	GG	USA	Examining	Chronic	Statistical Analysis	Strategic	Community Care
[106]	Onggo et al.	OR/MS	UK	Examining	Acute	Simulation	None	Single Hospital
[57]	Patrick	OR/MS	Canada	Examining and Forecasting	Chronic	Markov and Simulation	Strategic	Multiple Hospitals and Community
[58]	Ragab et al.	Other	Ireland	Improving	Chronic	Simulation	Strategic	Multiple Hospitals and Community
[76]	Rashwan et al.	OR/MS	Ireland	Improving	Acute	Simulation	Strategic	Single Hospital
[77]	Rossille et al.	HPS	France	Improving	Acute	Machine Leaning	None	Single Hospital
[73]	Shaw and Marshall	OR/MS	UK	Forecasting	Acute	Markov	Strategic	Single Hospital
[107]	Silverster et al.	GG	UK	Improving	Acute	Statistical Analysis	Tactical	Single Hospital
[102]	Tao et al.	Other	China	Improving	Chronic	Optimisation	Strategic	Community Care
[67]	Taylor et al.	Other	UK	Forecasting	Chronic	Markov	Strategic	Single Hospital and Community
[78]	Trevisan et al.	GG	Italy	Improving	Acute	Machine Learning	Tactical	Single Hospital
[59]	Walker and Haslett	Other	Australia	Examining	Chronic	Simulation	None	Multiple Hospitals and Community
[75]	Wallace et al.	GG	USA	Examining	Acute	Statistical Analysis	Strategic	Single Hospital
[87]	Welberry et al.	GG	Australia	Examining	Chronic	Machine Learning	Strategic	Community Care
[103]	Xie et al.	Other	UK	Forecasting	Chronic	Markov	Strategic	Community Care
[93]	Yalçindağ et al.	IE	Italy	Forecasting	Chronic	Routing	Tactical	Community Care
[88]	Zhang and Puterman	HPS	Canada	Forecasting	Chronic	Simulation	Strategic	Community Care
[108]	Zhang et al.	OR/MS	Canada	Forecasting	Chronic	Simulation	Strategic	Single Hospital
[60]	Zychlinski et al.	OR/MS	USA	Improving	Chronic	Fluid Model	Strategic	Multiple Hospitals and Community

Table B.1b: Summary of papers identified through the Scopus search in the first literature review, with a total of 62 papers.

JCR Category		Total
Geriatrics and Gerontology (GG)	[51, 69, 70, 71, 72, 75, 78, 87, 95, 96, 97, 107]	12
Health Policy and Services (HPS)	[53, 63, 65, 66, 77, 79, 88, 94, 99, 100, 101, 111, 114]	16
Industrial Engineering (IE)	[85, 89, 91, 93, 113]	5
Medical Informatics (MI)	[61, 65, 68, 112]	4
Operations Research and Management Sciences (OR/MS)	[52, 54, 55, 56, 57, 60, 73, 76, 80, 83, 86, 90, 106, 108, 110]	15
Other	[58, 59, 62, 67, 74, 82, 84, 98, 102, 103]	10

Table B.2: Number of papers that fall into each JCR category for the first literature review.

Medical Setting		Total
Community Care	[51, 68, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103]	23
Community Care & Multiple Hospitals	[52, 53, 54, 55, 56, 57, 58, 59, 60]	9
Single Hospital	[69, 70, 71, 72, 74, 76, 77, 79, 81, 106, 110, 111, 112, 113] [73, 75, 78, 107, 108]	19
Single Hospital & Community Care	[61, 62, 63, 64, 65, 66, 67]	7
Multiple Hospitals	[80, 82, 109, 114]	4

Table B.3: Number of papers that fall into each hospital setting for the first literature review.

Condition		Total
Acute	[69, 70, 71, 72, 73, 75, 76, 77, 78, 106, 107]	10
Chronic	[51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 74, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 108, 109, 110, 111, 112, 113, 114]	51

Table B.4: Number of papers that fall into each condition area for the first literature review.

Markov Method		Total
Continuous time	[56, 61, 62, 64, 65, 67, 69, 73, 79, 82, 103, 110, 111, 113]	14
Discrete time	[57, 63, 66]	3

Table B.5: Number of Markov model papers identified in the first literature review.

Research Aim		Total
Examining	[51, 53, 56, 59, 61, 62, 64, 68, 69, 70, 71, 72, 74, 75, 80, 84, 87, 92, 96, 97, 98, 106, 114]	23
Forecasting	[52, 54, 57, 63, 65, 66, 67, 70, 73, 79, 81, 82, 83, 85, 86, 88, 91, 93, 94, 95, 99, 100, 101, 103, 108, 109, 110, 111]	28
Improving	[55, 57, 58, 60, 63, 76, 77, 78, 89, 90, 102, 107, 112, 113]	14

Table B.6: Number of papers that fall into each research aim for the first literature review.

Planning Decision		Total
Strategic	[51, 52, 53, 54, 56, 57, 58, 60, 61, 65, 66, 67, 68, 69, 70, 73, 75, 76, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 92, 94, 96, 97, 98, 99, 101, 102, 103, 108, 109, 111, 112]	42
Tactical	[55, 63, 65, 78, 93, 95, 100, 107, 110, 113, 114]	10
Operational	[64, 72, 90, 91]	4
No Decision	[59, 62, 71, 74, 77, 106]	6

Table B.7: Number of papers that fall into each planning decision for the first literature review.

B.2 Hierarchical Prediction Models for Patients' Lengths of Stay - Tables (Chapter 2.3).

Ref	Authors	JCR	Country	Research Aim	Condition	Method	Planning Decision	Setting
[70]	Abe et al.	GG	Japan	Examining and Forecasting	Acute	Machine Learning	Strategic	Single Hospital
[169]	Adamis et al.	GG	Ireland	Examining	Chronic	Linear and Logistic Regression	Operational	Single Hospital
[139]	Agasi-Idenburg et al.	GG	The Netherlands	Examining	Surgical	Logistic Regression	Operational	Multiple Hospitals
[149]	Alyahya et al.	MI	Jordan	Forecasting	Acute	Decision Tree	Operational	Single Hospital
[150]	Antonelli et al.	GG	Italy	Forecasting	Chronic	Logistic Regression	Operational	Multiple Hospitals
[124]	Bahrmann et al.	GG	Germany	Forecasting	Acute	Kaplan-Meier	Operational	Single Hospital
[151]	Basic and Khoo	HPS	Australia	Forecasting	Acute	Logistic Regression	Operational	Single Hospital
[168]	Basic and Khoo	GG	Australia	Improving	Acute	Cox Regression	Tactical	Single Hospital
[174]	Basic and Shanley	HPS	UK	Forecasting	Chronic	Cox and Logistic Regression	Operational	Single Hospital
[185]	Beauchet et al.	GG	France	Examining	Acute	Linear Regression	Operational	Single Hospital
[152]	Beauchet et al.	GG	Canada	Forecasting	Chronic	Regression and Kaplan-Meier	Tactical	Single Hospital
[198]	Beauchet et al.	GG	Canada	Examining	Acute	Statistical Analysis	Tactical	Single Hospital
[153]	Bo et al.	GG	Finland	Forecasting	Chronic	Multi-dimensional Analysis	Operational	Single Hospital
[154]	Bo et al.	GG	Italy	Forecasting	Acute	Statistical	Operational	Single Hospital
[123]	Cacciatore et al.	GG	Italy	Forecasting	Surgical	Linear Regression	Operational	Single Hospital
[126]	Cai et al.	Other	USA	Examining	Chronic	Cox and Logistic Regression	Tactical	Community Care
[203]	Chen et al.	GG	China	Examining	Chronic	Statistical Analysis	Operational	Multiple Hospitals
[186]	Chua et al.	GG	UK	Examining	Chronic	Linear Regression	Operational	Single Hospital
[172]	Chung et al.	HPS	South Korea	Examining	Acute	Multivariate Regression	Operational	Multiple Hospitals
[155]	Curiati et al.	Other	Brazil	Forecasting	Acute	Logistic Regression	Operational	Single Hospital
[130]	Fan et al.	GG	China	Forecasting	Chronic	Logistic Regression	Strategic	Single Hospital and Community
[156]	Ferreira et al.	GG	Brazil	Forecasting	Chronic	Logistic Regression	Operational	Single Hospital
[164]	Feuerstadt et al.	GG	USA	Forecasting	Acute	Kaplan-Meier	Tactical	Multiple Hospitals
[205]	Garg et al.	Other	UK	Examining	Chronic	Markov	Operational	Multiple Hospitals
[65]	Gordon et al.	HPS	UK	Forecasting	Chronic	Markov	Operational	Single Hospital and Community
[199]	Greene et al.	GG	Ireland	Examining	Acute	Statistical	Operational	Single Hospital
[113]	Hamdani et al.	IE	France	Improving	Chronic	Markov	Tactical	Single Hospital
[200]	Harari et al.	GG	UK	Examining	Acute	Statistical Analysis	Operational	Single Hospital
[193]	Hartley et al.	GG	UK	Examining	Chronic	Cox Regression	Operational	Single Hospital
[143]	Harvey et al.	GG	Australia	Forecasting	Surgical	Logistic Regression	Strategic	Multiple Hospitals
[189]	Hasebe et al.	GG	Japan	Examining	Chronic	Linear Regression	Operational	Multiple Hospitals
[127]	Hoben et al.	GG	Canada	Forecasting	Chronic	Cox Regression	Strategic	Community Care
[157]	Hu et al.	MI	China	Forecasting	Chronic	Gradient Boosting	Strategic	Multiple Hospitals
[180]	Hubbard et al.	GG	Australia	Forecasting	Acute	Logistic Regression	Operational	Multiple Hospitals
[128]	Johnson et al.	GG	USA	Examining	Chronic	Logistic Regression and Kaplan-Meier	Strategic	Community Care
[144]	Jones et al.	GG	USA	Examining	Surgical	Logistic Regression	Operational	Single Hospital
[140]	Justo et al.	GG	Israel	Examining	Surgical	Linear Regression	Operational	Single Hospital
[145]	Kerr et al.	GG	UK	Examining	Surgical	Cox Regression	Operational	Single Hospital
[196]	Kidd et al.	GG	UK	Examining	Acute	Cox Regression	Operational	Multiple Hospitals
[190]	Kim and Lee	GG	South Korea	Forecasting	Acute	Linear Regression	Tactical	Multiple Hospitals
[136]	Kirfel et al.	GG	Germany	Examining	Surgical	Linear Regression	Operational	Single Hospital
[181]	Lang et al.	GG	France	Examining	Chronic	Logistic Regression	Strategic	Multiple Hospitals
[182]	Lang et al.	GG	Finland	Examining	Chronic	Logistic Regression	Strategic	Multiple Hospitals
[194]	Launay et al.	Other	France	Examining	Acute	Cox Regression	Tactical	Single Hospital
[158]	Launay et al.	Other	France	Forecasting	Acute	Neural Networks	Operational	Single Hospital
[159]	Launay et al.	GG	France	Forecasting	Acute	Logistic Regression	Operational	Single Hospital

Table B.8a: Summary of papers identified through the Scopus search in the first literature review, with a total of 90 papers.

Ref	Authors	JCR	Country	Research Aim	Condition	Method	Planning Decision	Setting
[204]	Le et al.	Other	Singapore	Examining	Chronic	Markov	Tactical	Single Hospital
[137]	Lee et al.	Other	Canada	Forecasting	Surgical	Logistic Regression	Operational	Multiple Hospitals
[191]	Liotta et al.	GG	Italy	Examining	Acute	Linear Regression	Tactical	Multiple Hospitals
[160]	Lisk et al.	GG	UK	Forecasting	Chronic	Linear and Logistic Regression	Operational	Multiple Hospitals
[175]	Lisk et al.	GG	UK	Examining	Acute	Logistic Regression	Operational	Single Hospital
[207]	Lisk et al.	Other	UK	Forecasting	Acute	ROC Curve	Operational	Single Hospital
[142]	MacDonald et al.	Other	Canada	Examining	Surgical	Logistic Regression	Tactical	Multiple Hospitals
[134]	Marano et al.	GG	Italy	Forecasting	Surgical	Statistical Analysis	Operational	Single Hospital
[201]	Marin et al.	Other	Spain	Examining	Acute	Statistical Analysis	Operational	Single Hospital
[110]	Marshall and McClean	HPS	UK	Forecasting	Chronic	Markov	Strategic	Single Hospital
[206]	Möllers et al.	Other	Germany	Examining	Chronic	Multivariate Regression	Operational	Single Hospital
[167]	Morandi et al.	GG	USA	Forecasting	Chronic	Kaplan-Meier	Operational	Single Hospital
[170]	Motohashi et al.	HPS	Japan	Examining	Acute	Linear and Logistic Regression	Operational	Multiple Hospitals
[173]	Motzek et al.	GG	Germany	Forecasting	Chronic	Regressions	Tactical	Multiple Hospitals
[171]	Naouri et al.	GG	France	Examining	Acute	Linear and Logistic Regression	Operational	Multiple Hospitals
[176]	Nishida et al.	GG	Japan	Examining	Chronic	Logistic Regression	Operational	Single Hospital
[161]	Nishino et al.	GG	Japan	Forecasting	Acute	CART	Operational	Single Hospital
[195]	Ono et al.	GG	Finland	Examining	Chronic	Cox Regression	Operational	Single Hospital
[129]	Park et al.	Other	USA	Examining	Chronic	Cox Regression	Strategic	Community Care
[197]	Pilotto et al.	GG	Finland	Examining	Chronic	Cox Regression	Operational	Multiple Hospitals
[138]	Pustavoitau et al.	GG	USA	Forecasting	Surgical	Linear Regression	Operational	Single Hospital
[135]	Raab et al.	GG	USA	Forecasting	Surgical	Linear Regression	Operational	Single Hospital
[133]	Rajamaki et al.	GG	Finland	Forecasting	Acute	Logistic Regression	Operational	Multiple Hospitals
[183]	Rubens et al.	GG	USA	Forecasting	Acute	Logistic Regression	Operational	Multiple Hospitals
[184]	Shebeshi et al.	GG	Australia	Forecasting	Acute	Logistic Regression	Operational	Multiple Hospitals
[177]	Shen et al.	GG	China	Examining	Chronic	Logistic Regression	Operational	Single Hospital
[187]	Snowden et al.	GG	USA	Examining	Acute	Linear Regression	Operational	Single Hospital
[162]	Sommerfeld and Arbin	Other	Sweden	Forecasting	Acute	Kaplan Meier	Operational	Single Hospital
[132]	Sommerfeld et al.	GG	Sweden	Forecasting	Acute	Cox Regression	Tactical	Single Hospital
[163]	Takahashi et al.	GG	Japan	Forecasting	Acute	Logistic Regression	Operational	Multiple Hospitals
[125]	Tal	GG	Israel	Examining	Acute	Statistical Analysis	Operational	Single Hospital
[178]	Toh et al.	Other	Singapore	Examining	Chronic	Logistic Regression	Operational	Single Hospital
[165]	Tropea et al.	GG	Australia	Forecasting	Chronic	Logistic Regression	Strategic	Single Hospital
[166]	Volpato et al.	GG	Italy	Forecasting	Chronic	Cox Regression	Operational	Multiple Hospitals
[179]	Volpato et al.	GG	Italy	Examining	Chronic	Logistic Regression	Operational	Single Hospital
[131]	Walsh et al.	HPS	Ireland	Improving	Chronic	Regressions	Tactical	Single Hospital and Community
[87]	Welberry et al.	GG	Australia	Examining	Chronic	Machine Learning	Strategic	Community Care
[141]	Willems et al.	GG	The Netherlands	Forecasting	Surgical	Linear Regression	Operational	Single Hospital
[192]	Wong and Miller	GG	Canada	Forecasting	Acute	Linear Regression	Operational	Multiple Hospitals
[202]	Wright et al.	GG	UK	Examining	Acute	Statistical Analysis	Tactical	Single Hospital
[188]	Yu et al	Other	Australia	Examining	Chronic	Linear Regression	Operational	Single Hospital
[146]	Zattoni et al.	GG	Italy	Forecasting	Surgical	Logistic Regression	Operational	Single Hospital
[147]	Zhao et al.	GG	China	Forecasting	Surgical	Logistic Regression	Tactical	Single Hospital
[148]	Zhao et al.	GG	China	Forecasting	Surgical	Logistic Regression	Operational	Single Hospital

Table B.8b: Summary of papers identified through the Scopus search in the first literature review, with a total of 90 papers.

JCR Category		Total
Geriatrics and Gerontology (GG)	[70, 87, 123, 124, 125, 127, 128, 130, 132, 133, 134, 135, 136, 138, 139, 140, 141, 143, 144, 145, 146, 147, 148, 150, 152, 153, 154, 156, 159, 160, 161, 163, 164, 165, 166, 167, 168, 169, 171, 173, 175, 177, 179, 180, 181, 182, 183, 184, 185, 186, 187, 189, 190, 191, 192, 193, 195, 196, 197, 198, 199, 200, 202, 203]	65
Health Policy and Services (HPS)	[65, 110, 131, 151, 170, 172, 174]	7
Industrial Engineering (IE)	[113]	1
Medical Informatics (MI)	[149, 157]	2
Other	[123, 129, 137, 142, 155, 158, 162, 178, 188, 194, 201, 204, 205, 206, 207]	15

Table B.9: Number of papers that fall into each JCR category for the second literature review.

Medical Setting		Total
Community Care	[87, 126, 127, 128, 129]	5
Single Hospital	[70, 110, 113, 123, 124, 125, 132, 134, 135, 136, 138, 140, 141, 144, 145, 146, 147, 148, 149, 151, 152, 153, 154, 155, 156, 158, 159, 161, 162, 165, 167, 168, 169, 174, 175, 176, 177, 178, 179, 185, 186, 187, 188, 193, 194, 195, 198, 199, 200, 201, 202, 204, 206, 207]	54
Single Hospital & Community Care	[65, 130, 131]	3
Multiple Hospitals	[133, 137, 139, 142, 143, 150, 156, 157, 160, 163, 164, 166, 170, 171, 172, 173, 180, 181, 182, 183, 184, 189, 190, 191, 192, 196, 197, 203]	28

Table B.10: Number of papers that fall into each hospital setting for the second literature review.

Condition		Total
Acute	[70, 124, 125, 132, 133, 149, 151, 154, 155, 158, 159, 161, 162, 163, 164, 168, 170, 171, 172, 175, 180, 183, 184, 185, 187, 190, 191, 192, 194, 196, 198, 199, 200, 201, 202, 207]	36
Chronic	[65, 87, 110, 113, 126, 127, 128, 129, 130, 131, 150, 152, 153, 156, 157, 160, 165, 166, 167, 169, 173, 174, 176, 177, 177, 178, 179, 181, 182, 186, 188, 189, 193, 195, 197, 197, 203, 204, 205, 206]	38
Surgical	[123, 134, 135, 136, 137, 138, 139, 140, 141, 142, 143, 144, 145, 146, 147, 148]	16

Table B.11: Number of papers that fall into each condition area for the second literature review.

Research Aim		**Total**
Examining	[70, 87, 125, 126, 128, 129, 136, 139, 140, 142, 144, 145, 169, 170, 171, 172, 175, 176, 177, 178, 179, 181, 182, 185, 186, 187, 188, 189, 191, 193, 194, 195, 196, 197, 198, 199, 200, 201, 202, 203, 204, 205, 206]	45
Forecasting	[65, 70, 110, 123, 124, 127, 130, 132, 133, 134, 135, 137, 138, 141, 143, 146, 147, 148, 149, 150, 151, 152, 153, 154, 155, 156, 157, 158, 159, 160, 161, 162, 163, 164, 165, 166, 167, 173, 174, 180, 183, 184, 190, 192, 207]	43
Improving	[113, 131, 168]	3

Table B.12: Number of papers that fall into each research aim for the second literature review.

Planning Decisions		**Total**
Strategic	[70, 87, 110, 127, 128, 129, 130, 143, 157, 165, 181, 182]	12
Tactical	[113, 126, 131, 132, 142, 147, 152, 164, 168, 173, 190, 191, 194, 198, 202, 204]	16
Operational	[65, 123, 124, 125, 133, 134, 135, 136, 137, 138, 139, 140, 141, 144, 145, 146, 148, 149, 150, 151, 153, 154, 155, 156, 158, 159, 160, 161, 162, 163, 166, 167, 169, 170, 171, 172, 174, 175, 176, 177, 178, 179, 180, 183, 184, 185, 186, 187, 188, 189, 192, 193, 195, 196, 197, 199, 200, 201, 203, 205, 206, 207]	62

Table B.13: Number of papers that fall into each planning decision for the second literature review.

Appendix C

Experimental Analysis Further Material

This appendix contains the material relating to Chapter 5. Section C.1 contains the data relating to hospital admissions and scan information for patients admitted between 1^{st} April 2017 and 31^{st} March 2020. Section C.2 contains the regression tree output data.

C.1 ABUHB Admission and Scan Data - Attributes

Attribute	Data type	Distinct attribute values or bins	Documentation		
			at admission	during admission	at discharge
Admission Date	Ordinal	1,096 (e.g. 01/04/2017)	✓		
Admission Method	Nominal	17 (e.g. Elective waiting list)	✓		
Admission Source	Nominal	26 (e.g. Usual place of residence)	✓		
Admission Time	Continuous	1440 ({hh:mm})	✓		
Borough	Nominal	174 (e.g. Newport LHB, Monmouthshire LHB)	✓		
Date of birth	Ordinal	12037 (e.g. 01/01/1940)	✓		
Diagnosis	Nominal	2758 (e.g. Fracture of neck of femur, Congestive heart failure)	✓		
Discharge Date	Ordinal	1154 (e.g. 01/04/2017)			✓

Attribute	**Data type**	**Distinct attribute values or bins**	**Documentation**		
			at admission	**during admission**	**at discharge**
Discharge Destination	Nominal	26 (e.g. Death, Own home, Patient transfer within same health board/trust)			✓
Discharge Time	Continuous	1306 ({hh:mm:ss})			✓
Hospital	Nominal	14 (e.g. Chepstow Community Hospital)	✓		
NHS Number	Nominal	66251 (e.g. 4900000000)	✓		
Postcode	Nominal	13819 (e.g. CF72 8XR)	✓		
Registered GP	Nominal	1313 (e.g. G9041668)	✓		
Registered GP Practice	Nominal	618 (e.g. W93012)	✓		
Scan Attendance Date	Ordinal	1097 (e.g. (01/04/2017)		✓	
Scan Attendance Time	Continuous	11417 ({hh:mm:ss})		✓	
Scan Exam	Nominal	293 (e.g. CT Neck and thorax)		✓	

Attribute	**Data type**	**Distinct attribute values or bins**	**Documentation**		
			at admission	**during admission**	**at discharge**
Scan Exam Code	Nominal	295 (e.g. XCHES, XABDO, CSKUH)		✓	
Scan Location Name	Nominal	74 (e.g. Medical Assessment, Intensive Care Unit)		✓	
Scan Procedure Code	Nominal	16 (e.g. R, CT, MR)		✓	
Scan Requested Date	Ordinal	1090 (e.g. (01/04/2017)		✓	
Scan Specialty Code	Nominal	37 (e.g. Gastro, Neuro)		✓	
Specialty	Nominal	30 (e.g. Care of the Elderly, Neurology)	✓		

Table C.1: List of attributes used for the CART analysis.

Variable	Equation
Age on Arrival	$Y = 0.3751x - 22.635$
Frailty Score	$Y = 2.26961x + 5.2631$
Number of Scans	$Y = 1.8414x + 6.2159$

Table C.2: Linear regression results for continuous variables.

Variable	Parameter	Coefficient
Admission Method	Elective - booked	0.2487
	Elective - planned	0.1686
	Elective - waiting list	0.7532
	Emergency - GP	9.7958
	Emergency - NHS Direct	15.0000
	Emergency - bed bureau	15.3636
	Emergency - casualty	8.7958
	Emergency - consultant OP clinic	8.5536
	Emergency - dom. visit by consultant	11.3774
	Emergency - other means	10.6267
	Maternity - ante-partum	10.5000
	Maternity - post-partum	7.0000
	Not applicable	0.0000
	Not known	0.0000
	Other - babies born outside hospital	75.0000
	Other - transferred from another hospital	25.7189
Admission Source	Babies born in or on the way to hospital	1.5000
	Babies born in or on the way to hospital (Baby Act	75.0000
	Hospice	2.0000
	L.A. Part3 residential acc. where care provided	10.9083
	No fixed abode	0.0000
	Non NHS (other than L.A.) run hospice	27.0000
	Non NHS (other than L.A.) run nursing home	8.3851
	Non NHS (other then L.A.) run res.care home	9.0758
	Non-NHS run hospital	5.2308
	Other NHS provider - general ward	25.2320
	Other NHS provider - maternity ward	5.3333
	Other NHS provider - mental health ward	1.5000
	Own Home	4.6228
	Patient transfer from non NHS hospital, includes p	19.0909
	Patient transfer from other health board/ trust.	26.4283
	Patient transfer within the same health board/trus	26.4740
	Penal establishment	2.9333

Variable	Parameter	Coefficient
Admission Source	Penal establishment or police custody suite	5.0000
	Permanent residence at nursing home, residential c	10.0706
	Same Trust- Mentally ill or learning disablilties	15.3571
	Same Trust-General or young phys.disabled	25.3286
	Same Trust-maternity/neonates	15.1429
	Temporary place of residence	7.6400
	Temporary residence at nursing home, residential c	7.5556
	USUAL PLACE OF RESIDENCE	0.0000
	Usual place of residence	4.8091
Age Group	65-69	3.3561
	70-74	3.9699
	75-79	5.6214
	80-84	7.6940
	85-89	10.2833
	90-94	12.8864
	95+	13.8917
Day of Admission	1	8.0375
	2	6.1530
	3	6.0376
	4	6.0256
	5	5.8866
	6	6.7008
	7	7.6869
Frailty Group	0	5.1121
	1	8.8235
	2	11.4974
Hospital	Chepstow Community Hospital(CCH)	36.7876
	County Hospital (CH)	29.3299
	Monnow Vale Health and Social Care Facility (MVHSCF)	39.6596
	Nevill Hall Hospital (NHH)	5.4669
	Offsite	0.0667
	Outsource	0.0000
	Outsource - CareUK	0.0025
	Rhymney Integrated Health & Social Care Centre (RIHSC)	38.7814
	Royal Gwent Hospital (RGH)	5.1409
	St Woolos Acute Hospital (STWAH)	1.2380
	St Woolos Community Hospital (STWCH)	31.9553
	University Hospital Of Wales (UHW)	0.6984
	Ysbyty Aneurin Bevan (YAB)	23.7867

Variable	Parameter	Coefficient
	Ysbyty Ystrad Fawr (YYF)	7.1026
ICD10 - First Letter	0	10.9157
	A	11.3868
	B	7.2211
	C	2.8919
	D	1.3822
	E	10.0316
	F	16.5323
	G	8.0823
	H	0.2263
	I	9.7608
	J	9.5938
	K	2.5100
	L	2.5274
	M	5.4658
	N	5.7637
	Q	3.4783
	R	6.8832
	S	13.5641
	T	8.4221
	Z	0.4610
Scan Y/N	N	6.2178
	Y	8.5000
Month	1	6.4166
	2	6.2411
	3	6.2600
	4	6.9825
	5	6.4336
	6	6.2120
	7	6.3876
	8	6.5247
	9	6.4493
	10	6.1330
	11	5.8964
	12	6.7994
Specialty	Accident & Emergency	2.2673
	Anaesthetics	14.9517
	Cardiology	4.6470
	Care Of The Elderly	12.1101

Variable	**Parameter**	**Coefficient**
Specialty	Community Medicine	34.2350
	Dermatology	0.2616
	Diabetes And Endocrinology	11.6161
	Ear Nose & Throat	2.7500
	GP Other	39.2603
	Gastroenterology	2.1382
	General Medicine	8.4519
	General Surgery	3.7149
	Gynaecology	1.6536
	Haematology	0.7974
	Infectious Diseases	11.6289
	Intermediate Care	14.3725
	Maxillo-Facial	0.6018
	Neurology	5.6131
	Ophthalmology	0.1307
	Pain	0.0080
	Plastic Surgery	0.1128
	Radiology	0.3548
	Radiotherapy And Oncology	13.6667
	Rehabilitation	28.7732
	Respiratory	7.7985
	Restorative Dentistry	0.0000
	Rheumatology	2.3333
	Trauma & Orthopaedic	6.6658
	Urology	0.9932

Table C.3: Linear regression results for categorical variables.

Variable	**Equation**
Age on Arrival	(Age $\times$ 0.00681) - 5.0746
Frailty Score	(Frailty_score $\times$ 0.5490) - 0.0618
Number of Scans	(No_scans $\times$ 1.4368) + 0.0798

Table C.4: Logistic regression results for continuous variables.

Variable	**Parameter**	**Log Odds Ratio**
Admission Method	Intercept	-3.1303
	Elective - planned	-0.3037
	Elective - waiting list	1.5800
	Emergency - GP	5.4922
	Emergency - NHS Direct	7.5391
	Emergency - bed bureau	4.9504
	Emergency - casualty	5.2280
	Emergency - consultant OP clinic	5.4610
	Emergency - dom. visit by consultant	32.8722
	Emergency - other means	5.0776
	Maternity - ante-partum	6.3958
	Maternity - post-partum	3.1225
	Not applicable	-0.2002
	Not known	-0.2002
	Other - babies born outside hospital	3.9658
	Other - transferred from another hospital	7.3295
Admission Source	Intercept	2.1882
	Babies born in or on the way to hospital (Baby Act	0.7221
	Hospice	0.7221
	L.A. Part3 residential acc. where care provided	0.2676
	No fixed abode	-3.7222
	Non NHS (other than L.A.) run hospice	1.3450
	Non NHS (other than L.A.) run nursing home	-0.5895
	Non NHS (other then L.A.) run res.care home	-0.7584
	Non-NHS run hospital	1.4362
	Other NHS provider - general ward	1.9328
	Other NHS provider - maternity ward	1.8957
	Other NHS provider - mental health ward	1.3450
	Own Home	-2.1817
	Patient transfer from non NHS hospital, includes p	4.9402
	Patient transfer from other health board/ trust	1.9041
	Patient transfer within the same health board/trus	2.2659
	Penal establishment	-1.7844
	Penal establishment or police custody suite	-0.6642
	Permanent residence at nursing home, residential c	-0.5487
	Same Trust- Mentally ill or learning disablilties	0.4053
	Same Trust-General or young phys.disabled	1.8935
	Same Trust-maternity/neonates	3.6399

Variable	Parameter	Log Odds Ratio
Admission Source	Temporary place of residence	-0.4217
	Temporary residence at nursing home, residential c	-0.4128
	USUAL PLACE OF RESIDENCE	-27.3023
	Usual place of residence	-2.0817
Age Group	Intercept	-0.3582
	70-74	0.1263
	75-79	0.4389
	80-84	0.7609
	85-89	1.2366
	90-94	1.8219
	95+	2.2298
Day of Admission	Intercept	1.2226
	2	-1.1597
	3	-1.2003
	4	-1.1932
	5	-1.2565
	6	-1.0601
	7	-0.2626
Frailty Group	Intercept	-0.1167
	1	1.1179
	2	1.3066
Hospital	Intercept	6.7546
	County Hospital (CH)	-3.3521
	Monnow Vale Health and Social Care Facility (MVHSCF)	-0.7102
	Nevill Hall Hospital (NHH)	-6.3055
	Offsite	-9.4029
	Outsource	-17.5290
	Outsource - CareUK	-13.0346
	Rhymney Integrated Health & Social Care Centre (RIHSC)	-2.7499
	Royal Gwent Hospital (RGH)	-6.5426
	St Woolos Acute Hospital (STWAH)	-7.8409
	St Woolos Community Hospital (STWCH)	-0.1758
	University Hospital Of Wales (UHW)	-6.2012
	Ysbyty Aneurin Bevan (YAB)	-0.9729
	Ysbyty Ystrad Fawr (YYF)	-7.2621
ICD10 - First Letter	Intercept	1.0536
	A	1.1059
	B	-0.2475
	C	-1.8016
	D	-2.6683
	E	0.7074
	F	1.3826
	G	-0.9233
	H	-4.2614

Variable	Parameter	Log Odds Ratio
ICD10 - First Letter	I	0.1415
	J	1.3383
	K	-1.8507
	L	-2.4526
	M	-0.3327
	N	-0.8702
	Q	-0.6577
	R	-0.3957
	S	0.9451
	T	0.1577
	Z	-3.6601
Scan Y/N	Intercept	0.00772
	Y	1.7035
Month	Intercept	0.1607
	2	-0.0254
	3	0.0376
	4	0.0978
	5	0.0401
	6	0.0230
	7	0.0114
	8	0.0370
	9	0.0221
	10	-0.0333
	11	-0.0951
	12	0.1176
Specialty	Intercept	0.3028
	Anaesthetics	2.2258
	Cardiology	-0.0444
	Care Of The Elderly	2.5741
	Community Medicine	4.7692
	Dermatology	-4.0136
	Diabetes And Endocrinology	2.5748
	Ear Nose & Throat	0.5140
	GP Other	4.3984
	Gastroenterology	-1.5574
	General Medicine	1.6618
	General Surgery	-0.3303
	Gynaecology	-0.6350
	Haematology	-2.4446
	Infectious Diseases	2.5972
	Intermediate Care	4.9687
	Maxillo-Facial	-2.4471
	Neurology	1.5312
	Ophthalmology	-4.0320
	Pain	-5.1595

Variable	Parameter	Log Odds Ratio
Specialty	Plastic Surgery	-5.8436
	Radiology	-1.7967
	Radiotherapy And Oncology	-2.7438
	Rehabilitation	5.5195
	Respiratory	1.2408
	Restorative Dentistry	-1.0784
	Rheumatology	-1.6581
	Trauma & Orthopaedic	0.7239
	Urology	-1.3497

Table C.5: Logistic regression results for categorical variables.

C.2 Regression Tree Groupings

This section provides the regression groupings and the associated average LOS for each of the end nodes. Comments are shown in blue, with average LOS in hours displayed in red.

```
if admission_method_Other - transferred from another hospital <= 0.5:
   if admission_method_Elective - waiting list <= 0.5:
       if admission_method_Elective - booked <= 0.5:
           if specialty_Accident & Emergency <= 0.5:
               if admission_method_Elective - planned <= 0.5:
                   if hospital_Ysbyty Ystrad Fawr <= 0.5:
                       if Age_group_65-69 <= 0.5:
                           if Age_group_70-74 <= 0.5:
                               if specialty_Trauma & Orthopaedic <= 0.5:
                                   if Age_group_75-79 <= 0.5:
                                       if specialty_Care Of The Elderly <= 0.5:
                                           return [[9.81148237]]
                                       else:  # if specialty_Care Of The Elderly > 0.5
                                           return [[11.39308949]]
                                   else:  # if Age_group_75-79 > 0.5
                                       return [[8.83057656]]
                               else:  # if specialty_Trauma & Orthopaedic > 0.5
                                   return [[13.27625571]]
                           else: # if Age_group_70-74 > 0.5
                               return [[7.83854833]]
                       else:  # if Age_group_65-69 > 0.5
                           return [[7.35347877]]
                   else:  # if hospital_Ysbyty Ystrad Fawr > 0.5
                       if FrailtyGroup_2 <= 0.5:
                           if Age_group_90-94 <= 0.5:
                               return [[13.43980061]]
                           else:  # if Age_group_90-94 > 0.5
                               return [[19.]]
                       else:  # if FrailtyGroup_2 > 0.5
                           return [[20.35322777]]
               else:  # if admission_method_Elective - planned > 0.5
                   return [[0.19179104]]
           else:  # if specialty_Accident & Emergency > 0.5
               return [[2.28825294]]
       else:  # if admission_method_Elective - booked > 0.5
           return [[0.24649508]]
   else:  # if admission_method_Elective - waiting list > 0.5
```

```
            if specialty_Trauma & Orthopaedic <= 0.5:
                return [[0.4027881]]
            else:  # if specialty_Trauma & Orthopaedic > 0.5
                if hospital_Ysbyty Ystrad Fawr <= 0.5:
                    return [[3.56313566]]
                else:  # if hospital_Ysbyty Ystrad Fawr > 0.5
                    return [[0.16842962]]
    else:  # if admission_method_Other - transferred from another hospital > 0.5
        if hospital_Royal Gwent Hospital <= 0.5:
            if hospital_St Woolos Acute Hospital <= 0.5:
                if hospital_Nevill Hall Hospital <= 0.5:
                    if hospital_Ysbyty Aneurin Bevan <= 0.5:
                        if specialty_Care Of The Elderly <= 0.5:
                            if specialty_Diabetes And Endocrinology <= 0.5:
                                if specialty_Rehabilitation <= 0.5:
                                    return [[38.63267544]]
                                else:  # if specialty_Rehabilitation > 0.5
                                    if hospital_Chepstow Community Hospital <= 0.5:
                                        if hospital_Ysbyty Ystrad Fawr <= 0.5:
                                            if admission_source_Same Trust-General or
                                            young phys.disabled <= 0.5:
                                                return [[33.48491155]]
                                            else:  # if admission_source_Same Trust-General
                                            or young phys.disabled > 0.5
                                                return [[30.2618469]]
                                        else:  # if hospital_Ysbyty Ystrad Fawr > 0.5
                                            return [[35.2183755]]
                                    else:  # if hospital_Chepstow Community Hospital > 0.5
                                        return [[37.42204301]]
                            else:  # if specialty_Diabetes And Endocrinology > 0.5
                                return [[20.428]]
                        else:  # if specialty_Care Of The Elderly > 0.5
                            if hospital_County Hospital <= 0.5:
                                return [[22.40190476]]
                            else:  # if hospital_County Hospital > 0.5
                                return [[30.08433735]]
                    else:  # if hospital_Ysbyty Aneurin Bevan > 0.5
                        return [[23.80604134]]
                else:  # if hospital_Nevill Hall Hospital > 0.5
                    if specialty_Rehabilitation <= 0.5:
                        return [[11.72972973]]
                    else: # if specialty_Rehabilitation > 0.5
                        return [[29.56521739]]
            else: # if hospital_St Woolos Acute Hospital > 0.5
                return [[6.9055794]]
        else:  # if hospital_Royal Gwent Hospital > 0.5
            if specialty_General Surgery <= 0.5:
                if specialty_Trauma & Orthopaedic <= 0.5:
                    return [[9.71439936]]
                else: # if specialty_Trauma & Orthopaedic > 0.5
                    return [[16.85333333]]
            else:  # if specialty_General Surgery > 0.5
                return [[17.96407186]]
```

C.3 Classification Tree Groupings

This section provides the classification groupings and the LOS groupings for each of the end nodes. Comments are shown in blue, with number of patients falling into each category denoted in red in a [<1,≥1] structure.

```
if admission_method_Elective - waiting list <= 0.5:
    if admission_method_Elective - booked <= 0.5:
        if admission_method_Elective - planned <= 0.5:
            if specialty_Accident & Emergency <= 0.5:
                if specialty_General Medicine <= 0.5:
                    return [[ 2652. 49868.]]
                else:  # if specialty_General Medicine > 0.5
                    return [[1242. 9099.]]
            else:  # if specialty_Accident & Emergency > 0.5
                return [[1848. 2500.]]
        else:  # if admission_method_Elective - planned > 0.5
            if specialty_Trauma & Orthopaedic <= 0.5:
                return [[1289.   17.]]
            else:  # if specialty_Trauma & Orthopaedic > 0.5
                return [[ 5. 29.]]
    else:  # if admission_method_Elective - booked > 0.5
        if specialty_Trauma & Orthopaedic <= 0.5:
            return [[4555.  141.]]
        else:  # if specialty_Trauma & Orthopaedic > 0.5
            return [[22. 61.]]
else:  # if admission_method_Elective - waiting list > 0.5
    if specialty_Trauma & Orthopaedic <= 0.5:
        if specialty_General Surgery <= 0.5:
            if specialty_Urology <= 0.5:
                if specialty_Ear Nose & Throat <= 0.5:
                    if specialty_Gynaecology <= 0.5:
                        if specialty_Respiratory <= 0.5:
                            return [[30951. 700.]]
                        else:  # if specialty_Respiratory > 0.5
                            return [[794. 224.]]
                    else:  # if specialty_Gynaecology > 0.5
                        if Day_1 <= 0.5:
                            return [[951. 462.]]
                        else:  # if Day_1 > 0.5
                            return [[ 0. 72.]]
                else:  # if specialty_Ear Nose & Throat > 0.5
                    return [[264. 330.]]
            else:  # if specialty_Urology > 0.5
                if hospital_Royal Gwent Hospital <= 0.5:
                    return [[1050. 28.]]
                else:  # if hospital_Royal Gwent Hospital > 0.5
                    if Day_5 <= 0.5:
                        if diagnosis_Malignant neoplasm: Bladder, unspecified <= 0.5:
                            if diagnosis_Hyperplasia of prostate <= 0.5:
                                return [[2209.  856.]]
                            else:  # if diagnosis_Hyperplasia of prostate > 0.5
                                return [[103. 176.]]
                        else:  # if diagnosis_Malignant neoplasm: Bladder, unspecified > 0.5
                            return [[106. 196.]]
                    else:  # if Day_5 > 0.5
                        return [[1410.  245.]]
        else:  # if specialty_General Surgery > 0.5
            if diagnosis_Calculus of gallbladder with other cholecystitis <= 0.5:
                if No_Scans <= 0.5:
                    if diagnosis_Diverticular disease of large intestine without perforation
                    or abscess <= 0.5:
                        if hospital_Ysbyty Ystrad Fawr <= 0.5:
                            if diagnosis_Malignant neoplasm: Breast, unspecified <= 0.5:
```

```
                    if diagnosis_Unilateral or unspecified inguinal hernia, without
                    obstruction or gangrene <= 0.5:
                        return [[5121. 1686.]]
                    else:  # if diagnosis_Unilateral or unspecified inguinal hernia,
                    without obstruction or gangrene > 0.5
                        return [[176. 194.]]
                else:  # if diagnosis_Malignant neoplasm: Breast, unspecified > 0.5
                    return [[ 62. 116.]]
            else:  # if hospital_Ysbyty Ystrad Fawr > 0.5
                return [[1282.  119.]]
        else:  # if diagnosis_Diverticular disease of large intestine without
        perforation or abscess > 0.5
            return [[834.   9.]]
    else:  # if No_Scans > 0.5
        return [[ 18. 100.]]
else:  # if diagnosis_Calculus of gallbladder with other cholecystitis > 0.5
    return [[ 79. 223.]]
else:  # if specialty_Trauma & Orthopaedic > 0.5
    if hospital_Ysbyty Ystrad Fawr <= 0.5:
        if diagnosis_Gonarthrosis, unspecified <= 0.5:
            if diagnosis_Coxarthrosis, unspecified <= 0.5:
                if diagnosis_Carpal tunnel syndrome <= 0.5:
                    if diagnosis_Palmar fascial fibromatosis [Dupuytren] <= 0.5:
                        return [[1379. 1726.]]
                    else:  # if diagnosis_Palmar fascial fibromatosis [Dupuytren] > 0.5
                        return [[154.  17.]]
                else:  # if diagnosis_Carpal tunnel syndrome > 0.5
                    return [[317.   9.]]
            else:  # if diagnosis_Coxarthrosis, unspecified > 0.5
                return [[ 108. 1030.]]
        else:  # if diagnosis_Gonarthrosis, unspecified > 0.5
            return [[  74. 1577.]]
    else:  # if hospital_Ysbyty Ystrad Fawr > 0.5
        return [[1073.  156.]]
```

C.4 Deterministic Parameters

This section contains the deterministic parameters that are used within the prescriptive models discussed in Chapter 5.

Parameter	Deterministic Parameters
Bands ($\mathcal{B}$)	b = [band5, band6]
Specialties ($\mathcal{S}$)	s = [Accident and Emergency, Anaesthetics, Cardiology, Care of the Elderly, Community Medicine, Dermatology, Diabetes and Endocrinology, Ear, Nose and Throat, Gastroenterology, General Medicine, General Surgery, GP Other, Gynaecology, Haematology, Infectious Diseases, Intermediate Care, Maxillo-Facial, Neurology, Ophthalmology, Pain, Plastic Surgery, Radiology, Radiotherapy and Oncology, Rehabilitation, Respiratory, Restorative Dentistry, Rheumatology, Trauma and Orthopaedic, Urology]
Hospitals ($\mathcal{H}$)	h = [Royal Gwent Hospital, St Woolos Acute Hospital, St Woolos Community Hospital, Ysbyty Ystrad Fawr, Rhymney Integrated Health and Social Care Centre, Ysbyty Aneurin Bevan, County Hospital, Nevill Hall Hospital, Chepstow Community Hospital, Monnow Vale Integrated Health and Social Care Centre, University Hospital of Wales, Offsite, Outsource, Outsource - CareUK]
Regions ($\mathcal{R}$)	r = [Region1, Region2, Region3, Region4, Region5, Region6]
$c_b^{\text{staff, 1st}}$	[£338.88, £419.52]

$UB_h^{\text{max, bed, 1st}}$	[588, 28, 75, 163, 12, 80, 73, 309, 26, 16, 90]
$UB_s^{\text{max, staff, 1st}}$	[400,400]
$\mathrm{c}_{s,h}^{\text{bed, 1st}}$	$\begin{bmatrix} 345 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 149 & 0 & 0 & 0 \\ 526 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1516 & 0 & 0 & 0 \\ 396 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 895 & 0 & 0 & 551 \\ 457 & 0 & 755 & 493 & 0 & 542 & 0 & 472 & 743 & 577 & 0 & 0 \\ 0 & 0 & 0 & 942 & 0 & 0 & 0 & 1100 & 0 & 1021 & 0 & 0 \\ 1672 & 216 & 1232 & 2404 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 296 & 0 & 0 & 1270 & 0 & 0 & 0 & 0 & 1497 & 0 & 0 & 0 \\ 481 & 0 & 0 & 501 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 448 & 402 & 0 & 639 & 0 & 0 & 0 & 0 & 959 & 0 & 0 & 832 \\ 390 & 0 & 0 & 94 & 0 & 65 & 0 & 0 & 611 & 0 & 0 & 0 \\ 472 & 0 & 0 & 304 & 0 & 0 & 0 & 0 & 539 & 541 & 0 & 849 \\ 0 & 0 & 0 & 0 & 172 & 0 & 0 & 0 & 0 & 325 & 443 & 360 \\ 1007 & 0 & 0 & 528 & 0 & 0 & 0 & 0 & 292 & 0 & 0 & 241 \\ 1299 & 1277 & 0 & 1218 & 0 & 0 & 0 & 0 & 1070 & 0 & 0 & 1176 \\ 711 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 39 & 0 & 197 & 0 & 0 & 0 & 0 \\ 2026 & 1940 & 0 & 0 & 0 & 0 & 0 & 0 & 264 & 0 & 0 & 0 \\ 1273 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 957 & 0 & 0 & 369 & 0 & 0 & 0 & 0 & 1416 & 0 & 0 & 174 \\ 124 & 0 & 0 & 111 & 0 & 0 & 0 & 134 & 143 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1308 & 0 & 0 & 496 \\ 945 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1097 & 0 & 0 & 0 \\ 633 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1545 \\ 1973 & 1274 & 972 & 975 & 0 & 1021 & 0 & 1983 & 1987 & 1455 & 0 & 0 \\ 330 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 566 & 0 & 0 & 0 \\ 141 & 0 & 0 & 139 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 631 & 0 & 0 & 621 & 0 & 0 & 0 & 0 & 536 & 0 & 0 & 0 \\ 678 & 651 & 0 & 610 & 0 & 0 & 0 & 0 & 873 & 0 & 0 & 0 \\ 102 & 214 & 0 & 79 & 0 & 0 & 0 & 0 & 378 & 0 & 0 & 1122 \end{bmatrix}$

$R_{s,b}$	$\begin{bmatrix} 0.250 & 0.25 \\ 0.250 & 0.25 \\ 0.125 & 0.125 \\ 0.1 & 0.1 \\ 0.1 & 0.1 \\ 0.1 & 0.1 \\ 0.125 & 0.125 \\ 0.125 & 0.125 \\ 0.125 & 0.125 \\ 0.25 & 0.25 \\ 0.25 & 0.25 \\ 0.1 & 0.1 \\ 0.125 & 0.125 \\ 0.125 & 0.125 \\ 0.125 & 0.125 \\ 0.1 & 0.1 \\ 0.125 & 0.125 \\ 0.125 & 0.125 \\ 0.125 & 0.125 \\ 0.125 & 0.125 \\ 0.125 & 0.125 \\ 0.125 & 0.125 \\ 0.125 & 0.125 \\ 0.1 & 0.1 \\ 0.125 & 0.125 \\ 0.125 & 0.125 \\ 0.125 & 0.125 \\ 0.25 & 0.25 \\ 0.125 & 0.125 \end{bmatrix}$
$K_{s,h}$	$\begin{bmatrix} 588 & 0 & 0 & 0 & 0 & 0 & 0 & 309 & 0 & 0 & 0 \\ 588 & 0 & 0 & 0 & 0 & 0 & 0 & 309 & 0 & 0 & 0 \\ 588 & 0 & 0 & 0 & 0 & 0 & 0 & 309 & 0 & 0 & 80 \\ 588 & 0 & 75 & 163 & 0 & 80 & 73 & 309 & 26 & 0 & 0 \\ 588 & 0 & 0 & 163 & 0 & 0 & 73 & 0 & 26 & 0 & 0 \\ 0 & 28 & 75 & 163 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 588 & 0 & 0 & 163 & 0 & 0 & 0 & 309 & 0 & 0 & 0 \\ 588 & 0 & 0 & 163 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 588 & 28 & 0 & 163 & 0 & 0 & 0 & 309 & 0 & 0 & 80 \\ 588 & 0 & 0 & 163 & 0 & 80 & 0 & 309 & 0 & 0 & 0 \\ 588 & 0 & 0 & 163 & 0 & 0 & 0 & 309 & 26 & 0 & 80 \\ 0 & 0 & 0 & 0 & 10 & 0 & 0 & 0 & 26 & 16 & 80 \\ 588 & 0 & 0 & 163 & 0 & 0 & 0 & 309 & 0 & 0 & 80 \\ 588 & 28 & 0 & 163 & 0 & 0 & 0 & 309 & 0 & 0 & 80 \\ 588 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 80 & 73 & 0 & 0 & 0 & 0 \\ 588 & 28 & 0 & 0 & 0 & 0 & 0 & 309 & 0 & 0 & 0 \\ 588 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 588 & 0 & 0 & 163 & 0 & 0 & 0 & 309 & 0 & 0 & 80 \\ 588 & 0 & 0 & 163 & 0 & 0 & 73 & 309 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 309 & 0 & 0 & 80 \\ 588 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 588 & 0 & 0 & 0 & 0 & 0 & 0 & 309 & 0 & 0 & 80 \\ 588 & 28 & 75 & 163 & 0 & 80 & 73 & 309 & 26 & 0 & 0 \\ 588 & 0 & 0 & 0 & 0 & 0 & 0 & 309 & 0 & 0 & 0 \\ 588 & 0 & 0 & 163 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 588 & 0 & 0 & 163 & 0 & 0 & 0 & 309 & 0 & 0 & 0 \\ 588 & 28 & 0 & 163 & 0 & 0 & 0 & 309 & 0 & 0 & 0 \\ 588 & 28 & 0 & 163 & 0 & 0 & 0 & 309 & 0 & 0 & 80 \end{bmatrix}$

Table C.6: Data and parameters used within the deterministic model.

C.5 Two-Stage Stochastic Parameters

This section contains the two-stage stochastic parameters that are used within the prescriptive models discussed in Chapter 5.

Table C.7: Data and parameters used within two-stage stochastic model.

Parameter	**Two-Stage Stochastic Parameters**
Bands ($\mathcal{B}$)	b = [band5, band6]
Specialties ($\mathcal{S}$)	s = [Accident and Emergency, Anaesthetics, Cardiology, Care of the Elderly, Community Medicine, Dermatology, Diabetes and Endocrinology, Ear, Nose and Throat, Gastroenterology, General Medicine, General Surgery, GP Other, Gynaecology, Haematology, Infectious Diseases, Intermediate Care, Maxillo-Facial, Neurology, Ophthalmology, Pain, Plastic Surgery, Radiology, Radiotherapy and Oncology, Rehabilitation, Respiratory, Restorative Dentistry, Rheumatology, Trauma and Orthopaedic, Urology]
Hospitals ($\mathcal{H}$)	h = [Royal Gwent Hospital, St Woolos Acute Hospital, St Woolos Community Hospital, Ysbyty Ystrad Fawr, Rhymney Integrated Health and Social Care Centre, Ysbyty Aneurin Bevan, County Hospital, Nevill Hall Hospital, Chepstow Community Hospital, Monnow Vale Integrated Health and Social Care Centre, University Hospital of Wales, Offsite, Outsource, Outsource - CareUK]
Regions ($\mathcal{R}$)	r = [Region1, Region2, Region3, Region4, Region5, Region6]
Scenarios ($\mathcal{S}$)	s = [Scenario1, Scenario2, Scenario3, Scenario 4]
p_k	[$0.3\dot{3}$, $0.3\dot{3}$, $0.3\dot{3}$]
$\mathrm{c}_b^{\text{staff, 1st}}$	[£338.88, £419.52]
$\mathrm{c}_b^{\text{staff, 2nd}}$	[£454.80, £560.64]
$UB_h^{\text{max, bed, 1st}}$	[588, 28, 75, 163, 12, 80, 73, 309, 26, 16, 20, 20, 20, 20]
$UB_h^{\text{max, bed, 2nd}}$	[177, 9, 28, 49, 4, 24, 22, 93, 8, 5, 6, 6, 6, 6]
$UB_s^{\text{max, staff, 1st}}$	[400,400]
$UB_s^{\text{max, staff, 2nd}}$	[200,200]

$$
c^{\text{bed, 1st}}_{s,h} \quad
\begin{bmatrix}
345 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 149 & 0 & 0 & 0 \\
526 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1516 & 0 & 0 & 0 \\
396 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 895 & 0 & 0 & 551 \\
457 & 0 & 755 & 493 & 0 & 542 & 0 & 472 & 743 & 577 & 0 & 0 \\
0 & 0 & 0 & 942 & 0 & 0 & 0 & 1100 & 0 & 1021 & 0 & 0 \\
1672 & 216 & 1232 & 2404 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
296 & 0 & 0 & 1270 & 0 & 0 & 0 & 0 & 1497 & 0 & 0 & 0 \\
481 & 0 & 0 & 501 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
448 & 402 & 0 & 639 & 0 & 0 & 0 & 0 & 882 & 0 & 0 & 823 \\
390 & 0 & 0 & 94 & 0 & 65 & 0 & 0 & 611 & 0 & 0 & 0 \\
472 & 0 & 0 & 304 & 0 & 0 & 0 & 0 & 539 & 541 & 0 & 849 \\
0 & 0 & 0 & 0 & 172 & 0 & 0 & 0 & 0 & 325 & 443 & 360 \\
1007 & 0 & 0 & 528 & 0 & 0 & 0 & 0 & 292 & 0 & 0 & 241 \\
1299 & 1277 & 0 & 1218 & 0 & 0 & 0 & 0 & 1070 & 0 & 0 & 1176 \\
711 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 39 & 0 & 197 & 0 & 0 & 0 & 0 \\
2026 & 1940 & 0 & 0 & 0 & 0 & 0 & 0 & 264 & 0 & 0 & 0 \\
1273 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
957 & 0 & 0 & 369 & 0 & 0 & 0 & 0 & 1416 & 0 & 0 & 174 \\
124 & 0 & 0 & 111 & 0 & 0 & 0 & 134 & 143 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1308 & 0 & 0 & 496 \\
945 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1097 & 0 & 0 & 0 \\
633 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1545 \\
1973 & 1274 & 972 & 975 & 0 & 1021 & 0 & 1983 & 1987 & 1455 & 0 & 0 \\
330 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 566 & 0 & 0 & 0 \\
141 & 0 & 0 & 139 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
631 & 0 & 0 & 621 & 0 & 0 & 0 & 0 & 536 & 0 & 0 & 0 \\
678 & 651 & 0 & 610 & 0 & 0 & 0 & 0 & 873 & 0 & 0 & 0 \\
102 & 214 & 0 & 79 & 0 & 0 & 0 & 0 & 378 & 0 & 0 & 1122
\end{bmatrix}
$$

	414	0	0	0	0	0	0	0	178.80	0	0	0
	631.20	0	0	0	0	0	0	0	1819.20	0	0	0
	475.20	0	0	0	0	0	0	0	1074	0	0	661.20
	548.40	0	906	591.60	0	650.40	0	566.40	891.60	692.40	0	0
	0	0	0	1130.40	0	0	0	1320	0	1225.20	0	0
	2006.40	259.20	1478.40	2884.80	0	0	0	0	0	0	0	0
	355.20	0	0	1524	0	0	0	0	1796.40	0	0	0
	577.20	0	0	601.20	0	0	0	0	0	0	0	0
	537.60	482.40	0	766.80	0	0	0	0	1150.80	0	0	998.40
	468	0	0	112.80	0	78	0	0	733.20	0	0	0
	566.40	0	0	364.80	0	0	0	0	646.80	649.20	0	1018.80
	0	0	0	0	206.40	0	0	0	0	390	531.60	432
	1208.40	0	0	633.60	0	0	0	0	350.4	0	0	289.20
	1558.80	1532.40	0	1461.60	0	0	0	0	1284	0	0	1411.20
$c_{s,h}^{\text{bed, 2nd}}$	853.20	0	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	46.80	0	236.40	0	0	0	0
	2431.20	2328	0	0	0	0	0	0	316.80	0	0	0
	1527.60	0	0	0	0	0	0	0	0	0	0	0
	1148.40	0	0	442.80	0	0	0	0	1699.20	0	0	208.80
	148.80	0	0	133.20	0	0	0	160.80	171.60	0	0	0
	0	0	0	00	0	0	0	1569.60	0	0	595.20	
	1134	0	00	0	0	0	0	1316.40	0	0	0	
	759.60	0	00	0	0	0	0	0	0	0	1854	
	2367.60	1528.80	1166.40	1170	0	1225.20	0	2379.60	2384.40	1746	0	0
	396	0	00	0	0	0	0	679.20	0	0	0	
	169.20	00	166.80	0	0	0	0	0	0	0	0	
	757.20	00	745.20	0	0	0	0	643.20	0	0	0	
	813.60	781.20	0	7320	0	0	0	1047.60	0	0	0	
	122.40	256.80	0	94.80	0	0	0	0	453.60	0	0	1346.40

	0.250	0.25
	0.250	0.25
	0.125	0.125
	0.1	0.1
	0.1	0.1
	0.1	0.1
	0.125	0.125
	0.125	0.125
	0.125	0.125
	0.25	0.25
	0.25	0.25
	0.1	0.1
	0.125	0.125
	0.125	0.125
$\mathrm{R}_{s,b}$	0.125	0.125
	0.1	0.1
	0.125	0.125
	0.125	0.125
	0.125	0.125
	0.125	0.125
	0.125	0.125
	0.125	0.125
	0.125	0.125
	0.1	0.1
	0.125	0.125
	0.125	0.125
	0.125	0.125
	0.25	0.25
	0.125	0.125

$K_{s,h}$	$\begin{bmatrix} 588 & 0 & 0 & 0 & 0 & 0 & 0 & 309 & 0 & 0 & 0 \\ 588 & 0 & 0 & 0 & 0 & 0 & 0 & 309 & 0 & 0 & 0 \\ 588 & 0 & 0 & 0 & 0 & 0 & 0 & 309 & 0 & 0 & 80 \\ 588 & 0 & 75 & 163 & 0 & 80 & 73 & 309 & 26 & 0 & 0 \\ 588 & 0 & 0 & 163 & 0 & 0 & 73 & 0 & 26 & 0 & 0 \\ 0 & 28 & 75 & 163 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 588 & 0 & 0 & 163 & 0 & 0 & 0 & 309 & 0 & 0 & 0 \\ 588 & 0 & 0 & 163 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 588 & 28 & 0 & 163 & 0 & 0 & 0 & 309 & 0 & 0 & 80 \\ 588 & 0 & 0 & 163 & 0 & 80 & 0 & 309 & 0 & 0 & 0 \\ 588 & 0 & 0 & 163 & 0 & 0 & 0 & 309 & 26 & 0 & 80 \\ 0 & 0 & 0 & 0 & 10 & 0 & 0 & 0 & 26 & 16 & 80 \\ 588 & 0 & 0 & 163 & 0 & 0 & 0 & 309 & 0 & 0 & 80 \\ 588 & 28 & 0 & 163 & 0 & 0 & 0 & 309 & 0 & 0 & 80 \\ 588 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 80 & 73 & 0 & 0 & 0 & 0 \\ 588 & 28 & 0 & 0 & 0 & 0 & 0 & 309 & 0 & 0 & 0 \\ 588 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 588 & 0 & 0 & 163 & 0 & 0 & 0 & 309 & 0 & 0 & 80 \\ 588 & 0 & 0 & 163 & 0 & 0 & 73 & 309 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 309 & 0 & 0 & 80 \\ 588 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 588 & 0 & 0 & 0 & 0 & 0 & 0 & 309 & 0 & 0 & 80 \\ 588 & 28 & 75 & 163 & 0 & 80 & 73 & 309 & 26 & 0 & 0 \\ 588 & 0 & 0 & 0 & 0 & 0 & 0 & 309 & 0 & 0 & 0 \\ 588 & 0 & 0 & 163 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 588 & 0 & 0 & 163 & 0 & 0 & 0 & 309 & 0 & 0 & 0 \\ 588 & 28 & 0 & 163 & 0 & 0 & 0 & 309 & 0 & 0 & 0 \\ 588 & 28 & 0 & 163 & 0 & 0 & 0 & 309 & 0 & 0 & 80 \end{bmatrix}$

Appendix D

Linking Predictive and Prescriptive Paradigms Further Material

This section provides additional information and visual aids to supplement the analysis presented in the Section 6.2. Specifically, this appendix includes demand tables and heatmaps that provide a more detailed breakdown of the data used in the analysis and its results.

The demand tables are presented in a tabular format and provide detailed information on the level of demand for each specialty and region. These tables are generated using the CART model result end nodes. The heatmaps provide a visual representation of the data used in the analysis. Heatmaps are a graphical representation of data that use colour-coding to indicate the intensity of a particular variable. In the case of the heatmaps included in this appendix, the intensity of the colour represents the number of beds to be deployed to each hospital and specialty. The heatmaps are presented in a visual format that allows for quick and easy interpretation of the data.

D.1 Regression Trees - Average LOS

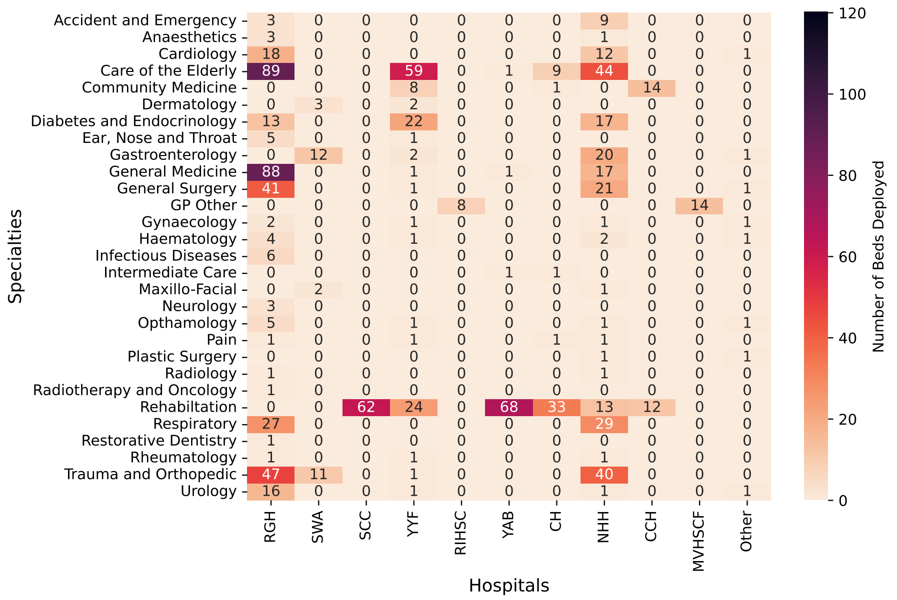


Figure D.1: Heatmap of bed locations for each specialty within each hospital for the deterministic model using the regression tree and average LOS over three years' worth of data.

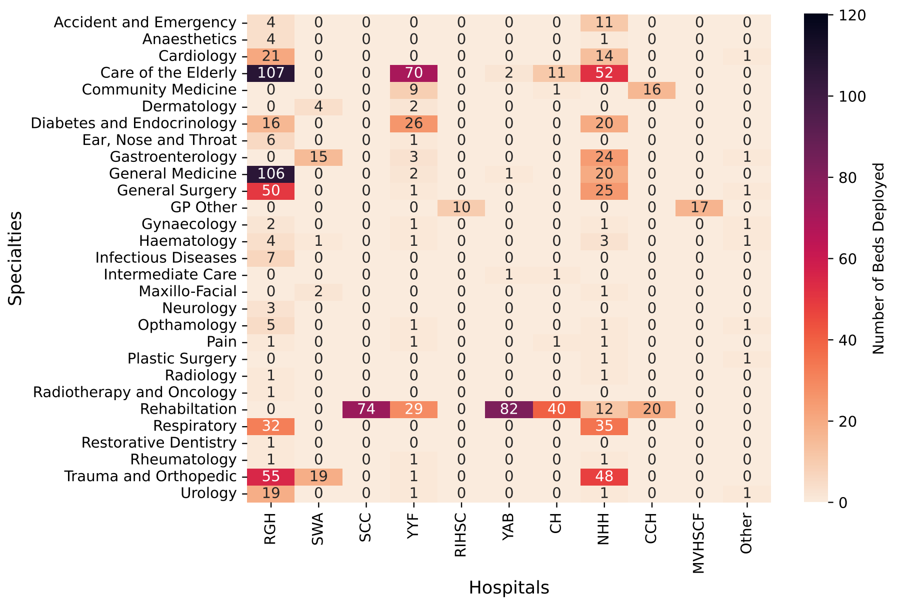


Figure D.2: Heatmap of bed locations for each specialty within each hospital for the two-stage stochastic model using the regression tree and average LOS over three years' worth of data.

Specialty	**Region 1**	**Region 2**	**Region 3**	**Region 4**	**Region 5**	**Region 6**
Accident & Emergency	2.9774	0.0000	0.0000	0.0000	8.5806	0.0000
Anaesthetics	2.6186	0.0000	0.0000	0.0000	0.4500	0.0000
Cardiology	17.4979	0.0000	0.0000	0.0000	11.6617	0.0004
Care of the Elderly	88.9469	58.0161	0.9593	8.6464	43.2786	0.0000
Community Medicine	0.0000	7.0858	0.0000	0.3525	13.1956	0.0000
Dermatology	2.9698	1.0093	0.0000	0.0000	0.0000	0.0000
Diabetes and Endocrinology	12.6673	21.6499	0.0000	0.0000	16.3229	0.0000
Ear Nose & Throat	4.2248	0.0051	0.0000	0.0000	0.0000	0.0000
Gastroenterology	11.7004	1.9326	0.0000	0.0000	19.8421	0.1022
General Medicine	87.7316	0.9656	0.0217	0.0000	16.1179	0.0000
General Surgery	40.9970	0.8121	0.0000	0.0000	20.7664	0.0011
GP Other	0.0000	7.9065	0.0000	0.0000	13.9381	0.0000
Gynaecology	1.6257	0.1061	0.0000	0.0000	0.8319	0.0004
Haematology	3.9436	0.0651	0.0000	0.0000	1.7608	0.0004
Infectious Diseases	5.4554	0.0000	0.0000	0.0000	0.0000	0.0000
Intermediate Care	0.0000	0.0000	0.2191	0.2345	0.0000	0.0000
Maxillo-Facial	1.2359	0.0000	0.0000	0.0000	0.0926	0.0000
Neurology	2.4671	0.0000	0.0000	0.0000	0.0000	0.0000
Ophthalmology	4.0402	0.0123	0.0000	0.0000	0.4399	0.7538
Pain	0.1110	0.0110	0.0000	0.0151	0.0452	0.0000
Plastic Surgery	0.0000	0.0000	0.0000	0.0000	0.0483	0.0004
Radiology	0.0233	0.0000	0.0000	0.0000	0.0109	0.0000
Radiotherapy and Oncology	0.0048	0.0000	0.0000	0.0000	0.0000	0.0000
Rehabilitation	61.4494	23.7536	67.9427	32.8620	24.2661	0.0000
Respiratory	26.5202	0.0000	0.0000	0.0000	28.8608	0.0000
Restorative Dentistry	0.0007	0.0000	0.0000	0.0000	0.0000	0.0000
Rheumatology	0.0004	0.0491	0.0000	0.0000	0.0072	0.0000
Trauma & Orthopaedic	57.7847	0.3512	0.0000	0.0000	39.6777	0.0000
Urology	15.2014	0.1453	0.0000	0.0000	0.5106	0.0202

Table D.1: The daily bed demands for each specialty grouped by regions within ABUHB for three years' worth of patient admissions, using the regression tree and average LOS.

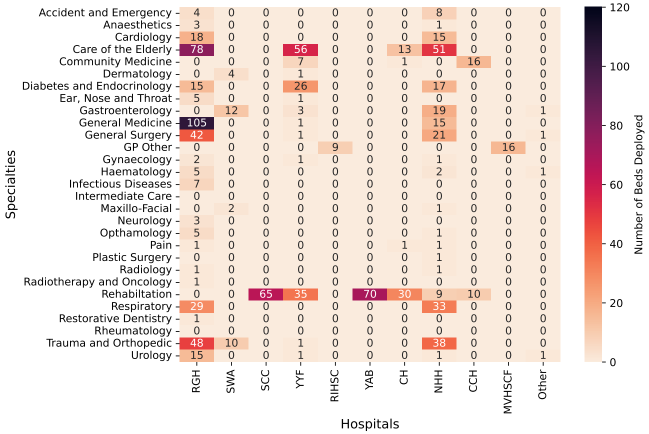


Figure D.3: Heatmap of bed locations for each specialty within each hospital for the deterministic model using the regression tree and average LOS for 2017-2018.

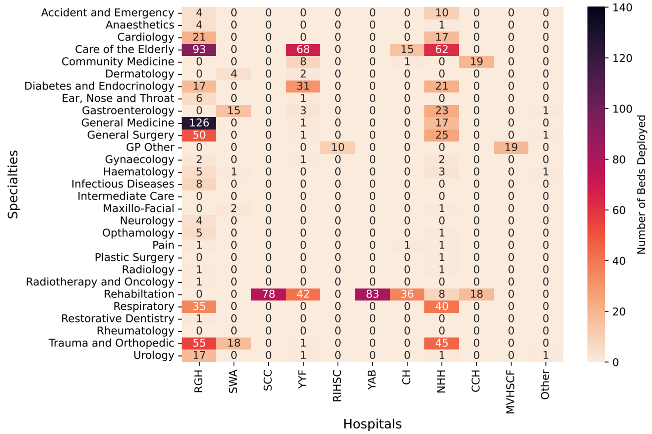


Figure D.4: Heatmap of bed locations for each specialty within each hospital for the two-stage stochastic model using the regression tree and average LOS for 2017-2018.

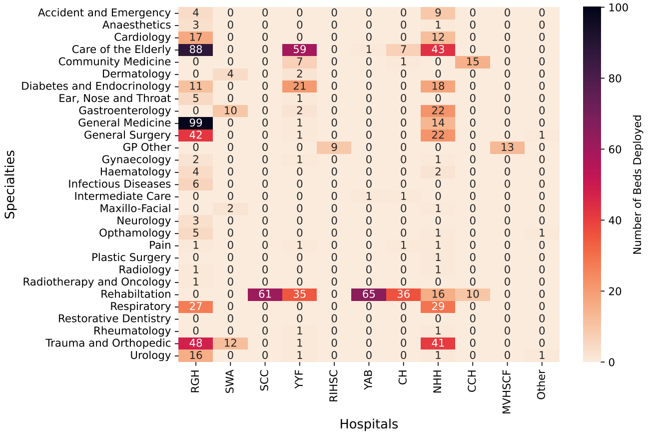


Figure D.5: Heatmap of bed locations for each specialty within each hospital for the deterministic model using the regression tree and average LOS for 2018-2019.

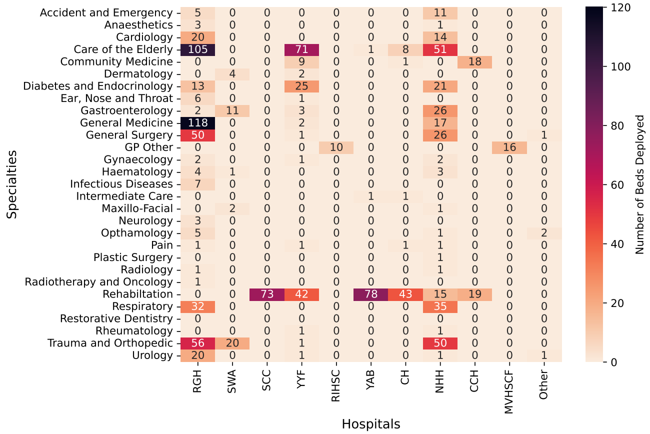


Figure D.6: Heatmap of bed locations for each specialty within each hospital for the two-stage stochastic model using the regression tree and average LOS for 2018-2019.

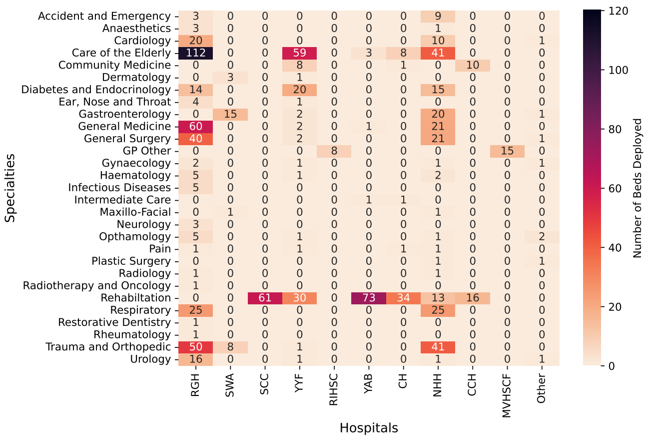


Figure D.7: Heatmap of bed locations for each specialty within each hospital for the deterministic model using the regression tree and average LOS for 2019-2020.

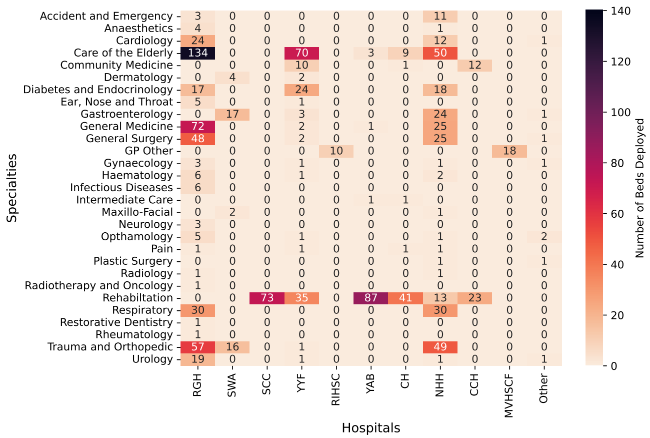


Figure D.8: Heatmap of bed locations for each specialty within each hospital for the two-stage stochastic model using the regression tree and average LOS for 2019-2020.

Specialty	**Region 1**			**Region 2**			**Region 3**		
	2017-2018	**2018-2019**	**2019-2020**	**2017-2018**	**2018-2019**	**2019-2020**	**2017-2018**	**2018-2019**	**2019-2020**
Accident & Emergency	3.2770	3.3545	2.2635	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Anaesthetics	2.8343	2.3643	2.6502	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Cardiology	17.0891	16.0723	19.2765	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Care of the Elderly	77.4710	87.2786	111.2406	55.8412	58.6116	58.0388	0.0000	0.2651	2.3439
Community Medicine	0.0000	0.0000	0.0000	6.6472	6.8756	7.8004	0.0000	0.0000	0.0000
Dermatology	3.1669	3.0209	2.8880	0.8694	1.2436	0.9720	0.0000	0.0000	0.0000
Diabetes and Endocrinology	14.1400	10.1330	13.5358	25.2148	20.3217	19.6372	0.0000	0.0000	0.0000
Ear Nose & Throat	4.6569	4.1676	3.8168	0.0104	0.0034	0.0022	0.0000	0.0000	0.0000
Gastroenterology	11.7496	9.1171	14.1127	2.0305	1.8778	1.9677	0.0000	0.0000	0.0000
General Medicine	104.7935	98.2893	59.7662	0.3669	0.8641	1.6425	0.0000	0.0000	0.0577
General Surgery	41.4285	41.4115	39.3531	0.5347	0.7045	1.2181	0.0000	0.0000	0.0000
GP Other	0.0000	0.0000	0.0000	8.0565	8.3022	7.5475	0.0000	0.0000	0.0000
Gynaecology	1.6483	1.3811	1.8413	0.1985	0.0748	0.0530	0.0000	0.0000	0.0000
Haematology	4.3233	3.3572	4.2947	0.0000	0.0000	0.1845	0.0000	0.0000	0.0000
Infectious Diseases	6.3557	5.0486	4.9143	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Intermediate Care	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0267	0.6029
Maxillo-Facial	1.3805	1.3863	0.9831	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Neurology	2.9267	2.2129	2.2365	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Ophthalmology	4.0896	4.1329	4.0669	0.0000	0.0000	0.0360	0.0000	0.0000	0.0000
Pain	0.1175	0.1290	0.0938	0.0000	0.0235	0.0099	0.0000	0.0000	0.0000
Plastic Surgery	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Radiology	0.0069	0.0311	0.0313	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Radiotherapy and Oncology	0.0083	0.0053	0.0033	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Rehabilitation	64.8212	60.5454	60.3122	34.5884	34.4483	29.0229	69.0037	64.8684	72.0949
Respiratory	28.3691	26.0232	24.8722	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Restorative Dentistry	0.0012	0.0000	0.0011	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Rheumatology	0.0000	0.0000	0.0011	0.0000	0.1473	0.0000	0.0000	0.0000	0.0000
Trauma and Orthopaedic	57.1085	59.2144	57.0404	0.3574	0.4668	0.2350	0.0000	0.0000	0.0000
Urology	14.0047	15.9501	15.2457	0.1520	0.1991	0.0735	0.0000	0.0000	0.0000

Table D.2: The daily bed demands for each specialty for regions one, two and three within ABUHB for three individual years' worth of patient admissions, using the regression tree and average LOS.

Specialty	Region 4			Region 5			Region 6		
	2017-2018	2018-2019	2019-2020	2017-2018	2018-2019	2019-2020	2017-2018	2018-2019	2019-2020
Accident & Emergency	0.0000	0.0000	0.0000	7.6197	8.9913	8.8694	0.0000	0.0000	0.0000
Anaesthetics	0.0000	0.0000	0.0000	0.3226	0.3526	0.7021	0.0000	0.0000	0.0000
Cardiology	0.0000	0.0000	0.0000	14.1081	11.6193	9.3513	0.0000	0.0000	0.0011
Care of the Elderly	12.4192	6.5233	7.4180	50.9444	42.2170	40.9048	0.0000	0.0000	0.0000
Community Medicine	0.4653	0.1991	0.4240	15.5867	14.5410	9.9916	0.0000	0.0000	0.0000
Dermatology	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Diabetes and Endocrinology	0.0000	0.0000	0.0000	16.6930	17.3699	14.6714	0.0000	0.0000	0.0000
Ear Nose & Throat	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Gastroenterology	0.0000	0.0000	0.0000	18.6975	21.2773	19.2673	0.3178	0.0000	0.0022
General Medicine	0.0000	0.0000	0.0000	14.0661	13.7832	20.0969	0.0000	0.0000	0.0000
General Surgery	0.0000	0.0000	0.0000	20.4384	21.0748	20.5769	0.0012	0.0011	0.0011
GP Other	0.0000	0.0000	0.0000	15.5313	12.5077	14.4248	0.0000	0.0000	0.0000
Gynaecology	0.0000	0.0000	0.0000	0.9696	0.8862	0.6536	0.0000	0.0000	0.0011
Haematology	0.0000	0.0000	0.0000	1.8745	1.7164	1.5286	0.0012	0.0000	0.0000
Infectious Diseases	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Intermediate Care	0.0000	0.3300	0.3561	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Maxillo-Facial	0.0000	0.0000	0.0000	0.1129	0.0761	0.0971	0.0000	0.0000	0.0000
Neurology	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Opthamology	0.0000	0.0000	0.0000	0.4120	0.4688	0.4645	0.0000	0.9782	1.2986
Pain	0.0173	0.0112	0.0177	0.0541	0.0449	0.0397	0.0000	0.0000	0.0000
Plastic Surgery	0.0000	0.0000	0.0000	0.0548	0.0471	0.0463	0.0000	0.0000	0.0011
Radiology	0.0000	0.0000	0.0000	0.0099	0.0011	0.0223	0.0000	0.0000	0.0000
Radiotherapy and Oncology	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Rehabilitation	29.8605	35.5390	33.8379	18.3697	25.2679	28.8564	0.0000	0.0000	0.0000
Respiratory	0.0000	0.0000	0.0000	32.6054	28.9470	24.7762	0.0000	0.0000	0.0000
Restorative Dentistry	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Rheumatology	0.0000	0.0000	0.0000	0.0000	0.0214	0.0000	0.0000	0.0000	0.0000
Trauma and Orthopedic	0.0000	0.0000	0.0000	37.1555	40.9789	40.7325	0.0000	0.0000	0.0000
Urology	0.0000	0.0000	0.0000	0.5282	0.5165	0.4400	0.0219	0.0181	0.0184

Table D.3: The daily bed demands for each specialty for regions four, five and six within ABUHB for three individual years' worth of patient admissions, using the regression tree and average LOS.

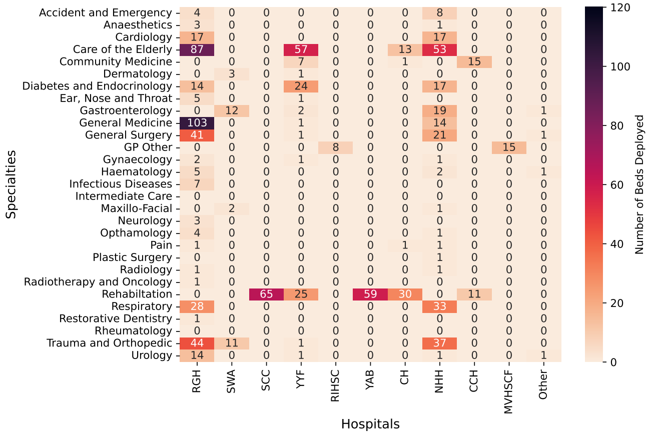


Figure D.9: Heatmap of bed locations for each specialty within each hospital for the deterministic model using the regression tree and average year LOS for 2017-2018.

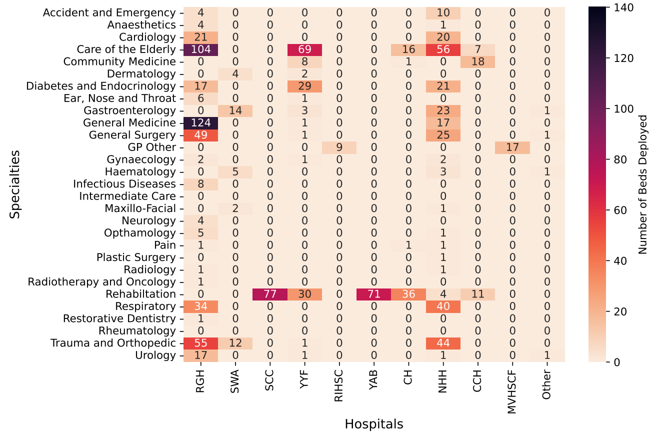


Figure D.10: Heatmap of bed locations for each specialty within each hospital for the two-stage stochastic model using the regression tree and average year LOS for 2017-2018.

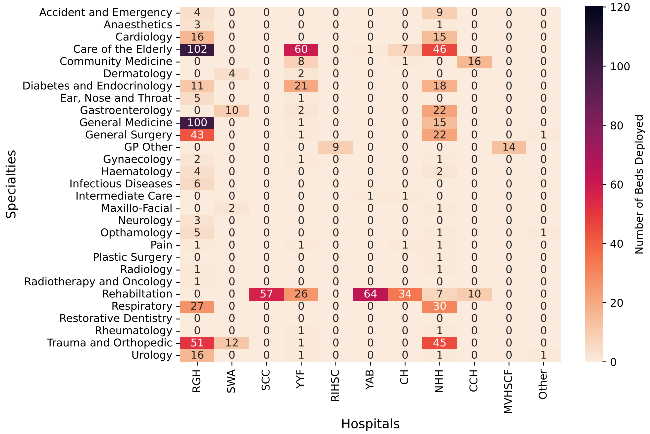


Figure D.11: Heatmap of bed locations for each specialty within each hospital for the deterministic model using the regression tree and average year LOS for 2018-2019.

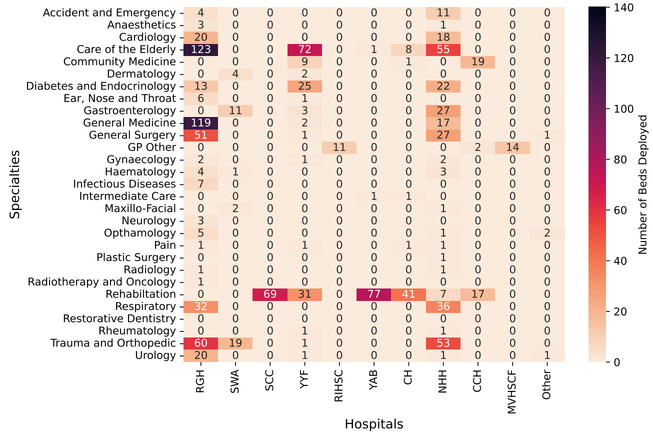


Figure D.12: Heatmap of bed locations for each specialty within each hospital for the two-stage stochastic model using the regression tree and average year LOS for 2018-2019.

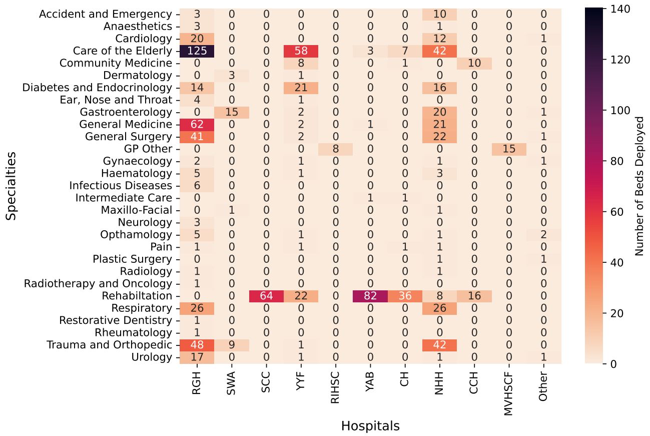


Figure D.13: Heatmap of bed locations for each specialty within each hospital for the deterministic model using the regression tree and average year LOS for 2019-2020.

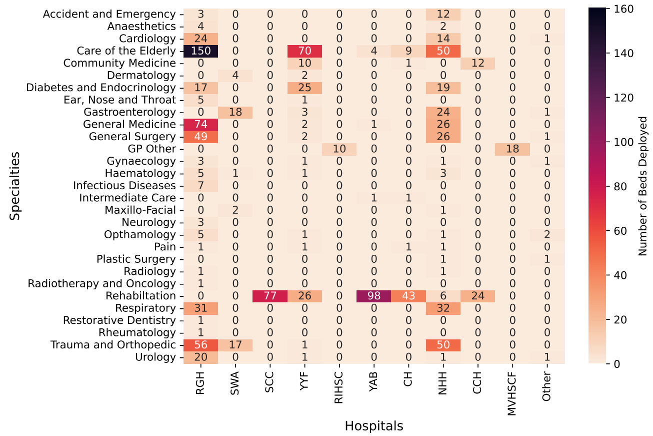


Figure D.14: Heatmap of bed locations for each specialty within each hospital for the two-stage stochastic model using the regression tree and average year LOS for 2019-2020.

Specialty	**Region 1**			**Region 2**			**Region 3**		
	2017-2018	**2018-2019**	**2019-2020**	**2017-2018**	**2018-2019**	**2019-2020**	**2017-2018**	**2018-2019**	**2019-2020**
Accident & Emergency	3.2862	3.1969	2.4505	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Anaesthetics	2.7907	2.3695	2.6954	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Cardiology	16.9908	15.9287	19.5686	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Care of the Elderly	86.2987	101.7951	124.1676	56.7726	59.3095	57.9663	0.0000	0.2609	2.6423
Community Medicine	0.0000	0.0000	0.0000	6.2182	7.2636	7.7736	0.0000	0.0000	0.0000
Dermatology	2.9946	3.0185	2.8964	0.8336	1.2245	0.9701	0.0000	0.0000	0.0000
Diabetes and Endocrinology	13.8852	10.2042	13.9090	23.7736	20.7721	20.4075	0.0000	0.0000	0.0000
Ear Nose & Throat	4.5786	4.1889	3.9077	0.0099	0.0033	0.0022	0.0000	0.0000	0.0000
Gastroenterology	11.5280	9.1310	14.4347	1.9533	1.8463	1.9979	0.0000	0.0000	0.0000
General Medicine	102.9001	99.0452	61.3219	0.3693	0.8681	1.6574	0.0000	0.0000	0.0650
General Surgery	40.7516	42.1739	40.0680	0.5145	0.6929	1.2279	0.0000	0.0000	0.0000
GP Other	0.0000	0.0000	0.0000	7.3663	8.8037	7.5505	0.0000	0.0000	0.0000
Gynaecology	1.6100	1.3820	1.8843	0.1921	0.0735	0.0529	0.0000	0.0000	0.0000
Haematology	4.1320	3.3443	4.3536	0.0000	0.0000	0.1950	0.0000	0.0000	0.0000
Infectious Diseases	6.2340	5.0822	5.0511	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Intermediate Care	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0269	0.6293
Maxillo-Facial	1.3366	1.3814	0.9905	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Neurology	2.8786	2.2274	2.2958	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Ophthalmology	3.9421	4.0835	4.0948	0.0000	0.0000	0.0367	0.0000	0.0000	0.0000
Pain	0.1126	0.1268	0.0936	0.0000	0.0232	0.0099	0.0000	0.0000	0.0000
Plastic Surgery	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Radiology	0.0066	0.0313	0.0318	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Radiotherapy and Oncology	0.0061	0.0049	0.0033	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Rehabilitation	64.1492	56.7391	63.4545	24.5786	25.2688	21.4197	58.7946	63.8572	81.1403
Respiratory	27.8692	26.1755	25.5186	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Restorative Dentistry	0.0011	0.0000	0.0011	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Rheumatology	0.0000	0.0000	0.0011	0.0000	0.1473	0.0000	0.0000	0.0000	0.0000
Trauma & Orthopaedic	54.3186	62.5405	56.4987	0.3162	0.4968	0.2410	0.0000	0.0000	0.0000
Urology	13.6899	15.8597	16.0522	0.1457	0.1941	0.0963	0.0000	0.0000	0.0000

Table D.4: The daily bed demands for each specialty for regions one, two and three within ABUHB for three individual years' worth of patient admissions, using the regression tree and the year specific average LOS.

Specialty	**Region 4**			**Region 5**			**Region 6**		
	2017-2018	**2018-2019**	**2019-2020**	**2017-2018**	**2018-2019**	**2019-2020**	**2017-2018**	**2018-2019**	**2019-2020**
Accident & Emergency	0.0000	0.0000	0.0000	7.6680	8.6804	9.9681	0.0000	0.0000	0.0000
Anaesthetics	0.0000	0.0000	0.0000	0.4638	0.4268	0.9722	0.0000	0.0000	0.0000
Cardiology	0.0000	0.0000	0.0000	16.1539	14.2806	11.1400	0.0000	0.0000	0.0011
Care of the Elderly	12.5281	6.5938	6.8223	52.3094	45.1508	41.2107	0.0000	0.0000	0.0000
Community Medicine	0.4234	0.2117	0.4222	14.1831	15.4549	9.9576	0.0000	0.0000	0.0000
Dermatology	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Diabetes and Endocrinology	0.0000	0.0000	0.0000	16.6698	17.8823	15.3511	0.0000	0.0000	0.0000
Ear Nose & Throat	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Gastroenterology	0.0000	0.0000	0.0000	18.7998	21.9374	19.9778	0.3047	0.0000	0.0022
General Medicine	0.0000	0.0000	0.0000	13.9936	14.0130	20.9445	0.0000	0.0000	0.0000
General Surgery	0.0000	0.0000	0.0000	20.4583	21.6905	21.4020	0.0011	0.0011	0.0011
GP Other	0.0000	0.0000	0.0000	14.1544	13.2848	14.3738	0.0000	0.0000	0.0000
Gynaecology	0.0000	0.0000	0.0000	0.9738	0.9246	0.6943	0.0000	0.0000	0.0011
Haematology	0.0000	0.0000	0.0000	1.8249	1.7582	2.0842	0.0011	0.0000	0.0000
Infectious Diseases	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Intermediate Care	0.0000	0.3333	0.3699	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Maxillo-Facial	0.0000	0.0000	0.0000	0.1061	0.0747	0.0969	0.0000	0.0000	0.0000
Neurology	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Ophthalmology	0.0000	0.0000	0.0000	0.3948	0.4613	0.4636	0.0000	0.9640	1.2960
Pain	0.0166	0.0110	0.0176	0.0519	0.0442	0.0396	0.0000	0.0000	0.0000
Plastic Surgery	0.0000	0.0000	0.0000	0.0521	0.0464	0.0462	0.0000	0.0000	0.0011
Radiology	0.0000	0.0000	0.0000	0.0091	0.0011	0.0225	0.0000	0.0000	0.0000
Radiotherapy and Oncology	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Rehabilitation	29.5478	33.4318	35.5989	10.7206	16.2437	23.0764	0.0000	0.0000	0.0000
Respiratory	0.0000	0.0000	0.0000	32.7445	29.8673	25.8748	0.0000	0.0000	0.0000
Restorative Dentistry	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Rheumatology	0.0000	0.0000	0.0000	0.0000	0.0215	0.0000	0.0000	0.0000	0.0000
Trauma & Orthopaedic	0.0000	0.0000	0.0000	36.3659	44.0053	41.3888	0.0000	0.0000	0.0000
Urology	0.0000	0.0000	0.0000	0.5179	0.4967	0.5494	0.0210	0.0169	0.0226

Table D.5: The daily bed demands for each specialty for regions four, five and six within ABUHB for three individual years' worth of patient admissions, using the regression tree and the year specific average LOS.

D.2 Regression Trees - Specific LOS

Specialty	Region 1	Region 2	Region 3	Region 4	Region 5	Region 6
Accident & Emergency	2.1168	0.0000	0.0000	0.0000	9.1761	0.0000
Anaesthetics	4.5894	0.0000	0.0000	0.0000	0.7719	0.0000
Cardiology	15.5265	0.0000	0.0000	0.0000	9.7901	0.0000
Care of the Elderly	93.9772	57.7354	0.7573	8.7856	46.4398	0.0000
Community Medicine	0.0000	7.0073	0.0000	0.3139	13.0137	0.0000
Dermatology & Endocrinology	2.2591	0.0000	0.0000	0.0000	0.0000	0.0000
Diabetes	14.4489	21.2409	0.0000	0.0000	17.2290	0.0000
Ear, Nose & Throat	3.1104	0.0009	0.0000	0.0000	0.0000	0.0000
Gastroenterology	12.3120	0.0985	0.0000	0.0000	19.7765	0.0009
General Medicine	84.4854	0.9818	0.0119	0.0000	14.1013	0.0000
General Surgery	45.5192	0.2318	0.0000	0.0000	21.3011	0.0000
GP Other	0.0000	9.8723	0.0000	0.0000	15.3102	0.0000
Gynaecology	2.0046	0.0721	0.0000	0.0000	1.0766	0.0000
Haematology	2.7792	0.0265	0.0000	0.0000	1.1332	0.0000
Infectious Diseases	7.1195	0.0000	0.0000	0.0000	0.0000	0.0000
Intermediate Care	0.0000	0.0000	0.3449	0.3239	0.0000	0.0000
Maxillo-Facial	1.0547	0.0000	0.0000	0.0000	0.0027	0.0000
Neurology	1.5620	0.0000	0.0000	0.0000	0.0000	0.0000
Ophthalmology	1.3704	0.0027	0.0000	0.0000	0.0009	0.0036
Pain	0.0018	0.0000	0.0000	0.0018	0.0000	0.0000
Plastic Surgery	0.0000	0.0000	0.0000	0.0000	0.0137	0.0000
Radiology	0.0100	0.0000	0.0000	0.0000	0.0000	0.0000
Radiotherapy and Oncology	0.2245	0.0000	0.0000	0.0000	0.0000	0.0000
Rehabilitation	63.0173	33.0922	69.4863	32.1077	24.5538	0.0000
Respiratory Dentistry	29.4717	0.0000	0.0000	0.0000	27.7290	0.0000
Restorative	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Rheumatology	0.0000	0.0000	0.0000	0.0000	0.0128	0.0000
Trauma	59.1807	0.2956	0.0000	0.0000	40.6989	0.0000
Urology	11.3257	0.0036	0.0000	0.0000	0.0255	0.0401

Table D.6: The daily bed demands for each specialty grouped by regions within ABUHB for three years' worth of patient admissions, using the regression tree and specific LOS.

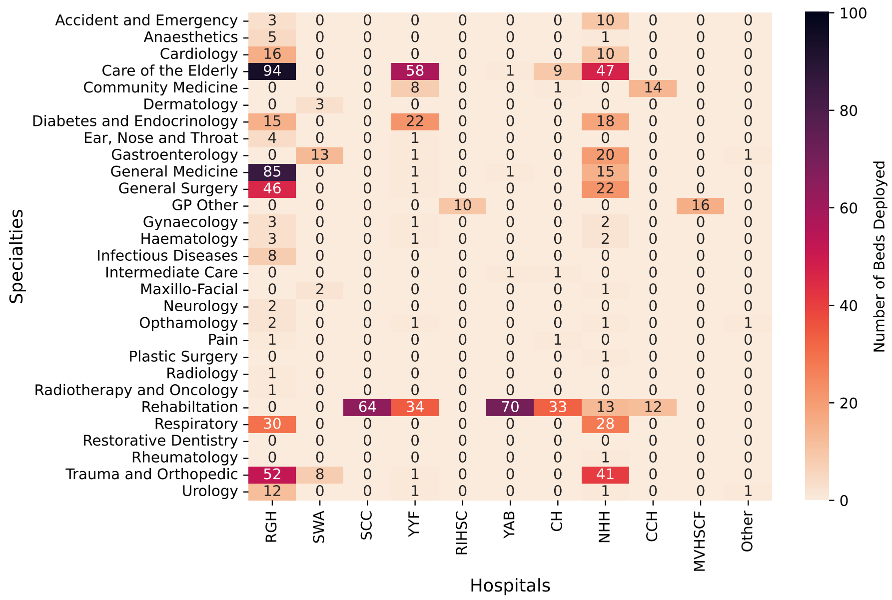


Figure D.15: Heatmap of bed locations for each specialty within each hospital for the deterministic model using the regression tree and specific LOS over three years' worth of data.

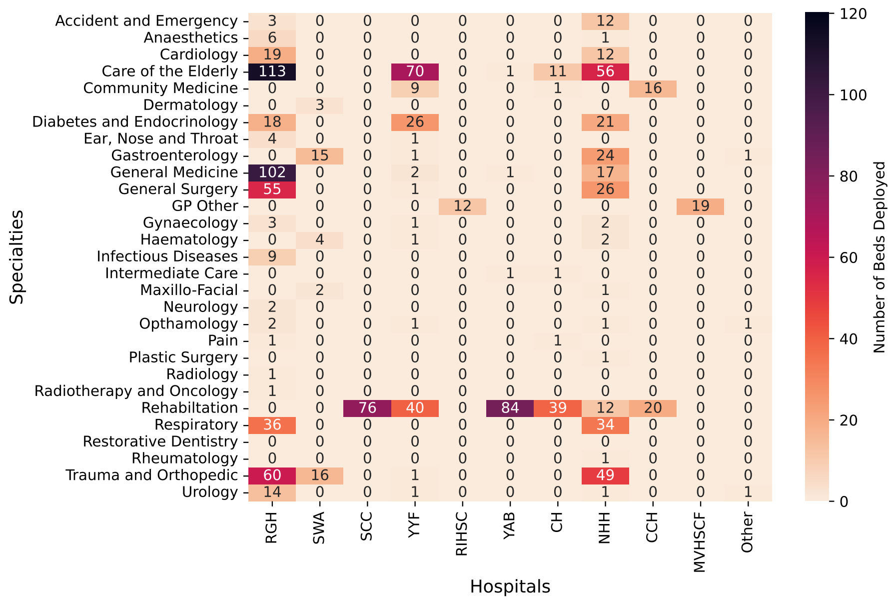


Figure D.16: Heatmap of bed locations for each specialty within each hospital for the two-stage stochastic model using the regression tree and specific LOS over three years' worth of data.

Specialty	Region 1			Region 2			Region 3		
	2017-2018	**2018-2019**	**2019-2020**	**2017-2018**	**2018-2019**	**2019-2020**	**2017-2018**	**2018-2019**	**2019-2020**
Accident & Emergency	2.2932	2.2000	1.8579	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Anaesthetics	4.4932	4.4712	4.8033	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Cardiology	16.7918	14.0630	15.7240	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Care of the Elderly	76.4575	88.0658	117.3443	55.8247	58.4877	58.8907	0.0000	0.4000	1.8689
Community Medicine	0.0000	0.0000	0.0000	7.1671	6.1014	7.7514	0.0000	0.0000	0.0000
Dermatology	2.6877	1.9918	2.0984	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Diabetes and Endocrinology	16.5863	11.4356	15.3224	24.7041	19.8055	19.2186	0.0000	0.0000	0.0000
Ear Nose & Throat	2.9616	3.6192	2.7514	0.0027	0.0000	0.0000	0.0000	0.0000	0.0000
Gastroenterology	11.4438	9.8219	15.6612	0.1260	0.0219	0.1475	0.0000	0.0000	0.0000
General Medicine	105.6712	96.0603	51.8142	0.2877	1.1479	1.5082	0.0000	0.0000	0.0355
General Surgery	47.9068	45.3479	43.3087	0.2000	0.2301	0.2650	0.0000	0.0000	0.0000
GP Other	0.0000	0.0000	0.0000	9.6795	10.1151	9.8224	0.0000	0.0000	0.0000
Gynaecology	2.4301	1.7233	1.8607	0.0795	0.0740	0.0628	0.0000	0.0000	0.0000
Haematology	3.1096	2.4384	2.7896	0.0000	0.0000	0.0792	0.0000	0.0000	0.0000
Infectious Diseases	8.0767	6.3918	6.8907	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Intermediate Care	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0082	1.0246
Maxillo-Facial	1.1068	0.9178	1.1393	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Neurology	1.4301	1.5534	1.7022	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Ophthalmology	1.4712	1.2685	1.3716	0.0000	0.0000	0.0082	0.0000	0.0000	0.0000
Pain	0.0000	0.0000	0.0055	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Plastic Surgery	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Radiology	0.0055	0.0164	0.0082	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Radiotherapy and Oncology	0.0000	0.0000	0.6721	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Rehabilitation	64.7671	63.1151	61.1749	34.9123	34.8384	29.5355	69.9671	65.3863	73.0956
Respiratory	29.8658	30.4137	28.1393	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Restorative Dentistry	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Rheumatology	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Trauma & Orthopaedic	56.0082	60.6164	60.9126	0.3808	0.2712	0.2350	0.0000	0.0000	0.0000
Urology	10.3671	12.5726	11.0383	0.0027	0.0055	0.0027	0.0000	0.0000	0.0000

Table D.7: The daily bed demands for each specialty for regions one, two and three within ABUHB for three individual years' worth of patient admissions, using the regression tree and the year specific LOS.

Specialty	**Region 4**			**Region 5**			**Region 6**		
	2017-2018	**2018-2019**	**2019-2020**	**2017-2018**	**2018-2019**	**2019-2020**	**2017-2018**	**2018-2019**	**2019-2020**
Accident & Emergency	0.0000	0.0000	0.0000	8.5452	10.0630	8.9208	0.0000	0.0000	0.0000
Anaesthetics	0.0000	0.0000	0.0000	0.4247	0.7753	1.1148	0.0000	0.0000	0.0000
Cardiology	0.0000	0.0000	0.0000	10.6603	10.2137	8.5000	0.0000	0.0000	0.0000
Care of the Elderly	12.4192	6.5233	7.4180	53.7178	44.1726	41.4426	0.0000	0.0000	0.0000
Community Medicine	0.3671	0.1397	0.4344	16.7151	14.0466	8.2923	0.0000	0.0000	0.0000
Dermatology	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Diabetes and Endocrinology	0.0000	0.0000	0.0000	17.8575	18.4000	15.4344	0.0000	0.0000	0.0000
Ear Nose & Throat	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Gastroenterology	0.0000	0.0000	0.0000	18.0164	21.3781	19.9344	0.0000	0.0000	0.0027
General Medicine	0.0000	0.0000	0.0000	11.9644	11.7342	18.5929	0.0000	0.0000	0.0000
General Surgery	0.0000	0.0000	0.0000	21.1041	22.6219	20.1803	0.0000	0.0000	0.0000
GP Other	0.0000	0.0000	0.0000	14.7781	13.8740	17.2732	0.0000	0.0000	0.0000
Gynaecology	0.0000	0.0000	0.0000	1.1753	1.4438	0.6120	0.0000	0.0000	0.0000
Haematology	0.0000	0.0000	0.0000	1.2932	1.0438	1.0628	0.0000	0.0000	0.0000
Infectious Diseases	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Intermediate Care	0.0000	0.6000	0.3716	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Maxillo-Facial	0.0000	0.0000	0.0000	0.0000	0.0000	0.0082	0.0000	0.0000	0.0000
Neurology	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Ophthalmology	0.0000	0.0000	0.0000	0.0000	0.0000	0.0027	0.0000	0.0055	0.0055
Pain	0.0055	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Plastic Surgery	0.0000	0.0000	0.0000	0.0000	0.0000	0.0410	0.0000	0.0000	0.0000
Radiology	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Radiotherapy and Oncology	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Rehabilitation	29.6877	33.0521	33.5792	18.3918	25.3288	29.9262	0.0000	0.0000	0.0000
Respiratory	0.0000	0.0000	0.0000	30.6411	29.4301	23.1284	0.0000	0.0000	0.0000
Restorative Dentistry	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Rheumatology	0.0000	0.0000	0.0000	0.0000	0.0384	0.0000	0.0000	0.0000	0.0000
Trauma & Orthopaedic	0.0000	0.0000	0.0000	41.8192	40.3178	39.9617	0.0000	0.0000	0.0000
Urology	0.0000	0.0000	0.0000	0.0110	0.0192	0.0464	0.0384	0.0384	0.0437

Table D.8: The daily bed demands for each specialty for regions four, five and six within ABUHB for three individual years' worth of patient admissions, using the regression tree and the year specific LOS.

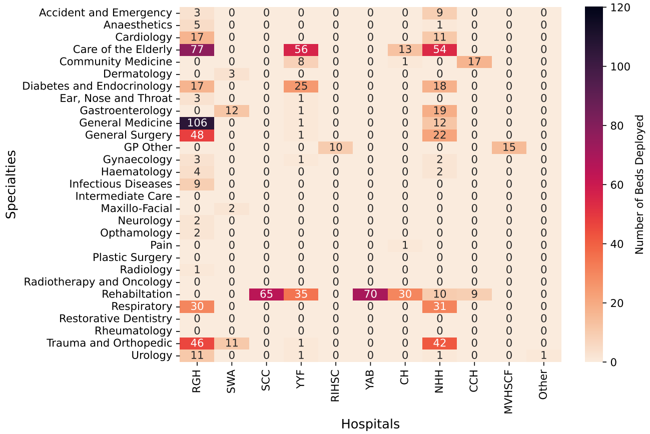


Figure D.17: Heatmap of bed locations for each specialty within each hospital for the deterministic model using the regression tree and specific LOS for 2017-2018.

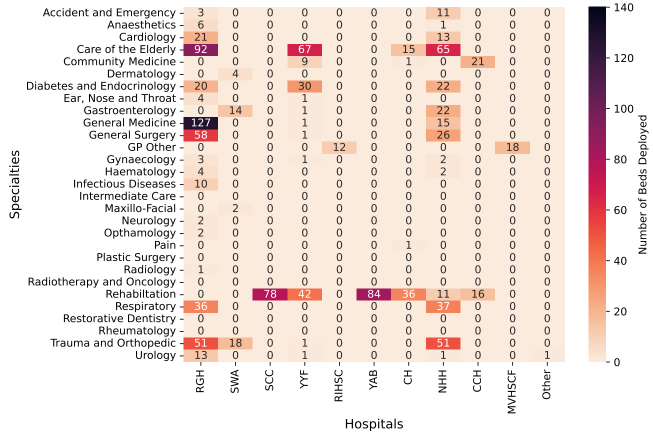


Figure D.18: Heatmap of bed locations for each specialty within each hospital for the two-stage stochastic model using the regression tree and specific LOS for 2017-2018.

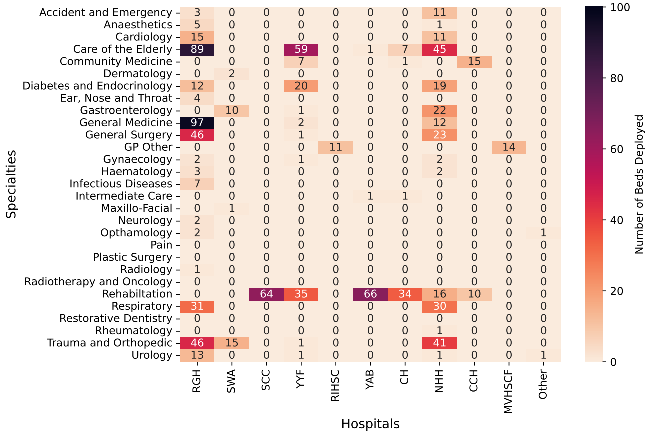


Figure D.19: Heatmap of bed locations for each specialty within each hospital for the deterministic model using the regression tree and specific LOS for 2018-2019.

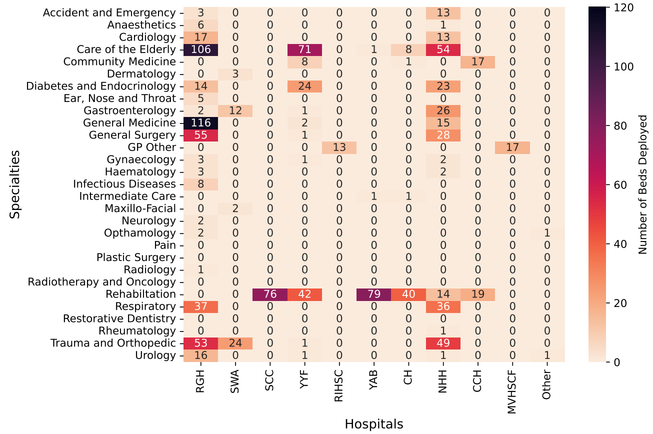


Figure D.20: Heatmap of bed locations for each specialty within each hospital for the two-stage stochastic model using the regression tree and specific LOS for 2018-2019.

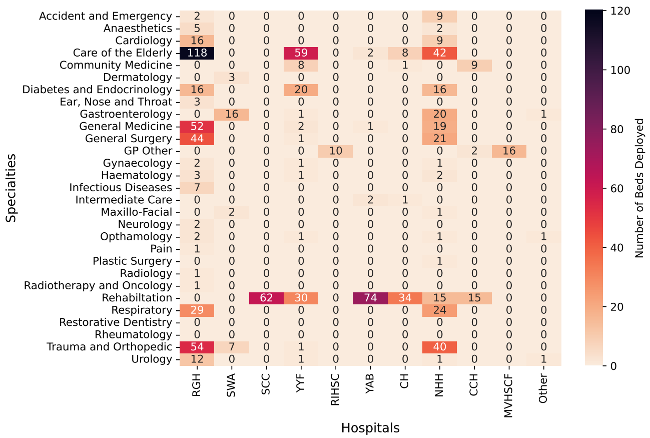


Figure D.21: Heatmap of bed locations for each specialty within each hospital for the deterministic model using the regression tree and specific LOS for 2019-2020.

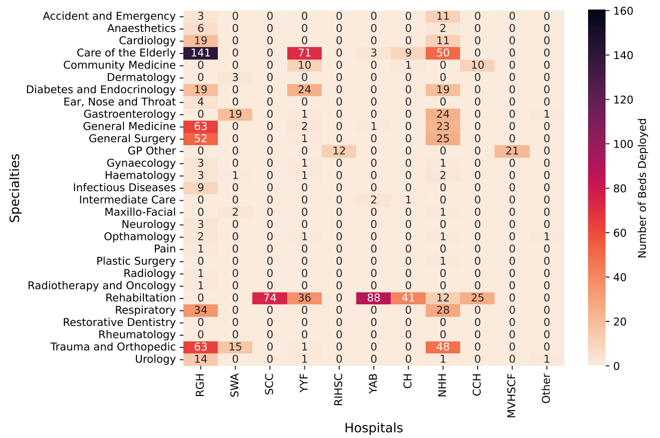


Figure D.22: Heatmap of bed locations for each specialty within each hospital for the two-stage stochastic model using the regression tree and specific LOS for 2019-2020.

D.3 Classification Trees - Average LOS

Specialty	Region 1	Region 2	Region 3	Region 4	Region 5	Region 6
Accident & Emergency	2.9437	0.0000	0.0000	0.0000	8.3492	0.0000
Anaesthetics	4.1522	0.0000	0.0000	0.0000	0.6332	0.0000
Cardiology	24.9347	0.0000	0.0000	0.0000	16.8229	0.0001
Care of the Elderly	119.0511	46.6365	0.5724	3.8356	58.7402	0.0000
Community Medicine	0.0000	3.1172	0.0000	0.1218	4.6880	0.0000
Dermatology	1.6374	0.3027	0.0000	0.0000	0.0000	0.0000
Diabetes and Endocrinology	18.8983	17.7172	0.0000	0.0000	24.1221	0.0000
Ear Nose & Throat	6.9407	0.0178	0.0000	0.0000	0.0000	0.0000
Gastroenterology	15.9809	0.6828	0.0000	0.0000	27.8963	0.0307
General Medicine	83.8888	0.4873	0.0077	0.0000	15.2006	0.0000
General Surgery	64.0241	0.3538	0.0000	0.0000	33.8609	0.0025
GP Other	0.0000	3.3973	0.0000	0.0000	5.1629	0.0000
Gynaecology	2.8824	0.2171	0.0000	0.0000	1.4492	0.0008
Haematology	5.4136	0.0635	0.0000	0.0000	2.1941	0.0001
Infectious Diseases	8.1705	0.0000	0.0000	0.0000	0.0000	0.0000
Intermediate Care	0.0000	0.0000	0.2922	0.3288	0.0000	0.0000
Maxillo-Facial	1.2233	0.0000	0.0000	0.0000	0.0288	0.0000
Neurology	3.6535	0.0000	0.0000	0.0000	0.0000	0.0000
Ophthalmology	2.4715	0.0122	0.0000	0.0000	0.1323	0.2449
Pain	0.0335	0.0033	0.0000	0.0045	0.0136	0.0000
Plastic Surgery	0.0000	0.0000	0.0000	0.0000	0.0146	0.0001
Radiology	0.0262	0.0000	0.0000	0.0000	0.0135	0.0000
Radiotherapy and Oncology	0.0026	0.0000	0.0000	0.0000	0.0000	0.0000
Rehabilitation	26.2528	13.1021	38.7217	14.1127	10.8981	0.0000
Respiratory	39.6264	0.0000	0.0000	0.0000	42.5619	0.0000
Restorative Dentistry	0.0002	0.0000	0.0000	0.0000	0.0000	0.0000
Rheumatology	0.0001	0.0487	0.0000	0.0000	0.0122	0.0000
Trauma & Orthopaedic	64.6427	0.2885	0.0000	0.0000	44.6021	0.0000
Urology	22.2711	0.0411	0.0000	0.0000	0.1389	0.0043

Table D.9: The daily bed demands for each specialty grouped by regions within ABUHB for three years' worth of patient admissions, using the classification tree and average LOS.

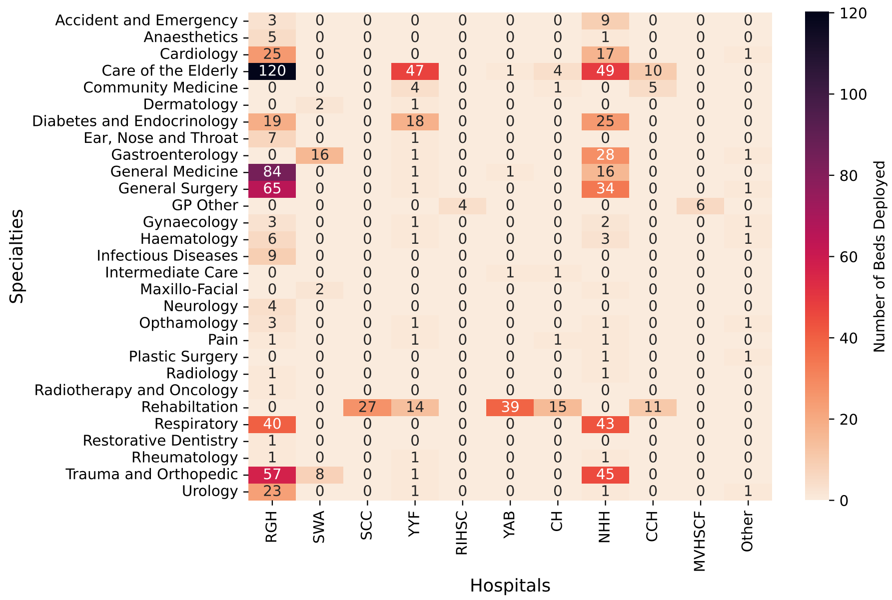


Figure D.23: Heatmap of bed locations for each specialty within each hospital for the deterministic model using the classification tree and average LOS over three years' worth of data.

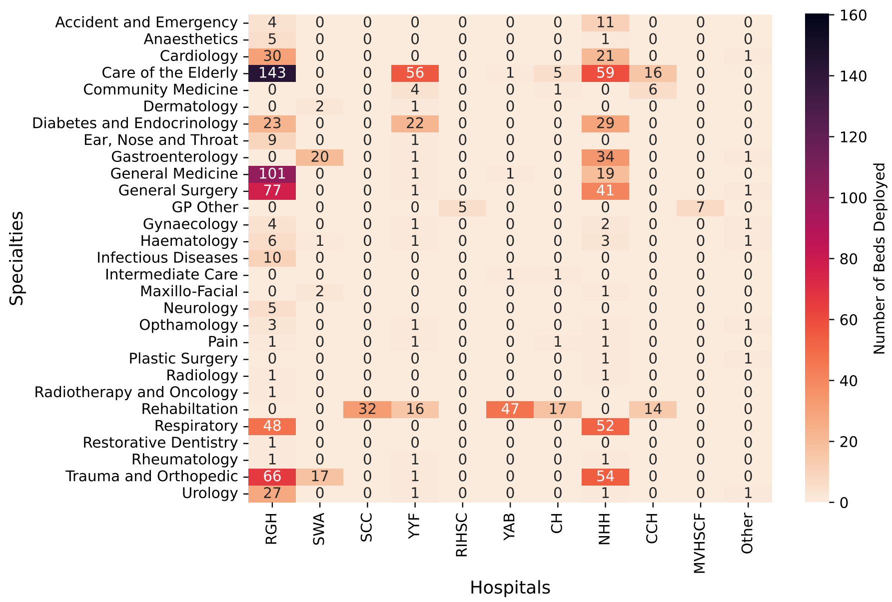


Figure D.24: Heatmap of bed locations for each specialty within each hospital for the two-stage stochastic model using the classification tree and average LOS over three years' worth of data.

Specialty	**Region 1**			**Region 2**			**Region 3**		
	2017-2018	**2018-2019**	**2019-2020**	**2017-2018**	**2018-2019**	**2019-2020**	**2017-2018**	**2018-2019**	**2019-2020**
Accident & Emergency	3.2363	3.1680	2.4283	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Anaesthetics	4.4607	3.7660	4.2297	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Cardiology	24.2501	22.6507	27.8952	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Care of the Elderly	99.2695	117.1490	140.6756	45.0459	48.8485	46.0167	0.0000	0.1463	1.5683
Community Medicine	0.0000	0.0000	0.0000	3.1079	2.9616	3.2817	0.0000	0.0000	0.0000
Dermatology	1.5668	1.8460	1.4996	0.2500	0.3673	0.2910	0.0000	0.0000	0.0000
Diabetes and Endocrinology	20.5858	15.1737	20.9300	19.0863	17.3310	16.7370	0.0000	0.0000	0.0000
Ear Nose & Throat	7.4638	6.8977	6.4620	0.0344	0.0115	0.0076	0.0000	0.0000	0.0000
Gastroenterology	15.5449	12.0964	20.2896	0.7179	0.6021	0.7282	0.0000	0.0000	0.0000
General Medicine	97.5751	94.9051	59.2536	0.2090	0.4413	0.8107	0.0000	0.0000	0.0232
General Surgery	63.3663	66.2451	62.4650	0.2488	0.3001	0.5120	0.0000	0.0000	0.0000
GP Other	0.0000	0.0000	0.0000	3.0713	3.5466	3.5734	0.0000	0.0000	0.0000
Gynaecology	3.0095	2.4954	3.1416	0.3757	0.1601	0.1158	0.0000	0.0000	0.0000
Haematology	5.7301	4.5135	5.9957	0.0000	0.0000	0.1902	0.0000	0.0000	0.0000
Infectious Diseases	9.3236	7.5686	7.6208	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Intermediate Care	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0366	0.8387
Maxillo-Facial	1.3817	1.3665	0.9226	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Neurology	4.2779	3.2917	3.3918	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Ophthalmology	2.4465	2.5135	2.4547	0.0000	0.0000	0.0365	0.0000	0.0000	0.0000
Pain	0.0338	0.0388	0.0281	0.0000	0.0070	0.0030	0.0000	0.0000	0.0000
Plastic Surgery	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Radiology	0.0020	0.0379	0.0388	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Radiotherapy and Oncology	0.0042	0.0025	0.0010	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Rehabilitation	28.3000	24.9361	25.5243	13.5650	13.8940	11.8506	33.5651	36.2707	46.3084
Respiratory	42.2210	39.5059	38.5387	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Restorative Dentistry	0.0003	0.0000	0.0003	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Rheumatology	0.0000	0.0000	0.0003	0.0000	0.1463	0.0000	0.0000	0.0000	0.0000
Trauma & Orthopaedic	59.6046	71.7496	62.5797	0.2491	0.3731	0.2435	0.0000	0.0000	0.0000
Urology	20.1980	23.1760	23.4361	0.0136	0.0981	0.0116	0.0000	0.0000	0.0000

Table D.10: The daily bed demands for each specialty for regions one, two and three within ABUHB for three individual years' worth of patient admissions, using the classification tree and the node average LOS.

Specialty	**Region 4**			**Region 5**			**Region 6**		
	2017-2018	**2018-2019**	**2019-2020**	**2017-2018**	**2018-2019**	**2019-2020**	**2017-2018**	**2018-2019**	**2019-2020**
Accident & Emergency	0.0000	0.0000	0.0000	7.6680	8.6804	9.9681	0.0000	0.0000	0.0000
Anaesthetics	0.0000	0.0000	0.0000	0.4638	0.4268	0.9722	0.0000	0.0000	0.0000
Cardiology	0.0000	0.0000	0.0000	16.1539	14.2806	11.1400	0.0000	0.0000	0.0011
Care of the Elderly	12.5281	6.5938	6.8223	52.3094	45.1508	41.2107	0.0000	0.0000	0.0000
Community Medicine	0.4234	0.2117	0.4222	14.1831	15.4549	9.9576	0.0000	0.0000	0.0000
Dermatology	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Diabetes and Endocrinology	0.0000	0.0000	0.0000	16.6698	17.8823	15.3511	0.0000	0.0000	0.0000
Ear Nose & Throat	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Gastroenterology	0.0000	0.0000	0.0000	18.7998	21.9374	19.9778	0.3047	0.0000	0.0022
General Medicine	0.0000	0.0000	0.0000	13.9936	14.0130	20.9445	0.0000	0.0000	0.0000
General Surgery	0.0000	0.0000	0.0000	20.4583	21.6905	21.4020	0.0011	0.0011	0.0011
GP Other	0.0000	0.0000	0.0000	14.1544	13.2848	14.3738	0.0000	0.0000	0.0000
Gynaecology	0.0000	0.0000	0.0000	0.9738	0.9246	0.6943	0.0000	0.0000	0.0011
Haematology	0.0000	0.0000	0.0000	1.8249	1.7582	2.0842	0.0011	0.0000	0.0000
Infectious Diseases	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Intermediate Care	0.0000	0.3333	0.3699	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Maxillo-Facial	0.0000	0.0000	0.0000	0.1061	0.0747	0.0969	0.0000	0.0000	0.0000
Neurology	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Ophthalmology	0.0000	0.0000	0.0000	0.3948	0.4613	0.4636	0.0000	0.9640	1.2960
Pain	0.0166	0.0110	0.0176	0.0519	0.0442	0.0396	0.0000	0.0000	0.0000
Plastic Surgery	0.0000	0.0000	0.0000	0.0521	0.0464	0.0462	0.0000	0.0000	0.0011
Radiology	0.0000	0.0000	0.0000	0.0091	0.0011	0.0225	0.0000	0.0000	0.0000
Radiotherapy and Oncology	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Rehabilitation	29.5478	33.4318	35.5989	10.7206	16.2437	23.0764	0.0000	0.0000	0.0000
Respiratory	0.0000	0.0000	0.0000	32.7445	29.8673	25.8748	0.0000	0.0000	0.0000
Restorative Dentistry	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Rheumatology	0.0000	0.0000	0.0000	0.0000	0.0215	0.0000	0.0000	0.0000	0.0000
Trauma & Orthopaedic	0.0000	0.0000	0.0000	36.3659	44.0053	41.3888	0.0000	0.0000	0.0000
Urology	0.0000	0.0000	0.0000	0.5179	0.4967	0.5494	0.0210	0.0169	0.0226

Table D.11: The daily bed demands for each specialty for regions four, five and six within ABUHB for three individual years' worth of patient admissions, using the classification tree and the node average LOS.

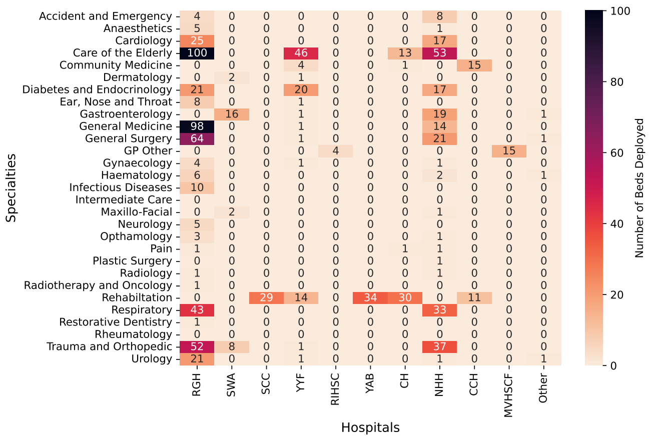


Figure D.25: Heatmap of bed locations for each specialty within each hospital for the deterministic model using the classification tree and average LOS for 2017-2018.

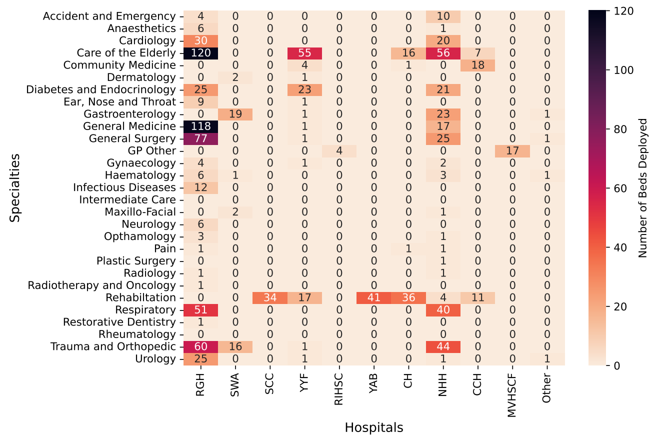


Figure D.26: Heatmap of bed locations for each specialty within each hospital for the two-stage stochastic model using the classification tree and average LOS for 2017-2018.

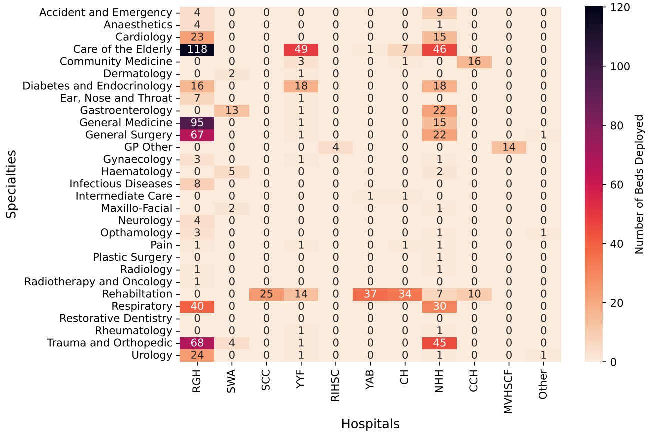


Figure D.27: Heatmap of bed locations for each specialty within each hospital for the deterministic model using the classification tree and average LOS for 2018-2019.

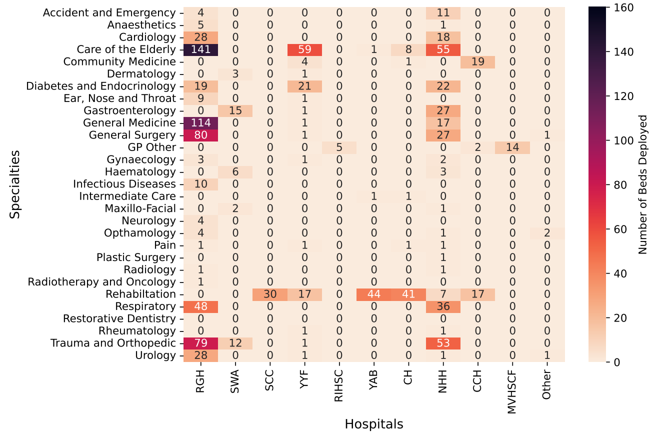


Figure D.28: Heatmap of bed locations for each specialty within each hospital for the two-stage stochastic model using the classification tree and average LOS for 2018-2019.

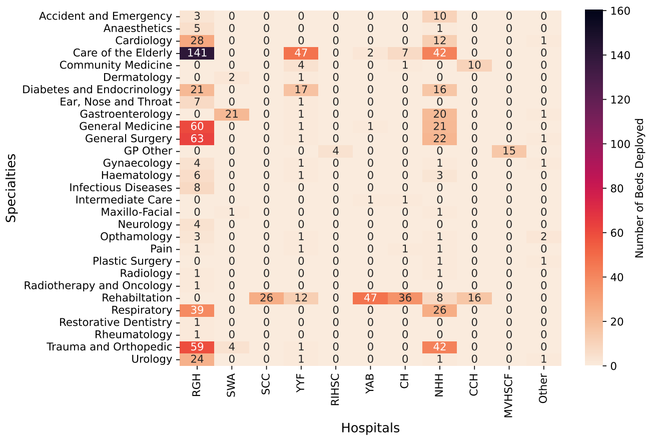


Figure D.29: Heatmap of bed locations for each specialty within each hospital for the deterministic model using the classification tree and average LOS for 2019-2020.

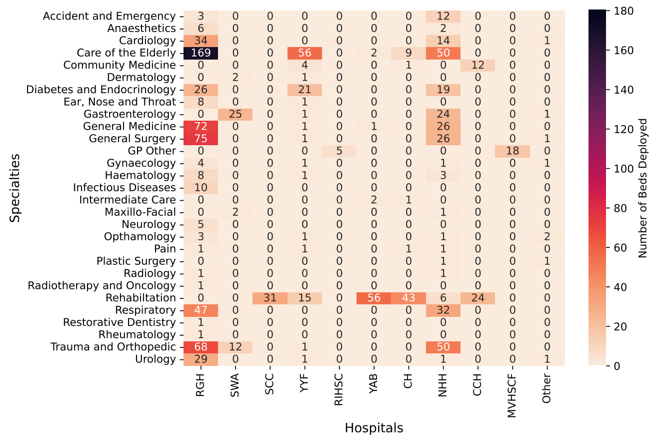


Figure D.30: Heatmap of bed locations for each specialty within each hospital for the two-stage stochastic model using the classification tree and average LOS for 2019-2020.

Specialty	**Region 1**			**Region 2**			**Region 3**		
	2017-2018	**2018-2019**	**2019-2020**	**2017-2018**	**2018-2019**	**2019-2020**	**2017-2018**	**2018-2019**	**2019-2020**
Accident & Emergency	3.2546	3.3498	2.2753	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Anaesthetics	4.5748	3.6941	4.2057	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Cardiology	24.8467	22.1565	27.8232	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Care of the Elderly	101.8083	114.9121	139.8770	46.1980	47.9158	45.7555	0.0000	0.1435	1.5595
Community Medicine	0.0000	0.0000	0.0000	3.1874	2.9051	3.2631	0.0000	0.0000	0.0000
Dermatology	1.5243	1.6713	1.7106	0.2269	0.2859	0.3767	0.0000	0.0000	0.0000
Diabetes and Endocrinology	21.1122	14.8840	20.8111	19.5744	17.0001	16.6421	0.0000	0.0000	0.0000
Ear Nose & Throat	7.8616	6.8260	6.1524	0.0430	0.0120	0.0054	0.0000	0.0000	0.0000
Gastroenterology	15.8905	11.7729	20.3049	0.6772	0.4819	0.8694	0.0000	0.0000	0.0000
General Medicine	103.9455	95.2600	53.5851	0.2227	0.4430	0.7331	0.0000	0.0000	0.0209
General Surgery	64.8113	65.3150	61.9319	0.3083	0.2920	0.4636	0.0000	0.0000	0.0000
GP Other	0.0000	0.0000	0.0000	3.1499	3.4789	3.5531	0.0000	0.0000	0.0000
Gynaecology	2.9168	2.5419	3.1850	0.3355	0.1718	0.1331	0.0000	0.0000	0.0000
Haematology	5.9642	4.4035	6.0321	0.0000	0.0000	0.1848	0.0000	0.0000	0.0000
Infectious Diseases	9.5621	7.4241	7.5776	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Intermediate Care	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0359	0.8339
Maxillo-Facial	1.4029	1.3015	0.9645	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Neurology	4.3873	3.2286	3.3727	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Ophthalmology	2.4090	2.2876	2.7137	0.0000	0.0000	0.0363	0.0000	0.0000	0.0000
Pain	0.0307	0.0305	0.0363	0.0000	0.0054	0.0038	0.0000	0.0000	0.0000
Plastic Surgery	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Radiology	0.0018	0.0369	0.0392	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Radiotherapy and Oncology	0.0060	0.0024	0.0013	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Rehabilitation	29.0237	24.4600	25.3794	13.9119	13.6287	11.7833	34.4235	35.5782	46.0456
Respiratory	42.8309	38.2732	37.8883	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Restorative Dentistry	0.0003	0.0000	0.0004	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Rheumatology	0.0000	0.0000	0.0004	0.0000	0.1435	0.0000	0.0000	0.0000	0.0000
Trauma & Orthopaedic	61.4394	70.5687	61.4178	0.2786	0.3522	0.2350	0.0000	0.0000	0.0000
Urology	20.3556	23.2087	23.0969	0.0141	0.0950	0.0083	0.0000	0.0000	0.0000

Table D.12: The daily bed demands for each specialty for regions one, two and three within ABUHB for three individual years' worth of patient admissions, using the classification tree and the yearly average LOS.

Specialty	**Region 4**			**Region 4**			**Region 6**		
	2017-2018	**2018-2019**	**2019-2020**	**2017-2018**	**2018-2019**	**2019-2020**	**2017-2018**	**2018-2019**	**2019-2020**
Accident & Emergency	0.0000	0.0000	0.0000	7.6680	8.6804	9.9681	0.0000	0.0000	0.0000
Anaesthetics	0.0000	0.0000	0.0000	0.4638	0.4268	0.9722	0.0000	0.0000	0.0000
Cardiology	0.0000	0.0000	0.0000	16.1539	14.2806	11.1400	0.0000	0.0000	0.0011
Care of the Elderly	12.5281	6.5938	6.8223	52.3094	45.1508	41.2107	0.0000	0.0000	0.0000
Community Medicine	0.4234	0.2117	0.4222	14.1831	15.4549	9.9576	0.0000	0.0000	0.0000
Dermatology	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Diabetes and Endocrinology	0.0000	0.0000	0.0000	16.6698	17.8823	15.3511	0.0000	0.0000	0.0000
Ear Nose & Throat	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Gastroenterology	0.0000	0.0000	0.0000	18.7998	21.9374	19.9778	0.3047	0.0000	0.0022
General Medicine	0.0000	0.0000	0.0000	13.9936	14.0130	20.9445	0.0000	0.0000	0.0000
General Surgery	0.0000	0.0000	0.0000	20.4583	21.6905	21.4020	0.0011	0.0011	0.0011
GP Other	0.0000	0.0000	0.0000	14.1544	13.2848	14.3738	0.0000	0.0000	0.0000
Gynaecology	0.0000	0.0000	0.0000	0.9738	0.9246	0.6943	0.0000	0.0000	0.0011
Haematology	0.0000	0.0000	0.0000	1.8249	1.7582	2.0842	0.0011	0.0000	0.0000
Infectious Diseases	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Intermediate Care	0.0000	0.3333	0.3699	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Maxillo-Facial	0.0000	0.0000	0.0000	0.1061	0.0747	0.0969	0.0000	0.0000	0.0000
Neurology	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Ophthalmology	0.0000	0.0000	0.0000	0.3948	0.4613	0.4636	0.0000	0.9640	1.2960
Pain	0.0166	0.0110	0.0176	0.0519	0.0442	0.0396	0.0000	0.0000	0.0000
Plastic Surgery	0.0000	0.0000	0.0000	0.0521	0.0464	0.0462	0.0000	0.0000	0.0011
Radiology	0.0000	0.0000	0.0000	0.0091	0.0011	0.0225	0.0000	0.0000	0.0000
Radiotherapy and Oncology	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Rehabilitation	29.5478	33.4318	35.5989	10.7206	16.2437	23.0764	0.0000	0.0000	0.0000
Respiratory	0.0000	0.0000	0.0000	32.7445	29.8673	25.8748	0.0000	0.0000	0.0000
Restorative Dentistry	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Rheumatology	0.0000	0.0000	0.0000	0.0000	0.0215	0.0000	0.0000	0.0000	0.0000
Trauma & Orthopaedic	0.0000	0.0000	0.0000	36.3659	44.0053	41.3888	0.0000	0.0000	0.0000
Urology	0.0000	0.0000	0.0000	0.5179	0.4967	0.5494	0.0210	0.0169	0.0226

Table D.13: The daily bed demands for each specialty for regions four, five and six within ABUHB for three individual years' worth of patient admissions, using the classification tree and the yearly average LOS.

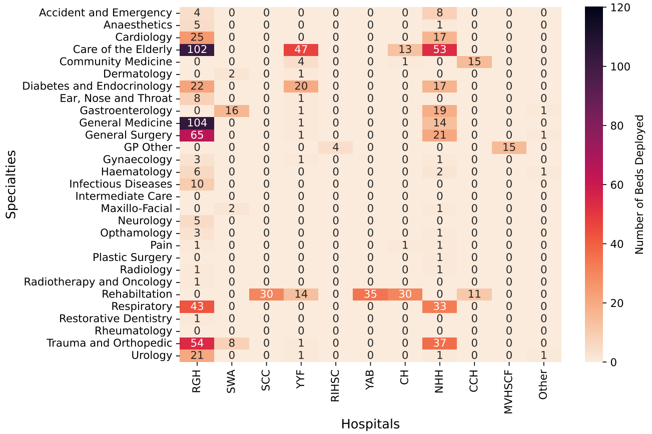


Figure D.31: Heatmap of bed locations for each specialty within each hospital for the deterministic model using the classification tree and average year LOS for 2017-2018.

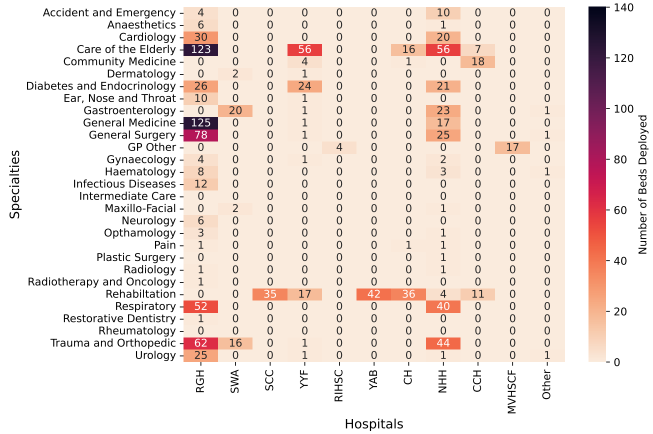


Figure D.32: Heatmap of bed locations for each specialty within each hospital for the two-stage stochastic model using the classification tree and average year LOS for 2017-2018.

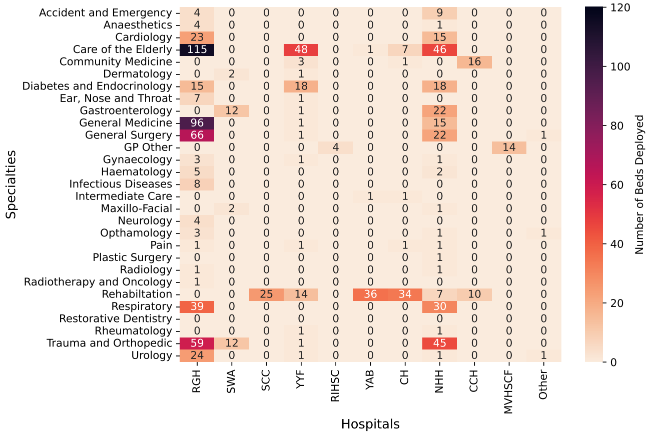


Figure D.33: Heatmap of bed locations for each specialty within each hospital for the deterministic model using the classification tree and average year LOS for 2018-2019.

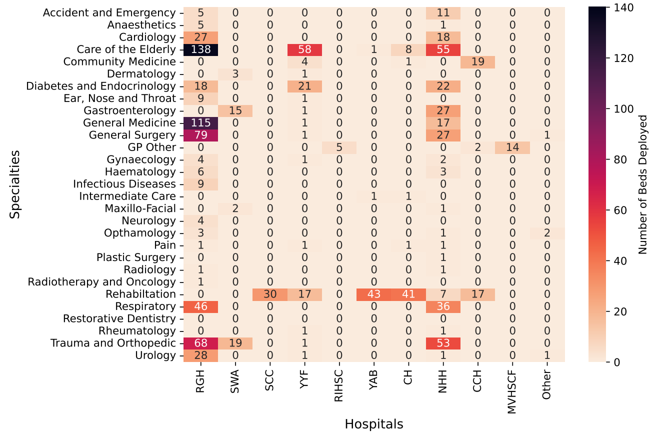


Figure D.34: Heatmap of bed locations for each specialty within each hospital for the two-stage stochastic model using the classification tree and average year LOS for 2018-2019.

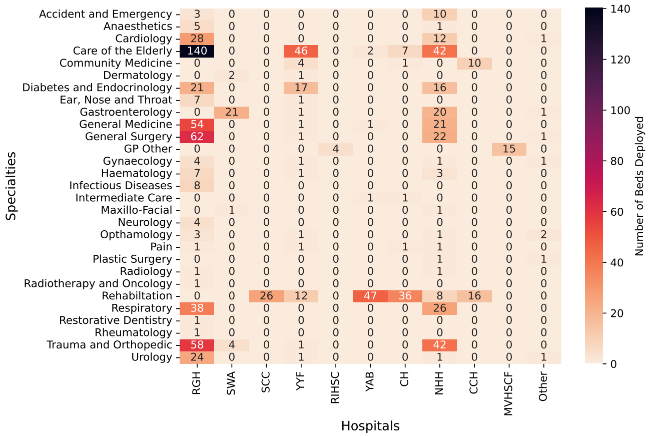


Figure D.35: Heatmap of bed locations for each specialty within each hospital for the deterministic model using the classification tree and average year LOS for 2019-2020.

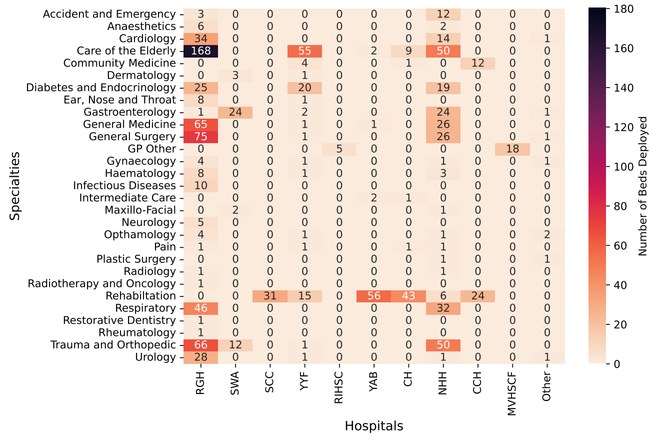


Figure D.36: Heatmap of bed locations for each specialty within each hospital for the two-stage stochastic model using the classification tree and average year LOS for 2019-2020.

D.4 Classification Trees - Specific LOS

Specialty	**Region 1**	**Region 2**	**Region 3**	**Region 4**	**Region 5**	**Region 6**
Accident & Emergency	2.1168	0.0000	0.0000	0.0000	9.1761	0.0000
Anaesthetics	4.5894	0.0000	0.0000	0.0000	0.7719	0.0000
Cardiology	15.5265	0.0000	0.0000	0.0000	9.7901	0.0000
Care of the Elderly	93.9772	57.7354	0.7573	8.7856	46.4398	0.0000
Community Medicine	0.0000	7.0073	0.0000	0.3139	13.0137	0.0000
Dermatology	2.2591	0.0000	0.0000	0.0000	0.0000	0.0000
Diabetes and Endocrinology	14.4489	21.2409	0.0000	0.0000	17.2290	0.0000
Ear Nose & Throat	3.1104	0.0009	0.0000	0.0000	0.0000	0.0000
Gastroenterology	12.3120	0.0985	0.0000	0.0000	19.7765	0.0009
General Medicine	84.4854	0.9818	0.0119	0.0000	14.1013	0.0000
General Surgery	45.5192	0.2318	0.0000	0.0000	21.3011	0.0000
GP Other	0.0000	9.8723	0.0000	0.0000	15.3102	0.0000
Gynaecology	2.0046	0.0721	0.0000	0.0000	1.0766	0.0000
Haematology	2.7792	0.0265	0.0000	0.0000	1.1332	0.0000
Infectious Diseases	7.1195	0.0000	0.0000	0.0000	0.0000	0.0000
Intermediate Care	0.0000	0.0000	0.3449	0.3239	0.0000	0.0000
Maxillo-Facial	1.0547	0.0000	0.0000	0.0000	0.0027	0.0000
Neurology	1.5620	0.0000	0.0000	0.0000	0.0000	0.0000
Ophthalmology	1.3704	0.0027	0.0000	0.0000	0.0009	0.0036
Pain	0.0018	0.0000	0.0000	0.0018	0.0000	0.0000
Plastic Surgery	0.0000	0.0000	0.0000	0.0000	0.0137	0.0000
Radiology	0.0100	0.0000	0.0000	0.0000	0.0000	0.0000
Radiotherapy and Oncology	0.2245	0.0000	0.0000	0.0000	0.0000	0.0000
Rehabilitation	63.0173	33.0922	69.4863	32.1077	24.5538	0.0000
Respiratory	29.4717	0.0000	0.0000	0.0000	27.7290	0.0000
Restorative Dentistry	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Rheumatology	0.0000	0.0000	0.0000	0.0000	0.0128	0.0000
Trauma & Orthopaedic	59.1807	0.2956	0.0000	0.0000	40.6989	0.0000
Urology	11.3257	0.0036	0.0000	0.0000	0.0255	0.0401

Table D.14: The daily bed demands for each specialty grouped by regions within ABUHB for three years' worth of patient admissions, using the classification tree and specific LOS.

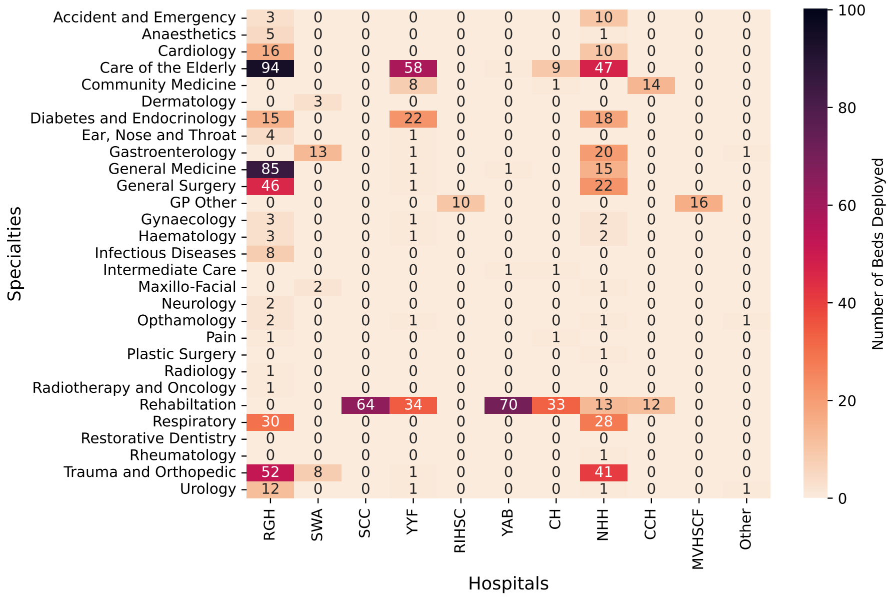


Figure D.37: Heatmap of bed locations for each specialty within each hospital for the deterministic model using the classification tree and specific LOS over three years' worth of data.

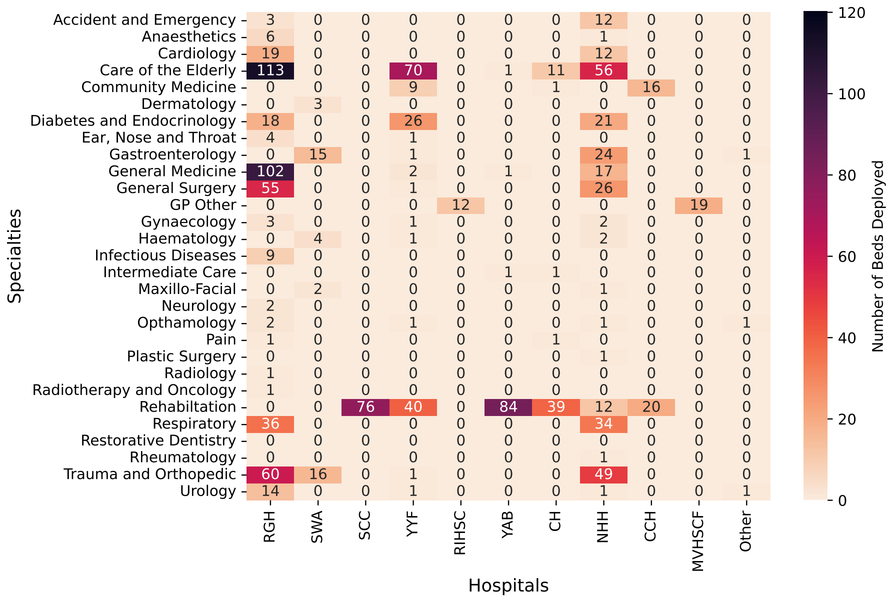


Figure D.38: Heatmap of bed locations for each specialty within each hospital for the two-stage stochastic model using the classification tree and specific LOS over three years' worth of data.

Specialty	**Region 1**			**Region 2**			**Region 3**		
	2017-2018	**2018-2019**	**2019-2020**	**2017-2018**	**2018-2019**	**2019-2020**	**2017-2018**	**2018-2019**	**2019-2020**
Accident & Emergency	2.2932	2.2000	1.8579	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Anaesthetics	4.4932	4.4712	4.8033	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Cardiology	16.7918	14.0630	15.7240	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Care of the Elderly	76.4575	88.0658	117.3443	55.8247	58.4877	58.8907	0.0000	0.4000	1.8689
Community Medicine	0.0000	0.0000	0.0000	7.1671	6.1014	7.7514	0.0000	0.0000	0.0000
Dermatology	2.6877	1.9918	2.0984	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Diabetes and Endocrinology	16.5863	11.4356	15.3224	24.7041	19.8055	19.2186	0.0000	0.0000	0.0000
Ear Nose & Throat	2.9616	3.6192	2.7514	0.0027	0.0000	0.0000	0.0000	0.0000	0.0000
Gastroenterology	11.4438	9.8219	15.6612	0.1260	0.0219	0.1475	0.0000	0.0000	0.0000
General Medicine	105.6712	96.0603	51.8142	0.2877	1.1479	1.5082	0.0000	0.0000	0.0355
General Surgery	47.9068	45.3479	43.3087	0.2000	0.2301	0.2650	0.0000	0.0000	0.0000
GP Other	0.0000	0.0000	0.0000	9.6795	10.1151	9.8224	0.0000	0.0000	0.0000
Gynaecology	2.4301	1.7233	1.8607	0.0795	0.0740	0.0628	0.0000	0.0000	0.0000
Haematology	3.1096	2.4384	2.7896	0.0000	0.0000	0.0792	0.0000	0.0000	0.0000
Infectious Diseases	8.0767	6.3918	6.8907	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Intermediate Care	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0082	1.0246
Maxillo-Facial	1.1068	0.9178	1.1393	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Neurology	1.4301	1.5534	1.7022	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Ophthalmology	1.4712	1.2685	1.3716	0.0000	0.0000	0.0082	0.0000	0.0000	0.0000
Pain	0.0000	0.0000	0.0055	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Plastic Surgery	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Radiology	0.0055	0.0164	0.0082	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Radiotherapy and Oncology	0.0000	0.0000	0.6721	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Rehabilitation	64.7671	63.1151	61.1749	34.9123	34.8384	29.5355	69.9671	65.3863	73.0956
Respiratory	29.8658	30.4137	28.1393	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Restorative Dentistry	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Rheumatology	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Trauma & Orthopaedic	56.0082	60.6164	60.9126	0.3808	0.2712	0.2350	0.0000	0.0000	0.0000
Urology	10.3671	12.5726	11.0383	0.0027	0.0055	0.0027	0.0000	0.0000	0.0000

Table D.15: The daily bed demands for each specialty for regions one, two and three within ABUHB for three individual years' worth of patient admissions, using the classification tree and the yearly specific LOS.

Specialty	**Region 4**			**Region 5**			**Region 6**		
	2017-2018	**2018-2019**	**2019-2020**	**2017-2018**	**2018-2019**	**2019-2020**	**2017-2018**	**2018-2019**	**2019-2020**
Accident & Emergency	0.0000	0.0000	0.0000	7.6680	8.6804	9.9681	0.0000	0.0000	0.0000
Anaesthetics	0.0000	0.0000	0.0000	0.4638	0.4268	0.9722	0.0000	0.0000	0.0000
Cardiology	0.0000	0.0000	0.0000	16.1539	14.2806	11.1400	0.0000	0.0000	0.0011
Care of the Elderly	12.5281	6.5938	6.8223	52.3094	45.1508	41.2107	0.0000	0.0000	0.0000
Community Medicine	0.4234	0.2117	0.4222	14.1831	15.4549	9.9576	0.0000	0.0000	0.0000
Dermatology	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Diabetes and Endocrinology	0.0000	0.0000	0.0000	16.6698	17.8823	15.3511	0.0000	0.0000	0.0000
Ear Nose & Throat	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Gastroenterology	0.0000	0.0000	0.0000	18.7998	21.9374	19.9778	0.3047	0.0000	0.0022
General Medicine	0.0000	0.0000	0.0000	13.9936	14.0130	20.9445	0.0000	0.0000	0.0000
General Surgery	0.0000	0.0000	0.0000	20.4583	21.6905	21.4020	0.0011	0.0011	0.0011
GP Other	0.0000	0.0000	0.0000	14.1544	13.2848	14.3738	0.0000	0.0000	0.0000
Gynaecology	0.0000	0.0000	0.0000	0.9738	0.9246	0.6943	0.0000	0.0000	0.0011
Haematology	0.0000	0.0000	0.0000	1.8249	1.7582	2.0842	0.0011	0.0000	0.0000
Infectious Diseases	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Intermediate Care	0.0000	0.3333	0.3699	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Maxillo-Facial	0.0000	0.0000	0.0000	0.1061	0.0747	0.0969	0.0000	0.0000	0.0000
Neurology	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Ophthalmology	0.0000	0.0000	0.0000	0.3948	0.4613	0.4636	0.0000	0.9640	1.2960
Pain	0.0166	0.0110	0.0176	0.0519	0.0442	0.0396	0.0000	0.0000	0.0000
Plastic Surgery	0.0000	0.0000	0.0000	0.0521	0.0464	0.0462	0.0000	0.0000	0.0011
Radiology	0.0000	0.0000	0.0000	0.0091	0.0011	0.0225	0.0000	0.0000	0.0000
Radiotherapy and Oncology	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Rehabilitation	29.5478	33.4318	35.5989	10.7206	16.2437	23.0764	0.0000	0.0000	0.0000
Respiratory	0.0000	0.0000	0.0000	32.7445	29.8673	25.8748	0.0000	0.0000	0.0000
Restorative Dentistry	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Rheumatology	0.0000	0.0000	0.0000	0.0000	0.0215	0.0000	0.0000	0.0000	0.0000
Trauma & Orthopaedic	0.0000	0.0000	0.0000	36.3659	44.0053	41.3888	0.0000	0.0000	0.0000
Urology	0.0000	0.0000	0.0000	0.5179	0.4967	0.5494	0.0210	0.0169	0.0226

Table D.16: The daily bed demands for each specialty for regions four, five and six within ABUHB for three individual years' worth of patient admissions, using the classification tree and the yearly specific LOS.

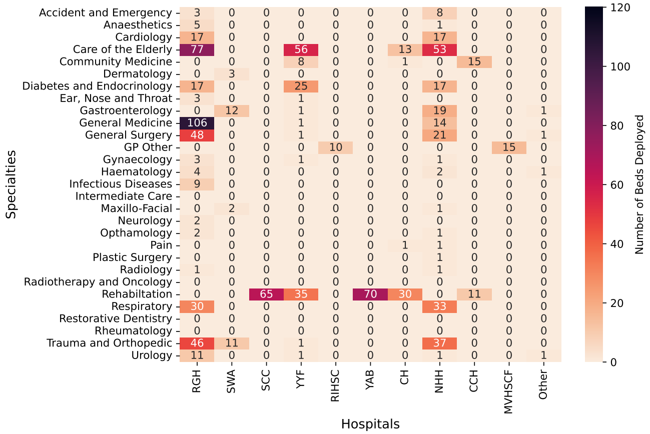


Figure D.39: Heatmap of bed locations for each specialty within each hospital for the deterministic model using the classification tree and specific LOS for 2017-2018.

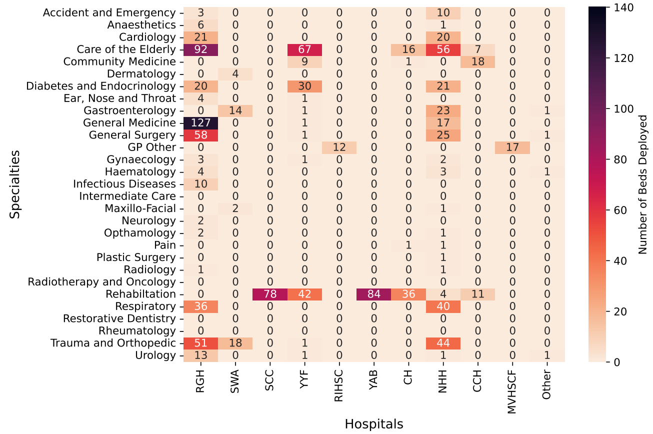


Figure D.40: Heatmap of bed locations for each specialty within each hospital for the two-stage stochastic model using the classification tree and specific LOS for 2017-2018.

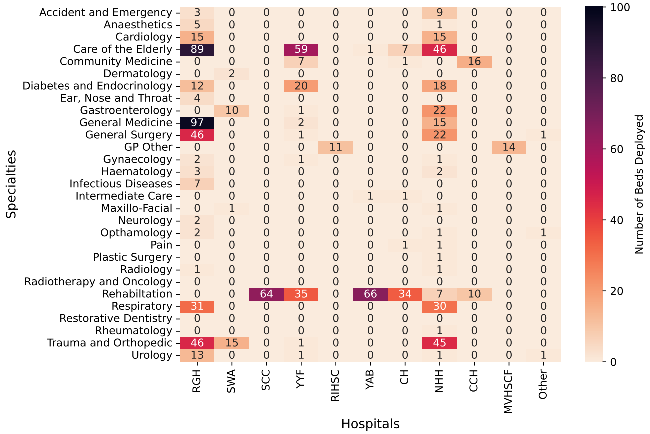


Figure D.41: Heatmap of bed locations for each specialty within each hospital for the deterministic model using the classification tree and specific LOS for 2018-2019.

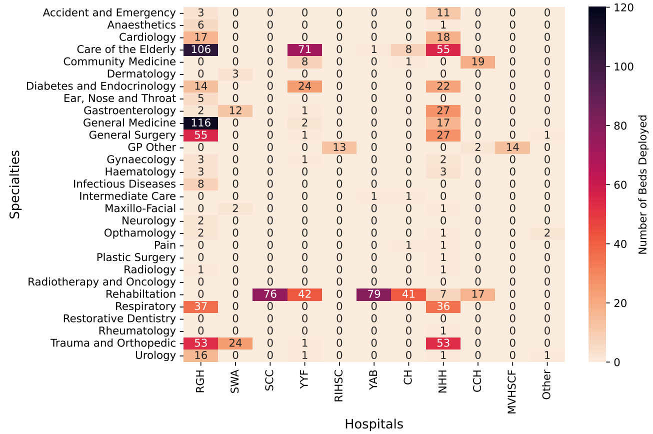


Figure D.42: Heatmap of bed locations for each specialty within each hospital for the two-stage stochastic model using the classification tree and specific LOS for 2018-2019.

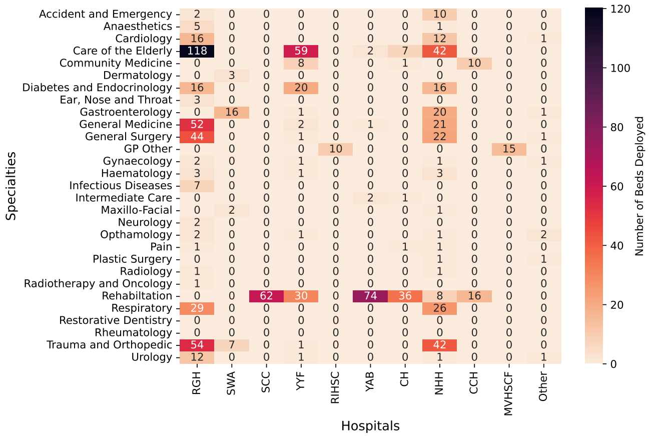


Figure D.43: Heatmap of bed locations for each specialty within each hospital for the deterministic model using the classification tree and specific LOS for 2019-2020.

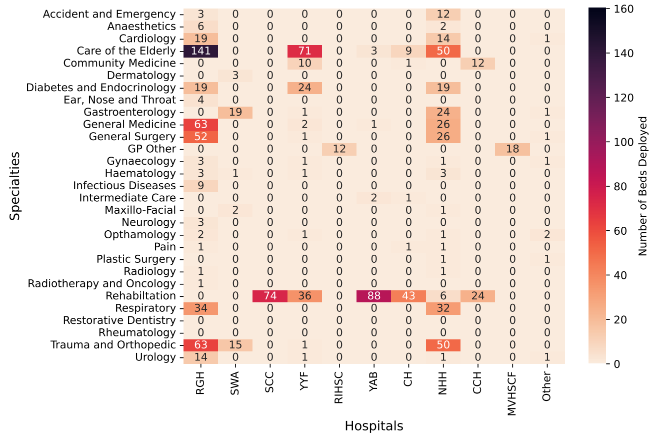


Figure D.44: Heatmap of bed locations for each specialty within each hospital for the two-stage stochastic model using the classification tree and specific LOS for 2019-2020.

D.5 Scenario Heatmaps

This section contains the heatmaps produced from performing various scenario analysis within Chapter 6.3.

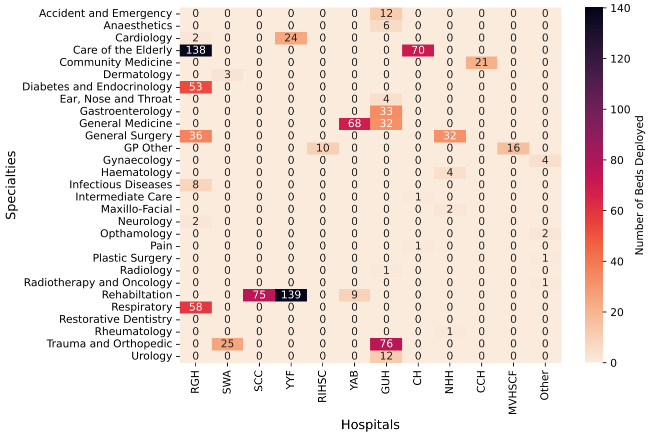


Figure D.45: Heatmap of bed locations for each specialty within each hospital for the deterministic model for Scenario 1 where GUH is added.

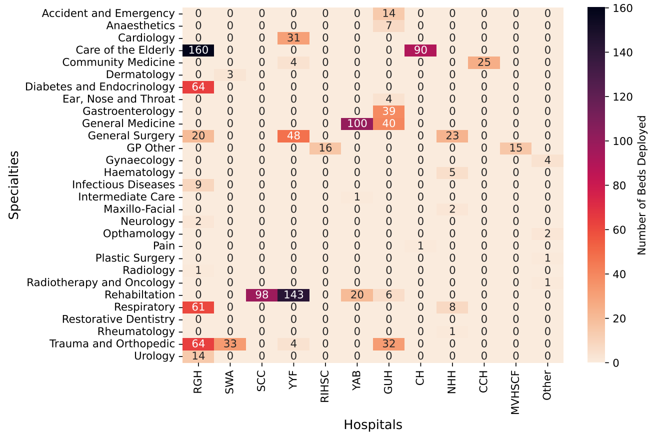


Figure D.46: Heatmap of bed locations for each specialty within each hospital for the two-stage stochastic model for Scenario 1 where GUH is added.

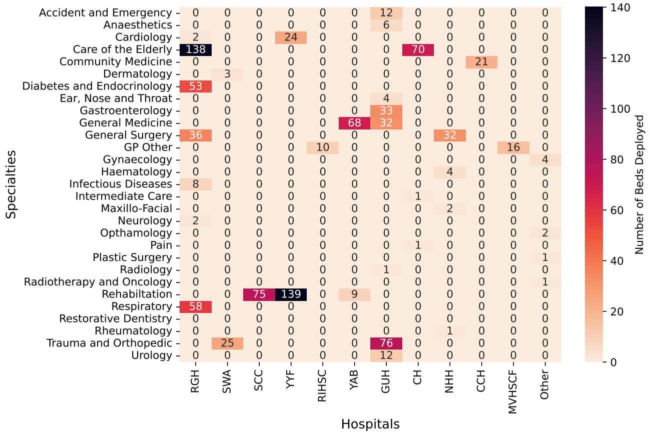


Figure D.47: Heatmap of bed locations for each specialty within each hospital for the deterministic model for Scenario 2 where the M-penalty method is added.

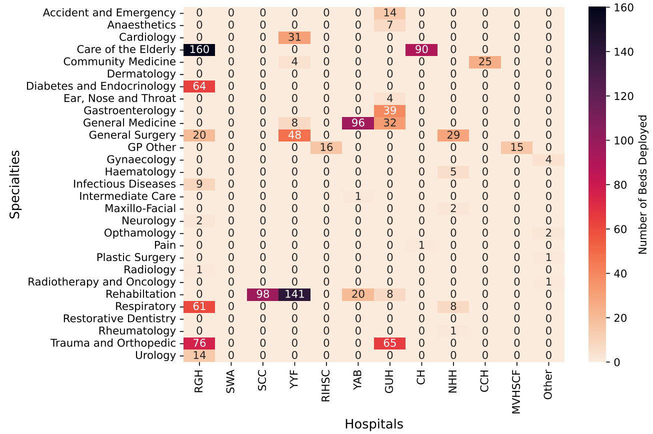


Figure D.48: Heatmap of bed locations for each specialty within each hospital for the two-stage stochastic model for Scenario 2 where the M-penalty method is added.

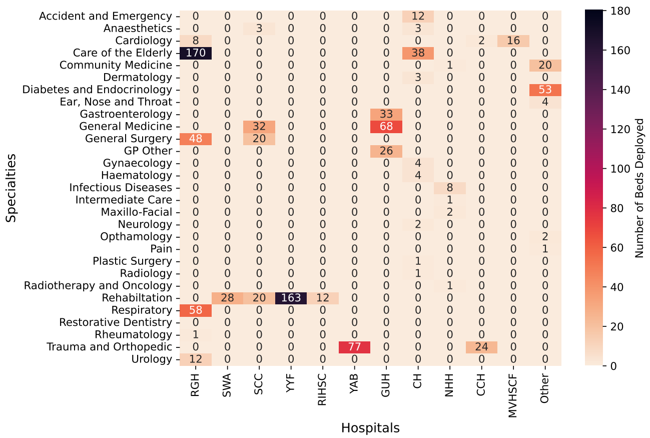


Figure D.49: Heatmap of bed locations for each specialty within each hospital for the deterministic model for Scenario 3 where the hospital setup is re-evaluated.

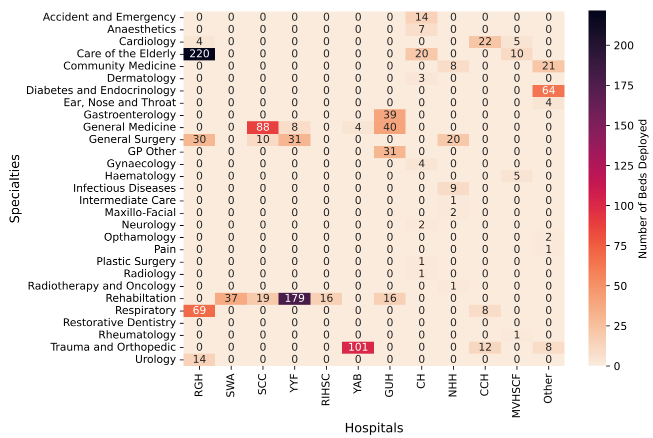


Figure D.50: Heatmap of bed locations for each specialty within each hospital for the two-stage stochastic model for Scenario 3 where the hospital setup is re-evaluated.

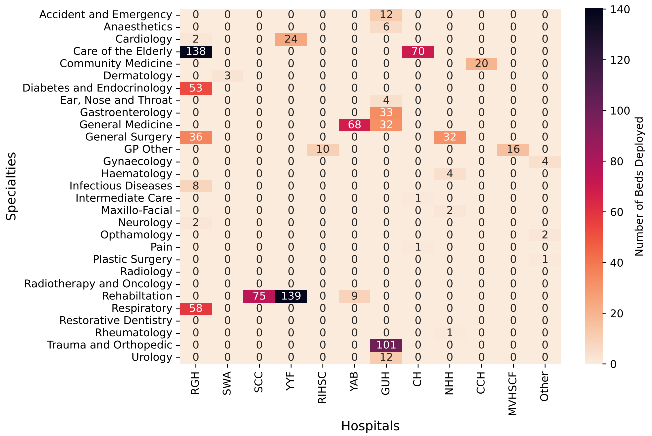


Figure D.51: Heatmap of bed locations for each specialty within each hospital for the deterministic model for Scenario 4 where the nursing capacity is reduced.

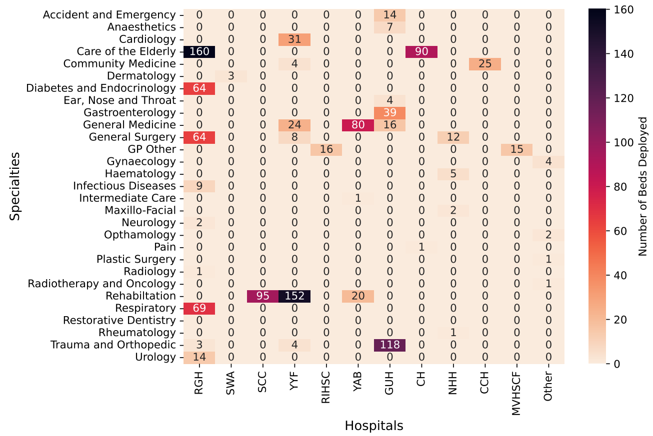


Figure D.52: Heatmap of bed locations for each specialty within each hospital for the two-stage stochastic model for Scenario 4 where the nursing capacity is reduced.

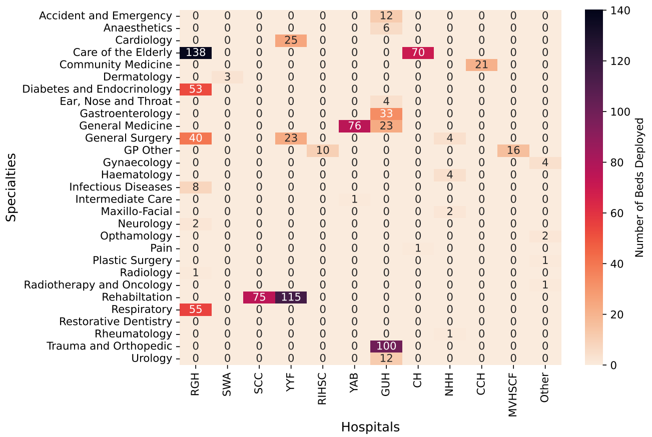


Figure D.53: Heatmap of bed locations for each specialty within each hospital for the deterministic model for Scenario 5 with the introduction of virtual wards.

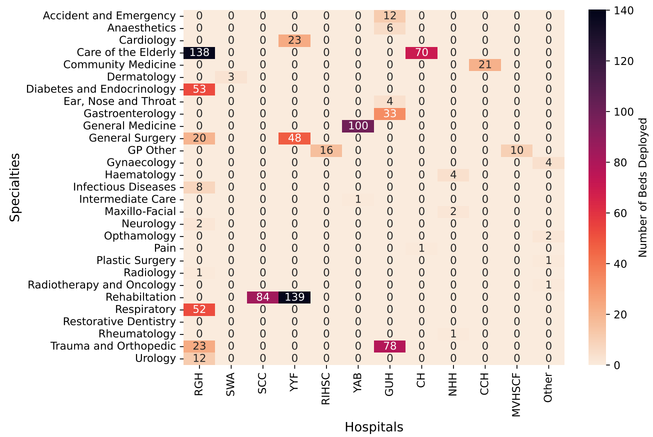


Figure D.54: Heatmap of bed locations for each specialty within each hospital for the two-stage stochastic model for Scenario 5 with the introduction of virtual wards.

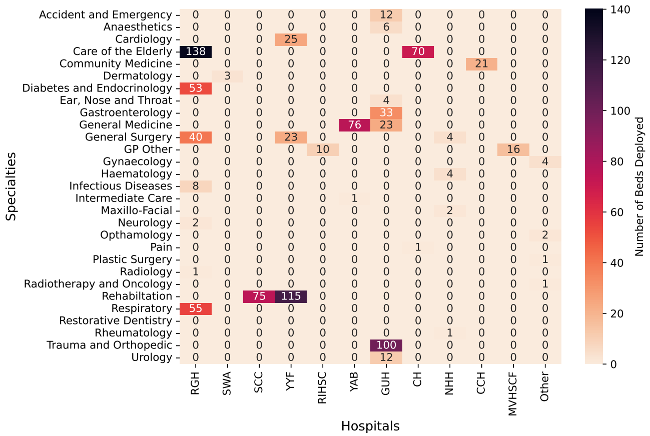


Figure D.55: Heatmap of bed locations for each specialty within each hospital for the deterministic model for Scenario 6 with the sudden increase in demand.

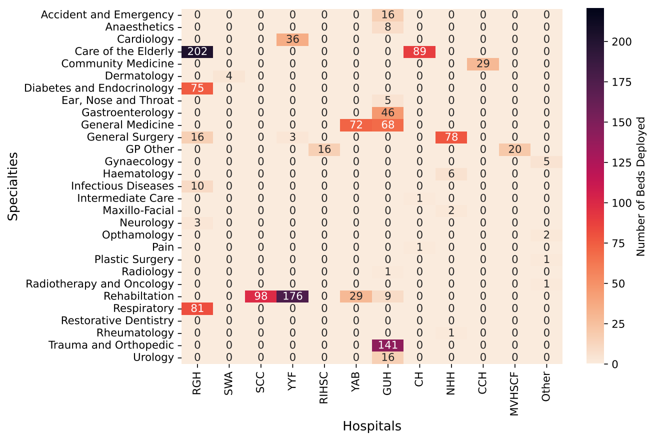


Figure D.56: Heatmap of bed locations for each specialty within each hospital for the two-stage model for Scenario 6 with the sudden increase in demand.

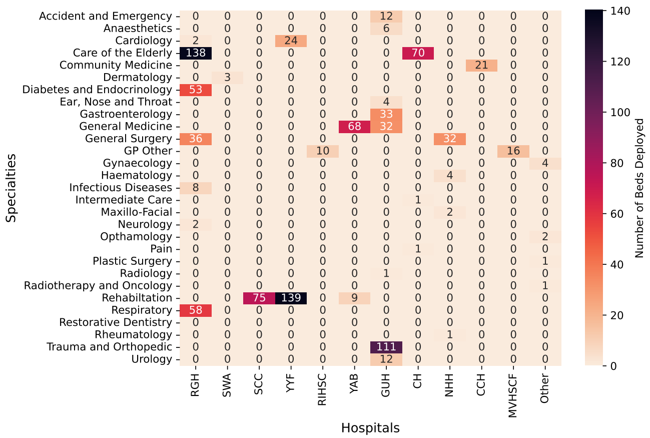


Figure D.57: Heatmap of bed locations for each specialty within each hospital for the deterministic model for Scenario 7 with an overall increase in T&O services of 10%.

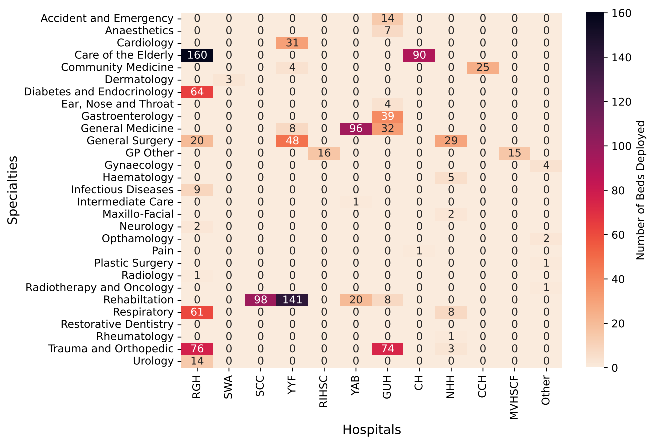


Figure D.58: Heatmap of bed locations for each specialty within each hospital for the two-stage stochastic model for Scenario 7 with an overall increase in T&O services of 10%.

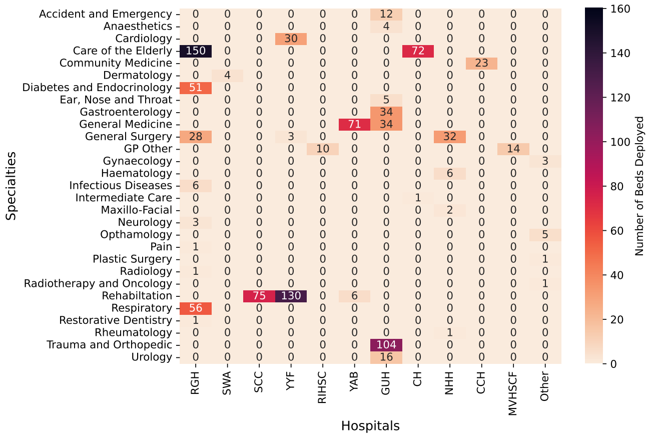


Figure D.59: Heatmap of bed locations for each specialty within each hospital for the deterministic model for Scenario 7 with targetting T&O specific nodes in the regression tree with a 10% increase.

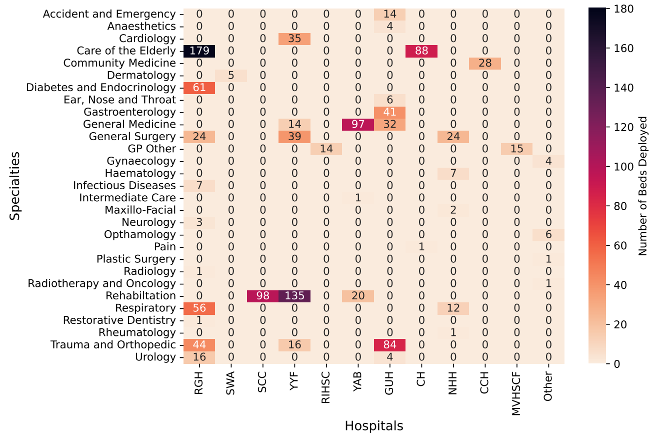


Figure D.60: Heatmap of bed locations for each specialty within each hospital for the two-stage stochastic model for Scenario 7 with targetting T&O specific nodes in the regression tree with a 10% increase.

Bibliography

[1] E. Williams, D. Gartner, and P. Harper. A survey of OR/MS models on care planning for frail and elderly patients. *Operations Research for Health Care*, 31:100325, 2021.

[2] D. Gartner, E. M. Williams, and P. R. Harper. Prescriptive Healthcare Analytics: A Tutorial on Discrete Optimization and Simulation. In *2022 IEEE 10th International Conference on Healthcare Informatics (ICHI)*, 2022.

[3] E. Williams, D. Gartner, and P. Harper. Linking Predictive and Prescriptive Analytics of Elderly and Frail Patient Hospital Services. In *2022 IEEE 10th International Conference on Healthcare Informatics (ICHI)*, 2022.

[4] StatsWales. National level population estimates by year, age and UK country. Available online at: "`https://statswales.gov.wales/Catalogue/Population-and-Migration/Population/Estimates/nationallevelpopulationestimates-by-year-age-ukcountry`". Last Accessed 19 July 2023, 2022.

[5] StatsWales. Population projections by year and age. Available online at: "`https://statswales.gov.wales/Catalogue/Population-and-Migration/Population/Projections/National/2020-based/2020basedpopulationprojections-by-year-age`". Last Accessed 19 July 2023, 2022.

[6] Aneurin Bevan University Health Board. Aneurin Bevan University Health Board - ANNUAL PLAN 2021/22. Available online at: "`https://abuhb.nhs.wales/files/key-documents/annual-plan-2021-2022/2021-22-health-board-annual-plan-pdf/`". Last Accessed 22 March 2023.

[7] N. C. Proudlove, K. Gordon, and R. Boaden. Can good bed management solve the overcrowding in accident and emergency departments? *Emergency Medicine Journal*, 20(2):149–155, 2003.

[8] Nuffield Trust. Hospitals at capacity: understanding delays in patient discharge. Available online at: "`https://www.nuffieldtrust.org.uk/news-item/hospitals-at-capacity-understanding-delays-in-patient-discharge`". Last Accessed 22 March 2023, 2022.

[9] Nuffield Trust. Why are delayed discharges happening? (April 2022). Available online at: "`https://bolt.nuffieldtrust.org.uk/public/chart/why-are-delayed-discharges-happening`". Last Accessed 19 July 2023, 2022.

[10] World Health Organisation. International Statistical Classification of Diseases and Related Health Problems 10th Revision. Available online at: "https://icd.who.int/browse10/2019/en". Last Accessed 19 July 2023, 2019.

[11] Public Health Scotland. Reports for Financial Year 2019 to 2020. Available online at: "https://beta.isdscotland.org/topics/finance/file-listings-fy-2019-to-2020/". Last Accessed 22 March 2023, 2021.

[12] NHS Employers. Hourly pay scales 2021/2022. Available online at "https://www.nhsemployers.org/articles/hourly-pay-scales-202122". Last Accessed 22 March 2023, 2021.

[13] Oxford Health NHS Foundation Trust. Pay rates. Available online at: "https://www.oxfordhealth.nhs.uk/careers/staff-bank/rates/". Last Accessed 22 March 2023, 2023.

[14] Aneurin Bevan University Health Board. Caring For You and Your Future. Available online at: "https://abuhb.nhs.wales/clinical-futures/caring-for-you-and-your-future/". Last Accessed 22 March 2023.

[15] Knowledge Economy Skills Scholarships . KESS2 East. Available online at: "https://kess2.ac.uk/". Last Accessed 22 March 2023, 2023.

[16] Aneurin Bevan University Health Board. What is Clinical Futures and what changes will it bring? Available online at: "https://abuhb.nhs.wales/clinical-futures/frequently-asked-questions-faqs/what-is-clinical-futures-and-what-changes-will-it-bring/". Last Accessed 22 March 2023.

[17] E. Welch. *The NHS at 70 A Living History*. Barnsley: Pen & Sword Books Limited, 2018.

[18] Office for National Statistics. Healthcare expenditure, UK Health Accounts provisional estimates: 2021. Available online at: "https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/healthcaresystem/bulletins/healthcareexpenditureukhealthaccountsprovisionalestimates/2021#government-healthcare-expenditure". Last Accessed 22 March 2023, 2022.

[19] United Nations. Ageing. Available online at: "https://www.un.org/en/sections/issues-depth/ageing/". Last Accessed 22 March 2023, 2019.

[20] D. E. Crews and S. Zavotka. Aging, Disability, and Frailty: Implications for Universal Design. *Journal of Physiological Anthropology*, 25(1):113–118, 2006.

[21] GOV.UK. Modern standards and service models: Older people. Available online at: "https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/198033/National_Service_Framework_for_Older_People.pdf". Last Accessed 22 March 2023, 2001.

[22] NHS England. Improving care for older people. Available online at : "https://www.england.nhs.uk/ourwork/clinical-policy/older-people/improving-care-for-older-people/". Last Accessed 22 March 2023.

[23] H. Orimo, H. Ito, T. Suzuki, A. Araki, T. Hosoi, and M. Sawabe. Reviewing the definition of "elderly". *Geriatrics and Gerontology International*, 6(3):149–158, 2006.

[24] Office for National Statistics. Voices of our ageing population: Living longer lives. Available online at: "https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/ageing/articles/voicesofourageingpopulation/livinglongerlives#:~:text=Over%2011%20million%20people%20%E2%80%93%2018.6,the%20previous%20census%20in%202011". Last Accessed 22 March 2023, 2022.

[25] Welsh Government. National population projections (interim data): 2020-based. Available online at: "https://gov.wales/national-population-projections-interim-data-2020-based#:~:text=Main%20points,3.17%20million%20to%203.30%20million.". Last Accessed 22 March 2023, 2022.

[26] E. Topinková. Aging, disability and frailty. *Annals of Nutrition and Metabolism*, 52(Suppl. 1):6–11, 2008.

[27] E. Dent, P. Kowal, and E. O. Hoogendijk. Frailty measurement in research and clinical practice: A review. *European Journal of Internal Medicine*, 31:3–10, 2016.

[28] K. Rockwood. A global clinical measure of fitness and frailty in elderly people. *Canadian Medical Association Journal*, 173(5):489–495, 2005.

[29] NHS England. Electronic Frailty Index. Available online at "https://www.england.nhs.uk/ourwork/clinical-policy/older-people/frailty/efi/". Last Accessed 22 March 2023.

[30] Age UK. Understanding frailty. Available online at: "https://www.ageuk.org.uk/our-impact/policy-research/frailty-in-older-people/understanding-frailty/". Last Accessed 22 March 2023, 2020.

[31] T. Gilbert, J. Neuburger, J Kraindler, E. Keeble, P. Smith, C. Ariti, S. Arora, A. Street, S. Parker, H. C. Roberts, M. Bardsley, and S. Conroy. Development and validation of a Hospital Frailty Risk Score focusing on older people in acute care settings using electronic hospital records: an observational study. *The Lancet*, 391(10132):1775–1782, 2018.

[32] J. Soong, A. J. Poots, S. Scott, K. Donald, T. Woodcock, D. Lovett, and D. Bell. Quantifying the prevalence of frailty in English hospitals. *BMJ open*, 5(10):e008456, October 2015.

[33] NHS Wales. ICD-10. Available online at: "https://dhcw.nhs.wales/information-services/information-standards/clinical-classifications-and-terminology-standards/about-clinical-classifications-and-coding/icd-10/". Last Accessed 22 March 2023.

[34] Aneurin Bevan University Health Board. The Health Board. Available online at: "https://abuhb.nhs.wales/about-us/the-health-board/". Last Accessed 22 March 2023.

[35] Aneurin Bevan University Health Board. Aneurin Bevan Continuous Improvement (ABCi). Available online at: "https://abuhb.nhs.wales/about-us/abci/". Last Accessed 22 March 2023.

[36] Welsh Government. The Well-being of Future Generations. Available online at: "https://gov.wales/well-being-of-future-generations-wales". Last Accessed 22 March 2023.

[37] A. Steventon, S. Deeny, R. Friebel, T. Gardner, and R. Thorlby. Briefing: Emergency hospital admissions in England: which may be avoidable and how? Available online at: "https://www.health.org.uk/sites/default/files/Briefing_Emergency%2520admissions_web_final.pdf". Last Accessed 22 March 2023, 2018.

[38] NHS England. Principle 5: Encourage a supported 'Home First' approach. Available online at: "https://www.england.nhs.uk/urgent-emergency-care/reducing-length-of-stay/reducing-long-term-stays/home-first/". Last Accessed 22 March 2023.

[39] National Health Service (Wales) Act 2006 (c.42). [Online]. The National Assembly of Wales. Available online at "https://www.legislation.gov.uk/anaw/2016/5/section/1/enacted". Last Accessed 22 March 2023.

[40] National Institute for Health and Care Excellence. Safe staffing for nursing in adult inpatient wards in acute hospitals. Available online at: "https://www.nice.org.uk/guidance/sg1/chapter/Recommendations#assessing-if-nursing-staff-available-on-the-day-meet-patients-nursing-needs". Last Accessed 22 March 2023, 2014.

[41] P. R. Harper and A. K. Shahani. Modelling for the planning and management of bed capacities in hospitals. *Journal of the Operational Research Society*, 53(1):11–18, 2002.

[42] E. Williams, T. Szakmany, I. Spernaes, B. Muthuswamy, and P. Holborn. Discrete-Event Simulation Modeling of Critical Care Flow: New Hospital, Old Challenges. *Critical Care Explorations*, 2(9):e0174, 2020.

[43] J. Webster and R. T. Watson. Analyzing the Past to Prepare for the Future: Writing a Literature Review. *MIS Quarterly*, 26(2):13–24, 2002.

[44] P. J. H. Hulshof, N. Kortbeek, R. J. Boucherie, E. W. Hans, and P. J. M. Bakker. Taxonomic classification of planning decision in health care: a structured review of the state of the art in OR/MS. *Health Systems*, 1:129–175, 2012.

[45] R. Palmer, N. J. Fulop, and M. Utley. A systematic literature review of operational research methods for modelling patient flow and outcomes within community healthcare and other settings. *Health Systems*, 7(1):29–50, 2018.

[46] K. de Luc and D. Kitchiner and A. Layton and E. Morris and Y. Murray and S. Overill. *Developing Care Pathways: the handbook*. Radcliffe Medical Press Ltd, Abingdon, 2001.

[47] D. Kitchiner, C. Davidson, and P. Bundred. Integrated care pathways: effective tools for continuous evaluation of clinical practice. *Journal of Evaluation in Clinical Practice*, 2(1):65–69, 1996.

[48] G. Berntsen, F. Strisland, K. Malm-Nicolaisen, B. Smaradottir, R. Fensli, and M. Røhne. The evidence base for an ideal care pathway for frail multimorbid elderly: Combined scoping and systematic intervention review. *Journal of Medical Internet Research*, 21(4), 2019.

[49] S. Freeman, K. Bishop, L. Spirgiene, E. Koopmans, F. C. Botelho, T. Fyfe, B. Xiong, S. Patchett, and M. MacLeod. Factors affecting residents transition from long term care facilities to the community: a scoping review. *BMC Health Services Research*, 17(1):689, 2017.

[50] J. E. Gaugler, S. Duval, K. A. Anderson, and R. L. Kane. Predicting nursing home admission in the U.S: A meta-analysis. *BMC Geriatrics*, 7, 2007.

[51] L. Kerpershoek, M. de Vugt, C. Wolfs, M. Orrell, B. Woods, H. Jelley, G. Meyer, A. Bieber, A. Stephan, G. Selbæk, M. Michelet, A. Wimo, R. Handels, K. Irving, L. Hopper, M. Gonçalves-Pereira, C. Balsinha, O. Zanetti, D. Portolani, F. Verhey, and the Actifcare Consortium. Is there equity in initial access to formal dementia care in Europe? The Andersen Model applied to the Actifcare cohort. *International Journal of Geriatric Psychiatry*, 35(1):45–52, 2020.

[52] S. Davari and T. Van Woensel. The elderly centre location problem. *Journal of the Operational Research Society*, 72(6):1207–1220, 2020.

[53] P. Intrevado, V. Verter, and L. Tremblay. Patient-centric design of long-term care networks. *Health Care Management Science*, 22(2):376–390, 2019.

[54] M. P. Johnson, W. L. Gorr, and S. Roehrig. Location of service facilities for the elderly. *Annals of Operations Research*, 136(1):329–349, 2005.

[55] A. Lim, Z. Zhang, and H. Qin. Pickup and Delivery Service with Manpower Planning in Hong Kong Public Hospitals. *Transportation Science*, 2:688–705, 2016.

[56] S. I. McClean and P. H. Millard. Where to treat the older patient? Can Markov models help us better understand the relationship between hospital and community care? *Journal of the Operational Research Society*, 58(2):255–261, 2007.

[57] J. Patrick. Access to long-term care: The true cause of hospital congestion? *Production and Operations Management*, 20(3):347–358, 2011.

[58] M. A. F. Ragab, W. Rashwan, W. Abo-Hamad, and A. Arisha. Using modelling and simulation to improve elderly care in Ireland: A case study. *International Journal on Advances in Life Sciences*, 5(1-2):89–102, 2013.

[59] B. Walker and T. Haslett. System dynamics and action research in aged care. *Australian health review : a publication of the Australian Hospital Association*, 24(1):183–191, 2001.

[60] N. Zychlinski, A. Mandelbaum, P. Momčilović, and I. Cohen. Bed blocking in hospitals due to scarce capacity in geriatric institutions—cost minimization via fluid models. *Manufacturing and Service Operations Management*, 22(2):396–411, 2020.

[61] M. J. Faddy and S. I. McClean. Markov chain modelling for geriatric patient care. *Methods of Information in Medicine*, 44(3):369–373, 2005.

[62] L. Garg, S. I. McClean, B. Meenan, and P. Millard. Non-homogeneous Markov models for sequential pattern mining of healthcare data. *IMA Journal of Management Mathematics*, 20(4):327–344, 2009.

[63] L. Garg, S. McClean, B. Meenan, and P. H. Millard. A non-homogeneous discrete time Markov model for admission scheduling and resource planning in a cost or capacity constrained healthcare system. *Health Care Management Science*, 13(2):155–169, 2010.

[64] A. S. Gordon, A. H. Marshall, and K. J. Cairns. A conditional approach for modelling patient readmissions to hospital using a mixture of Coxian phase-type distributions incorporating Bayes' theorem. *Statistics in Medicine*, 35(21):3810–3826, 2016.

[65] A. S. Gordon, A. H. Marshall, and M. Zenga. Predicting elderly patient length of stay in hospital and community care using a series of conditional Coxian phase-type distributions, further conditioned on a survival tree. *Health Care Management Science*, 21(2):269–280, 2018.

[66] W. L. Hare, A. Alimadad, H. Dodd, R. Ferguson, and A. Rutherford. A deterministic model of home and community care client counts in British Columbia. *Health Care Management Science*, 12(1):80–98, 2009.

[67] G. J. Taylor, S. I. McClean, and P. H. Millard. Stochastic models of geriatric patient bed occupancy behaviour. *Journal of the Royal Statistical Society. Series A: Statistics in Society*, 163(1):39–48, 2000.

[68] M. Cepoiu-Martin and D. P. Bischak. Policy choices in dementia care—An exploratory analysis of the Alberta continuing care system (ACCS) using system dynamics. *Journal of Evaluation in Clinical Practice*, 24(1):278–284, 2018.

[69] M. Franklin and R. M. Hunter. A modelling-based economic evaluation of primary-care-based fall-risk screening followed by fall-prevention intervention: A cohort-based Markov model stratified by older age groups. *Age and Ageing*, 49(1):57–66, 2019.

[70] N. Abe, T. Kakamu, T. Kumagai, T. Hidaka, Y. Masuishi, S. Endo, H. Kasuga, and T. Fukushima. Polypharmacy at admission prolongs length of hospitalization in gastrointestinal surgery patients. *Geriatrics and Gerontology International*, 20(11):1085–1090, 2020.

[71] N. Azad, F. Molnar, and A. Byszewski. Lessons learned from a multidisciplinary heart failure clinic for older women: A randomised controlled trial. *Age and Ageing*, 37(3):282–287, 2008.

[72] S. Kul, K. Vanhaecht, E. Savaş, Z. A. Öztürk, R. A. Parker, and M. Panella. Modeling of in-hospital treatment outcomes for elderly patients with heart failure: Care pathway versus usual care. *European Geriatric Medicine*, 4(2):86–90, 2013.

[73] B. Shaw and A. H. Marshall. Modelling the flow of congestive heart failure patients through a hospital system. *Journal of the Operational Research Society*, 58(2):212–218, 2007.

[74] L. A. Beaupre, J. G. Cinats, A. Senthilselvan, D. Lier, C. A. Jones, A. Scharfenberger, D. W. C. Johnston, and L. D. Saunders. Reduced morbidity for elderly patients with a hip fracture after implementation of a perioperative evidence-based clinical pathway. *Quality and Safety in Health Care*, 15(5):375–379, 2006.

[75] R. Wallace, L. D. G. Angus, S. Munnang, S. Shukry, J. C. DiGiacomo, and C. Ruotolo. Improved outcomes following implementation of a multidisciplinary care pathway for elderly hip fractures. *Aging Clinical and Experimental Research*, 31(2):273–278, 2019.

[76] W. Rashwan, W. Abo-Hamad, and A. Arisha. A system dynamics view of the acute bed blockage problem in the Irish healthcare system. *European Journal of Operational Research*, 247(1):276–293, 2015.

[77] D. Rossille, M. Cuggia, A. Arnault, J. Bouget, and P. Le Beux. Managing an emergency department by analysing HIS medical data: a focus on elderly patient clinical pathways. *Health Care Management Science*, 11(2):139–146, 2008.

[78] C. Trevisan, P. Di Gregorio, E. Debiasi, M. Pedrotti, M. La Guardia, E. Manzato, G. Sergi, and A. March. Decision tree for ward admissions of older patients at the emergency department after a fall. *Geriatrics and Gerontology International*, 18(9):1388–1392, 2018.

[79] G. Christodoulou and G. J. Taylor. Using a continuous time hidden Markov process, with covariates, to model bed occupancy of people aged over 65 years. *Health Care Management Science*, 4(1):21–24, 2001.

[80] T. Franck, V. Augusto, X. Xie, R. Gonthier, and E. Achour. Performance evaluation of an integrated care for geriatric departments using discrete-event simulation. pages 1331–1342, 2016.

[81] F. Gorunescu, S. I. McClean, and P. H. Millard. Using a queueing model to help plan bed allocation in a department of geriatric medicine. *Health Care Management Science*, 5(4):307–312, 2002.

[82] A. H. Marshall, H. Mitchell, and M. Zenga. Modelling the Length of Stay of Geriatric Patients in the Emilia Romagna Hospitals Using Coxian Phase-type Distributions with Covariates. *Studies in Theoretical and Applied Statistics, Selected Papers of the Statistical Societies*, pages 127–139, 2015.

[83] K. Bae, M. Jones, G. Evans, and D. Antimisiaris. Simulation modelling of patient flow and capacity planning for regional long-term care needs: a case study. *Health Systems*, 8(1):1–16, 2019.

[84] Z. D. Gassoumis, K. T. Fike, A. N. Rahman, S. M. Enguidanos, and K. H. Wilber. Who Transitions to the Community From Nursing Homes? Comparing Patterns and Predictors for Short-Stay and Long-Stay Residents. *Home Health Care Services Quarterly*, 32(2):75–91, 2013.

[85] Y. Li, Y. Zhang, N. Kong, and M. Lawley. Capacity planning for long-term care networks. *IIE Transactions (Institute of Industrial Engineers)*, 48(12):1098–1111, 2016.

[86] H. Mohammadi Bidhandi, J. Patrick, P. Noghani, and P. Varshoei. Capacity planning for a network of community health services. *European Journal of Operational Research*, 275(1):266–279, 2019.

[87] H. J. Welberry, H. Brodaty, B. Hsu, S. Barbieri, and L. R. Jorm. Impact of Prior Home Care on Length of Stay in Residential Care for Australians With Dementia. *Journal of the American Medical Directors Association*, 21(6):843–850.e5, 2020.

[88] Y. Zhang and M. L. Puterman. Developing an adaptive policy for long-term care capacity planning. *Health Care Management Science*, 16(3):271–279, 2013.

[89] M. Y. Lin, K. S. Chin, K. L. Tsui, and T. C. Wong. Genetic based discrete particle swarm optimization for Elderly Day Care Center timetabling. *Computers and Operations Research*, 65:125–138, 2016.

[90] P. Eveborn, M. Rönnqvist, H. Einarsdóttir, M. Eklund, K. Lidén, and M. Almroth. Operations research improves quality and efficiency in home care. *Interfaces*, 39(1):18–34, 2009.

[91] F. Grenouilleau, N. Lahrichi, and L.-M. Rousseau. New decomposition methods for home care scheduling with predefined visits. *Computers and Operations Research*, 115:104855, 2020.

[92] J. Guo, R. T. Konetzka, and W. G. Manning. The causal effects of home care use on institutional long-term care utilization and expenditures. *Health Economics (United Kingdom)*, 24(S1):4–17, 2015.

[93] S. Yalçindağ, P. Cappanera, M. G. Scutellá, E. Śahin, and A. Matta. Pattern-based decompositions for human resource planning in home health care service. *Computers and Operations Research*, 1:12–26, 2016.

[94] G. Arling, K. A. Abrahamson, V. Cooke, R. L. Kane, and T. Lewis. Facility and market factors affecting transitions from nursing home to community. *Medical Care*, 49(9):790–796, 2011.

[95] E. Borowiak, J. Kostka, and T. Kostka. Comparative analysis of the expected demands for nursing care services among older people from urban, rural, and institutional environments. *Clinical Interventions in Aging*, 10:405–412, 2015.

[96] N. Muramatsu, H. Yin, R. T. Campbell, R. L. Hoyem, M. A. Jacob, and C. O. Ross. Risk of nursing home admission among older Americans: Does states' spending on home- and community-based services matter? *Journals of Gerontology - Series B Psychological Sciences and Social Sciences*, 62(3):S169–S178, 2007.

[97] R. C. Ambagtsheer, J. Beilby, C. Seiboth, and E. Dent. Prevalence and associations of frailty in residents of Australian aged care facilities: findings from a retrospective cohort study. *Aging Clinical and Experimental Research*, 32(9):1849–1856, 2020.

[98] E. Arvelo, J. de Armas, and M. Guillen. Assessing the Distribution of Elderly Requiring Care: A Case Study on the Residents in Barcelona and the Impact of COVID-19. *International Journal of Environmental Research and Public Health*, 17(20):7486, 2020.

[99] M. S. Desai, M. L. Penn, S. Brailsford, and M. Chipulu. Modelling of Hampshire Adult Services - Gearing up for future demands. *Health Care Management Science*, 11(2):167–176, 2008.

[100] E. Eggink, I. Woittiez, and M. Ras. Forecasting the use of elderly care: a static micro-simulation model. *European Journal of Health Economics*, 17(6):681–691, 2016.

[101] K. Katsaliaki, S. Brailsford, D. Browning, and P. Knight. Mapping care pathways for the elderly. *Journal of Health, Organisation and Management*, 19(1):57–72, 2005.

[102] Z. Tao, Y. Cheng, T. Dai, and M. W. Rosenberg. Spatial optimization of residential care facility locations in Beijing, China: Maximum equity in accessibility. *International Journal of Health Geographics*, 13(1), 2014.

[103] H. Xie, T. J. Chaussalet, and P. H. Millard. A continuous time Markov model for the length of stay of elderly people in institutional long-term care. *Journal of the Royal Statistical Society. Series A: Statistics in Society*, 168(1):51–61, 2005.

[104] E. Aspland, D. Gartner, and P. Harper. Clinical pathway modelling: a literature review. *Health Systems*, 10(1):1–23, 2021.

[105] Y. Zhang, R. Padman, and J. E. Levin. Paving the COWpath: data-driven design of pediatric order sets. *Journal of the American Medical Informatics Association*, 21(e2):e304–e311, 2014.

[106] B. S. S. Onggo, N. C. Proudlove, S. A. D'Ambrogio, A. Calabrese, S. Bisogno, and N. Levialdi Ghiron. A BPMN extension to support discrete-event simulation for healthcare applications: An explicit representation of queues, attributes and data-driven decision points. *Journal of the Operational Research Society*, 69(5):788–802, 2018.

[107] K. M. Silvester, M. A. Mohammed, P. Harriman, A. Girolami, and T. W. Downes. Timely care for frail older people referred to hospital improves efficiency and reduces mortality without the need for extra resources. *Age and Ageing*, 43(4):472–477, 2014.

[108] Y. Zhang, M. L. Puterman, M. Nelson, and D. R. Atkins. A simulation optimization approach to long-term care capacity planning. *Operations Research*, 60(2):249–261, 2012.

[109] T. Heggestad. Do hospital length of stay and staffing ratio affect elderly patients' risk of readmission? A nation-wide study of Norwegian hospitals. *Health Services Research*, 37(3):647–665, 2002.

[110] A. H. Marshall and S. I. McClean. Conditional phase-type distributions for modelling patient length of stay in hospital. *International Transactions in Operational Research*, 10(6):565–576, 2003.

[111] A. H. Marshall and S. I. McClean. Using coxian phase-type distributions to identify patient characteristics for duration of stay in hospital. *Health Care Management Science*, 7(4):285–289, 2004.

[112] T. J. Chaussalet, H. Xie, and P. H. Millard. A closed queueing network approach to the analysis of patient flow in health care systems. *Methods of Information in Medicine*, 45(5):492–497, 2006.

[113] F. E. Hamdani, M. Masmoudi, A. Al Hanbali, F. Bouyahia, and A. A. Ouahman. Diagnostic and modeling of elderly flow in a French healthcare institution. *Computers and Industrial Engineering*, 112:675–689, 2017.

[114] G. W. Harrison. Implications of mixed exponential occupancy distributions and patient flow models for health care planning. *Health Care Management Science*, 4(1):37–45, 2001.

[115] T. K. Abe, B. M. Beamon, R. L. Storch, and J. Agus. Operations research applications in hospital operations: Part II. *Health Care Operations Management*, 6(2):96–109, 2016.

[116] T. K. Abe, B. M. Beamon, R. L. Storch, and J. Agus. Operations research applications in hospital operations: Part III. *Health Care Operations Management*, 6(3):175–191, 2016.

[117] C. Corke, E. de Leeuw, S. K. Lo, and C. George. Predicting Future Intensive Care Demand in Australia. *Critical Care and Resuscitation*, 11(4):257–260, 2019.

[118] J. Kahn. Matching supply and demand in critical care. *Current Opinion in Critical Care*, 17(6):618–619, 2011.

[119] World Health Organization. Coronavirus disease (COVID-19) pandemic. Available online at: "`https://www.who.int/europe/emergencies/situations/covid-19`". Last Accessed 22 March 2023, 2023.

[120] J. J. Luna. Hierarchical relation in simulation models. In *Proceedings of the 1993 Winter Simulation Conference*, 1993.

[121] C. Hunt-O'Connor, Z. Moore, D. Patton, L. Nugent, P. Avsar, and T. O'Connor. The effect of discharge planning on length of stay and readmission rates of older adults in acute hospitals: A systematic review and meta-analysis of systematic reviews. *Journal of Nursing Management*, 29(8):2697–2706, 2021.

[122] C. Pellett. Discharge planning: best practice in transitions of care. *British Journal of Community Nursing*, 21(11):542–548, 2016.

[123] F. Cacciatore, C. B. Anello, N. Ferrara, F. Mazzella, M. Manzi, U De Angelis, and Pasquale Abete. Determinants of prolonged intensive care unit stay after cardiac surgery in the elderly. *Aging Clinical and Experimental Research*, 24, 2012.

[124] A. Bahrmann, L. Benner, M. Christ, T. Bertsch, C. C. Sieber, H. Katus, and P. Bahrmann. The Charlson Comorbidity and Barthel Index predict length of hospital stay, mortality, cardiovascular mortality and rehospitalization in unselected older patients admitted to the emergency department. *Aging Clinical and Experimental Research*, 31(9):1233–1242, 2018.

[125] S. Tal. Length of hospital stay among oldest-old patients in acute geriatric ward. *Archives of Gerontology and Geriatrics*, 94:104352, 2021.

[126] Q. Cai, J. W. Salmon, and M. E. Rodgers. Factors Associated with Long-Stay Nursing Home Admissions Among the U.S. Elderly Population: Comparison of Logistic Regression and the Cox Proportional Hazards Model with Policy Implications for Social Work. *Social Work in Health Care*, 48(2):154–168, 2009.

[127] M. Hoben, S. A. Chamberlain, A. Gruneir, J. A. Knopp-Sihota, J. M. Sutherland, J. W. Poss, M. B. Doupe, V. Bergstrom, P. G. Norton, C. Schalm, K. McCarthy, K. Kashuba, F. Ackah, and C. A. Estabrooks. Nursing Home Length of Stay in 3 Canadian Health Regions: Temporal Trends, Jurisdictional Differences, and Associated Factors. *Journal of the American Medical Directors Association*, 20(9):1121–1128, 2019.

[128] K. S. Johnson, M. Kuchibhatla, and J. A. Tulsky. Racial Differences in Location Before Hospice Enrollment and Association with Hospice Length of Stay. *Journal of the American Geriatrics Society*, 59(4):732–737, 2011.

[129] N. S. Park, I. V. Carrion, B. S. Lee, D. Dobbs, H. J. Shin, and M. A. Becker. The Role of Race and Ethnicity in Predicting Length of Hospice Care among Older Adults. *Journal of Palliative Medicine*, 15(2):149–153, feb 2012.

[130] L. Fan, Y. Tian, J. Wang, Y. Ding, S. Wang, H. Xue, and W. Du. Frailty Predicts Increased Health Care Utilization Among Community-Dwelling Older Adults: A Longitudinal Study in China. *Journal of the American Medical Directors Association*, 22(9):1819–1824, 2021.

[131] B. Walsh, S. Lyons, S. Smith, M.-A. Wren, J. Eighan, and E. Morgenroth. Does formal home care reduce inpatient length of stay? *Health Economics*, 29(12):1620–1636, 2020.

[132] D. K. Sommerfeld, H. Johansson, A.-L. Jönsson, V. Murray, T. Wessari, L. W. Holmqvist, and M. von Arbin. Rivermead Mobility Index Can Be Used to Predict Length of Stay for Elderly Persons, 5 Days After Stroke Onset. *Journal of Geriatric Physical Therapy*, 34(2):64–71, 2011.

[133] B. Rajamaki, M. Koponen, S. Hartikainen, and A. m. Tolppanen. Length of hospital stay after hip fracture and readmission rates of persons with and without Alzheimer's disease: a matched cohort study. *BMC Geriatrics*, 20(1), 2020.

[134] L. Marano, L. Carbone, G. Edoardo Poto, M. Gambelli, L. L. N. Noudem, G. Grassi, F. Manasci, G. Curreri, A. Giuliani, R. Piagnerelli, V. Savelli, D. Marrelli, F. Roviello, and V. Boccardi. Handgrip strength predicts length of hospital stay in an abdominal surgical setting: the role of frailty beyond age. *Aging Clinical and Experimental Research*, 34(4):811–817, 2022.

[135] G. Raab, D. Restifo, S. M. McBride, R. J. Wong, N. Y. Lee, A. Shahrokni, and Kaveh K. Outcomes following head and neck cancer surgery among older adults as determined by an electronic geriatric assessment. *Journal of Geriatric Oncology*, 13(2):228–233, 2022.

[136] A. Kirfel, J. Menzenbach, V. Guttenthaler, J. Feggeler, A. Mayr, M. Coburn, and M. Wittmann. Postoperative delirium after cardiac surgery of elderly patients as an independent risk factor for prolonged length of stay in intensive care unit and in hospital. *Aging Clinical and Experimental Research*, 33(11):3047–3056, 2021.

[137] D. H. Lee, K. J. Buth, B.-J. Martin, A. M. Yip, and G. M. Hirsch. Frail Patients Are at Increased Risk for Mortality and Prolonged Institutional Care After Cardiac Surgery. *Circulation*, 121(8):973–978, mar 2010.

[138] A. Pustavoitau, V. Barodka, N. E. Sharpless, C. Torrice, D. Nyhan, D. E. Berkowitz, A. S. Shah, K. J. B. Roche, and J. D. Walston. Role of senescence marker p16INK4a measured in peripheral blood T-lymphocytes in predicting length of hospital stay after coronary artery bypass surgery in older adults. *Experimental Gerontology*, 74:29–36, 2016.

[139] S. C. Agasi-Idenburg, C. J. A. Punt, N. K. Aaronson, and M. M. Stuiver. The association between preoperative fatigue and instrumental activities in daily living with complications and length of hospital stay in patients undergoing colorectal surgery. *Aging Clinical and Experimental Research*, 32(2):257–264, 2019.

[140] D. Justo, N. Vislapu, V. Shvedov, M. Fickte, A. Danylesko, P. Kimelman, C. Merdler, and Y. Lerman. Admission Norton scale scores (ANSS) correlate with rehabilitation outcome and length in elderly patients following hip arthroplasty. *Archives of Gerontology and Geriatrics*, 53(1):e33–e36, 2011.

[141] J. M. Willems, A. J. M. de Craen, R. G. H. H. Nelissen, P. A. van Luijt, R. G. J. Westendorp, and G. J. Blauw. Haemoglobin predicts length of hospital stay after hip fracture surgery in older patients. *Maturitas*, 72(3):225–228, 2012.

[142] V. MacDonald, P. Ottem, M. Wasdell, and R. Spiwak. Predictors of prolonged hospital stays following hip and knee arthroplasty. *International Journal of Orthopaedic and Trauma Nursing*, 14(4):198–205, 2010.

[143] L. A. Harvey, B. Toson, C. Norris, I. A. Harris, R. C. Gandy, and J. J. C. T. Close. Does identifying frailty from ICD-10 coded data on hospital admission improve prediction of adverse outcomes in older surgical patients? A population-based study. *Age and Ageing*, 50(3):802–808, 2020.

[144] T. S. Jones, E. L. Jones, V. Richardson, J. B. Finley, J. L. Franklin, D. L. Gore, C. P. Horney, A. Kovar, T. L. Morin, and T. N. Robinson. Preliminary data demonstrate the Geriatric Surgery Verification program reduces postoperative length of stay. *Journal of the American Geriatrics Society*, 69(7):1993–1999, 2021.

[145] A. Kerr, H. E. Syddall, C. Cooper, G. F. Turner, R. S. Briggs, and A. Aihie Sayer. Does admission grip strength predict length of stay in hospitalised older patients? *Age and Ageing*, 35(1):82–84, 2006.

[146] D. Zattoni, I. Montroni, N. M. Saur, A. Garutti, M. L. B. Reggiani, C. Galetti, P. Calogero, and V. Tonini. A Simple Screening Tool to Predict Outcomes in Older Adults Undergoing Emergency General Surgery. *Journal of the American Geriatrics Society*, 67(2):309–316, 2018.

[147] Y. Zhao, N. Ge, D. Xie, L. Gao, Y. Wang, Y. Liao, and J. Yue. The geriatric nutrition risk index versus the mini-nutritional assessment short form in predicting postoperative delirium and hospital length of stay among older non-cardiac surgical patients: a prospective cohort study. *BMC Geriatrics*, 20(1), 2020.

[148] Y. Zhao, X. Xia, D. Xie, Y. Liao, Y. Wang, L. Chen, N. Ge, and J. Yue. Geriatric Nutritional Risk Index can predict postoperative delirium and hospital length of stay in elderly patients undergoing non-cardiac surgery. *Geriatrics International*, 20(8):759–764, 2020.

[149] M. S. Alyahya, H. H. Hijazi, H. A. Alshraideh, and A. D. Al-Nasser. Using decision trees to explore the association between the length of stay and potentially avoidable readmissions: A retrospective cohort study. *Informatics for Health and Social Care*, 42(4):361–377, 2017.

[150] R. Antonelli, C. Pedone, G. Onder, M. Pahor, and P. U. Carbonin. Predicting length of stay of older patients with exacerbated chronic obstructive pulmonary disease. *Aging Clinical and Experimental Research*, 13(1):49–57, 2001.

[151] D. Basic and A. Khoo. Admission variables predicting short lengths of stay of acutely unwell older patients: relevance to emergency and medical short-stay units. *Australian Health Review*, 33(3):502, 2009.

[152] O. Beauchet, S. Fung, C. P. Launay, J. Afilalo, P. Herbert, M. Afilalo, and J. Chabot. Predicting a long hospital stay after admission to a geriatric assessment unit: Results from an observational retrospective cohort study. *Maturitas*, 115:110–114, 2018.

[153] M. Bo, G. Fonte, F. Pivaro, M. Bonetto, C. Comi, V. Giorgis, L. Marchese, G. Isaia, G. Maggiani, E. Furno, Y. Falcone, and G. C. Isaia. Prevalence

of and factors associated with prolonged length of stay in older hospitalized medical patients. *Geriatrics &amp Gerontology International*, 16(3):314–321, 2015.

[154] M. Bo, M. Bonetto, G. Bottignole, P. Porrino, E. Coppo, M. Tibaldi, G. Ceci, S. Raspo, G. Cappa, and G. Bellelli. Length of Stay in the Emergency Department and Occurrence of Delirium in Older Medical Patients. *Journal of the American Geriatrics Society*, 64(5):1114–1119, 2016.

[155] P. K. Curiati, L. A. Gil-Junior, C. V. Morinaga, F. Ganem, J. A. E. Curiati, and T. J. Avelino-Silva. Predicting Hospital Admission and Prolonged Length of Stay in Older Adults in the Emergency Department: The PRO-AGE Scoring System. *Annals of Emergency Medicine*, 76(3):255–265, 2020.

[156] M. S. Ferreira, F. G. de Melo Franco, P. S. Rodrigues, V. M. S. P. Correa, S. T. G. Akopian, G. G. Cucato, R. M. R. Dias, M. S. Cendoroglo, C. N. França, and J. A. M. de Carvalho. Impaired chair-to-bed transfer ability leads to longer hospital stays among elderly patients. *BMC Geriatrics*, 19(1), 2019.

[157] Z. Hu, H. Qiu, L. Wang, and M. Shen. Network analytics and machine learning for predicting length of stay in elderly patients with chronic diseases at point of admission. *BMC Medical Informatics and Decision Making*, 22(1), 2022.

[158] C. P. Launay, H. Rivière, A. Kabeshova, and O. Beauchet. Predicting prolonged length of hospital stay in older emergency department users: Use of a novel analysis method, the Artificial Neural Network. *European Journal of Internal Medicine*, 26(7):478–482, 2015.

[159] C. P. Launay, A. Kabeshova, A. Lanoé, J. Chabot, E. J. Levinoff, and O. Beauchet. Age effect on the prediction of risk of prolonged length hospital stay in older patients visiting the emergency department: results from a large prospective geriatric cohort study. *BMC Geriatrics*, 18(1), 2018.

[160] R. Lisk, M. Uddin, A. Parbhoo, K. Yeong, D. Fluck, P. Sharma, M. E. J. Lean, and T. S. Han. Predictive model of length of stay in hospital among older patients. *Aging Clinical and Experimental Research*, 31(7):993–999, 2018.

[161] M. Nishino, Y. Matsuhiro, H. Nakamura, K. Yasumoto, K. Yasumura, A. Tanaka, Y. Matsunaga-Lee, D. Nakamura, M. Yano, M. Yamato, Y. Egami, R. Shutta, and J. Tanouchi. Which factors are associated with length of stay in older patients with acute decompensated heart failure with preserved ejection fraction?: AURORA study. *Geriatrics & Gerontology International*, 19(11):1084–1087, 2019.

[162] D. K. Sommerfeld and M. H. von Arbin. Disability test 10 days after acute stroke to predict early discharge home in patients 65 years and older. *Clinical Rehabilitation*, 15(5):528–534, 2001.

[163] C. Takahashi, K. Fushimi, and S. Matsuda. Factors associated with a protracted hospital stay after hip fracture surgery in Japan. *Geriatrics & Gerontology International*, 11(4):474–481, 2011.

[164] P. Feuerstadt, W. W. Nelson, E. M. Drozd, J. Dreyfus, D. N. Dahdal, A. C. Wong, I. Mohammadi, C. Teigland, and A. Amin. Mortality, Health Care Use, and Costs of Clostridioides difficile Infections in Older Adults. *Journal of the American Medical Directors Association*, 23(10):1721–1728.e19, 2022.

[165] J. Tropea, D. LoGiudice, D. Liew, A. Gorelik, and C. Brand. Poorer outcomes and greater healthcare costs for hospitalised older people with dementia and delirium: a retrospective cohort study. *International Journal of Geriatric Psychiatry*, 32(5):539–547, 2016.

[166] S. Volpato, S. Bazzano, A. Fontana, L. Ferrucci, and A. Pilotto. Multidimensional Prognostic Index Predicts Mortality and Length of Stay During Hospitalization in the Older Patients: A Multicenter Prospective Study. *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences*, 70(3):325–331, 2014.

[167] A. Morandi, G. Bellelli, E. E. Vasilevskis, R. Turco, F. Guerini, T. Torpilliesi, S. Speciale, V. Emiliani, S. Gentile, J. Schnelle, and M. Trabucchi. Predictors of Rehospitalization Among Elderly Patients Admitted to a Rehabilitation Hospital: The Role of Polypharmacy, Functional Status, and Length of Stay. *Journal of the American Medical Directors Association*, 14(10):761–767, 2013.

[168] D. Basic and A. Khoo. New medical diagnoses and length of stay of acutely unwell older patients: Implications for funding models. *Australasian Journal on Ageing*, 34(3):160–165, 2015.

[169] D. Adamis, D. Meagher, S. Rooney, O. Mulligan, and G. McCarthy. A comparison of outcomes according to different diagnostic systems for delirium (DSM-5, DSM-IV, CAM, and DRS-R98). *International Psychogeriatrics*, 30(4):591–596, 2017.

[170] T. Motohashi, H. Hamada, J. Lee, M. Sekimoto, and Y. Imanaka. Factors associated with prolonged length of hospital stay of elderly patients in acute care hospitals in Japan: A multilevel analysis of patients with femoral neck fracture. *Health Policy*, 111(1):60–67, 2013.

[171] D. Naouri, N. Pelletier-Fleury, N. Lapidus, and Y. Yordanov. The effect of direct admission to acute geriatric units compared to admission after an emergency department visit on length of stay, postacute care transfers and ED return visits. *BMC Geriatrics*, 22(1), 2022.

[172] W. Chung, S. M. Oh, T. Suh, Y. M. Lee, B. H. Oh, and C. W. Yoon. Determinants of length of stay for psychiatric inpatients: Analysis of a national database covering the entire Korean elderly population. *Health Policy*, 94(2):120–128, 2010.

[173] T. Motzek, A. Werblow, F. Tesch, G. Marquardt, and J. Schmitt. Determinants of hospitalization and length of stay among people with dementia – An analysis of statutory health insurance claims data. *Archives of Gerontology and Geriatrics*, 76:227–233, 2018.

[174] D. Basic and C. Shanley. Frailty in an older inpatient population: Using the clinical frailty scale to predict patient outcomes. *Journal of Aging and Health*, 27(4):670–685, 2014.

[175] R. Lisk, K. Yeong, P. Enwere, J. Jenkinson, J. Robin, M. Irvin-Sellers, D. Fluck, A. Osmani, R. Sharmin, P. Sharma, C. H. Fry, and T. S. Han. Associations of 4AT with mobility, length of stay and mortality in hospital and discharge destination among patients admitted with hip fractures. *Age and Ageing*, 49(3):411–417, 2019.

[176] S. Nishida, Y. Hayashi, A. Suzuki, R. Kobayashi, T. Inuzuka, and Y. Itoh. Relationship between number of drugs and duration of hospital stay in older patients with neuromuscular diseases. *Geriatrics &amp Gerontology International*, 18(7):1018–1024, 2018.

[177] Y. Shen, Q. Hao, S. Liu, L. Su, X. Sun, J. H. Flaherty, B. Dong, and L. Cao. The impact of functional status on LOS and readmission in older patients in geriatrics department: a cohort study. *Aging Clinical and Experimental Research*, 32(10):1977–1983, 2019.

[178] H. J. Toh, Z. Y. Lim, P. Yap, and T. Tang. Factors associated with prolonged length of stay in older patients. *Singapore Medical Journal*, 58(3):134–138, 2017.

[179] S. Volpato, R. Custureri, M. Puntoni, L. Bianchi, J. Daragjati, S. Garaboldi, M. Simonato, A. Greco, E. Rizzo, P. D. Santo, A. Mangiacotti, E. Zigoura, G. Siri, and A. Pilotto. Effects of oral amino acid supplementation on Multidimensional Prognostic Index in hospitalized older patients: a multicenter randomized, double-blind, placebo-controlled pilot study. *Clinical Interventions in Aging*, Volume 13:633–640, 2018.

[180] R. E. Hubbard, N. M. Peel, M. Samanta, L. C. Gray, A. Mitnitski, and K. Rockwood. Frailty status at admission to hospital predicts multiple adverse outcomes. *Age and Ageing*, 46(5):801–806, 2017.

[181] P.-O. Lang, D. Heitz, G. Hédelin, M. Dramé, N. Jovenin, J. Ankri, D. Somme, J.-L. Novella, J. B. Gauvain, P. Couturier, T. Voisin, B. De Wazère, R. Gonthier, C. Jeandel, D. Jolly, O. Saint-Jean, and F. Blanchard. Early Markers of Prolonged Hospital Stays in Older People: A Prospective, Multicenter Study of 908 Inpatients in French Acute Hospitals. *Journal of the American Geriatrics Society*, 54(7):1031–1039, 2006.

[182] P.-O. Lang, D. Zekry, J.-P. Michel, M. Drame, J.-L. Novella, D. Jolly, and F. Blanchard. Early markers of prolonged hospital stay in demented inpatients: A multicentre and prospective study. *The Journal of Nutrition, Health & Aging*, 14(2):141–147, 2009.

[183] M. Rubens, A. Cristian, V. Ramamoorthy, A. Saxena, P. McGranaghan, R. Tonse, and E. Veledar. Effect of frailty on hospital outcomes among patients with cancer in the United States: Results from the National Inpatient Sample. *Journal of Geriatric Oncology*, 13(7):1043–1049, 2022.

[184] D. S. Shebeshi, X. Dolja-Gore, and J. Byles. Validation of hospital frailty risk score to predict hospital use in older people: Evidence from the Australian Longitudinal Study on Women's Health. *Archives of Gerontology and Geriatrics*, 92:104282, 2021.

[185] O. Beauchet, C. Launay, L. De Decker, B. Fantino, A. Kabeshova, and C. Annweiler. Who is at risk of long hospital stay among patients admitted to geriatric acute care unit? Results from a prospective cohort study. *The Journal of Nutrition, Health & Aging*, 17(8):695–699, 2013.

[186] M. Chua, G. E. Hoyle, and R. L. Soiza. Prognostic implications of hyponatremia in elderly hospitalized patients. *Archives of Gerontology and Geriatrics*, 45(3):253–258, 2007.

[187] M. B. Snowden, A. Walaszek, J. E. Russo, K. A. Comtois, D. S. Srebnik, R. K. Ries, and P. P. Roy-Byrne. Geriatric Patients Improve as Much as Younger Patients from Hospitalization on General Psychiatric Units. *Journal of the American Geriatrics Society*, 52(10):1676–1680, 2004.

[188] S. T. S. Yu, T. Brown, M.-L. Yu, and H. Andrews. Association between Older Adults' Estimated Length of Hospital Stay and Cognitive Performance. *Physical &amp Occupational Therapy In Geriatrics*, 36(2-3):136–155, 2018.

[189] K. Hasebe, R. Momosaki, M. Sawabe, M. Chono, A. Sawaguchi, S. Kasuga, D. Asanuma, S. Suzuki, N. Miyauchi, and M. Abo. Effectiveness of weekend physical rehabilitation for functional recovery in geriatric patients with hip fracture. *Geriatrics & Gerontology International*, 18(8):1143–1146, 2018.

[190] J.-H. Kim and Y. Lee. Implementation of long-term care and hospital utilization: Results of segmented regression analysis of interrupted time series study. *Archives of Gerontology and Geriatrics*, 78:221–226, 2018.

[191] G. Liotta, S. Mancinelli, P. Scarcella, and L. E. Gialloreti. Determinants of acute hospital care use by elderly patients in Italy from 1996 to 2006. *Archives of Gerontology and Geriatrics*, 54(3):e364–e369, 2012.

[192] R. Y. Wong and W. C. Miller. Adverse outcomes following hospitalization in acutely ill older patients. *BMC Geriatrics*, 8(1), 2008.

[193] P. J. Hartley, V. L. Keevil, L. Alushi, R. L. Charles, E. B. Conroy, P. M. Costello, B. Dixon, A. M. Dolinska-Grzybek, D. Vajda, and R. Romero-Ortuno. Earlier Physical Therapy Input Is Associated With a Reduced Length of Hospital Stay and Reduced Care Needs on Discharge in Frail Older Inpatients: An Observational Study. *Journal of Geriatric Physical Therapy*, 42(2):E7–E14, 2019.

[194] C. P. Launay, L. de Decker, A. Kabeshova, C. Annweiler, and O. Beauchet. Screening for Older Emergency Department Inpatients at Risk of Prolonged Hospital Stay: The Brief Geriatric Assessment Tool. *PLoS ONE*, 9(10):e110135, 2014.

[195] T. Ono, A. Tamai, D. Takeuchi, Y. Tamai, H. Iseki, H. Fukushima, and S. Kasahara. Predictors of length of stay in a ward for demented elderly: gender differences. *Psychogeriatrics*, 10(3):153–159, 2010.

[196] A. C. Kidd, P. Musonda, R. L. Soiza, C. Butchart, C. J. Lunt, Y. Pai, Y. Hameed, C. Fox, J. F. Potter, and P. K. Myint. The relationship between total anticholinergic burden (ACB) and early in-patient hospital mortality and length of stay in the oldest old aged 90 years and over admitted with an acute illness. *Archives of Gerontology and Geriatrics*, 59(1):155–161, 2014.

[197] A. Pilotto, D. Sancarlo, F. Pellegrini, F. Rengo, N. Marchionni, S. Volpato, and L. Ferrucci. The Multidimensional Prognostic Index predicts in-hospital length of stay in older patients: a multicentre prospective study. *Age and Ageing*, 45(1):90–96, 2016.

[198] O. Beauchet, J. Lubov, K. Galery, M. Afilalo, and C. P. Launay. Emergency room evaluation and recommendations for older emergency department users: results of the ER^2 experimental study. *European Geriatric Medicine*, 12(5):921–929, jun 2021.

[199] E. Greene, C. J. Cunningham, A. Eustace, N. Kidd, A. W. Clare, and B. A. Lawlor. Recurrent falls are associated with increased length of stay in elderly psychiatric inpatients. *International Journal of Geriatric Psychiatry*, 16(10):965–968, 2001.

[200] D. Harari, F. C. Martin, A. Buttery, S. O'Neill, and A. Hopper. The older persons' assessment and liaison team 'OPAL': evaluation of comprehensive geriatric assessment in acute medical inpatients. *Age and Ageing*, 36(6):670–675, nov 2007.

[201] J. S. Marín, F. J. F. Martínez, C. F. Such, J. M. S. Ripoll, D. O. Beltrán, M. C. C. Munuera, J. F. M. López, and J. C. M. Campos. Risk factors for high length of hospital stay and in-hospital mortality in hip fractures in the elderly. *Revista Española de Cirugía Ortopédica y Traumatología (English Edition)*, 65(5):322–330, 2021.

[202] P. N. Wright, G. Tan, S. Iliffe, and D. Lee. The impact of a new emergency admission avoidance system for older people on length of stay and same-day discharges. *Age and Ageing*, 43(1):116–121, 2013.

[203] H. Chen, Y. Hara, N. Horita, Y. Saigusa, Y. Hirai, and T. Kaneko. Declined Functional Status Prolonged Hospital Stay for Community-Acquired Pneumonia in Seniors. *Clinical Interventions in Aging*, Volume 15:1513–1519, 2020.

[204] T. V. Le, C. K. Kwoh, E. S. Teo, and K. H. Lee. Analyzing Trends of Hospital Length of Stay Using Phase-Type Distributions. In *2011 IEEE 11th International Conference on Data Mining Workshops*. IEEE, dec 2011.

[205] L. Garg, S. McClean, B. J. Meenan, and P. Millard. ”Phase-Type Survival Trees and Mixed Distribution Survival Trees for Clustering Patients’ Hospital Length of Stay”. *Informatica*, 22(1):55–72, 2011.

[206] T. Möllers, L. Perna, P. Ihle, I. Schubert, J. Bauer, and H. Brenner. Factors associated with length of stay in hospital patients with and without dementia. *Journal of Alzheimer's Disease*, 67(3):1055–1065, feb 2019.

[207] R. Lisk, K. Yeong, D. Fluck, C. H. Fry, and T. S. Han. The Ability of the Nottingham Hip Fracture Score to Predict Mobility, Length of Stay and Mortality in Hospital, and Discharge Destination in Patients Admitted with a Hip Fracture. *Calcified Tissue International*, 107(4):319–326, jul 2020.

[208] M. P. Quinn, A. E. Courtney, D. G. Fogarty, D. O'Reilly, C. Cardwell, and P. T. McNamee. Influence of prolonged hospitalization on overall bed occupancy: a five-year single-centre study. *QJM*, 100(9):561–566, 2007.

[209] Organisation for Economic Cooperation and Development (OECD). Hospital beds and occupancy. Available online at: "`https://www.oecd-ilibrary.org/sites/e5a80353-en/index.html?itemId=/content/component/e5a80353-en`". Last Accessed 22 March 2023, 2023.

[210] British Medical Association. NHS hospital beds data analysis. Available online at: "`https://www.bma.org.uk/advice-and-support/nhs-delivery-and-workforce/pressures/nhs-hospital-beds-data-analysis#:~:text=Since%202010%2C%20average%20bed%20occupancy,and%20efficiency%20are%20at%20risk`". Last Accessed 22 March 2023, 2023.

[211] Centers for Disease Control and Prevention. Hospitals, beds, and occupancy rates, by type of ownership and size of hospital: United states, selected years 1975–2015. Available online at "`https://www.cdc.gov/nchs/data/hus/2017/089.pdf`." Accessed 6 January 2023, 2017.

[212] NHS Jobs. Working in the NHS. Available online at: "`https://www.jobs.nhs.uk/about_nhs.html`". Last Accessed 22 March 2023, 2022.

[213] Office for National Statistics. Overview of the UK population: 2020. Available online: "`https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/articles/overviewoftheukpopulation/2020`". Last Accessed 22 March 2023, 2022.

[214] L. H. Aiken, D. M. Sloane, J. Ball, L. Bruyneel, A. M. Rafferty, and P. Griffiths. Patient satisfaction with hospital care and nurses in England: an observational study. *BMJ Open*, 8(1):e019189, 2021.

[215] V. Kumar and M. L. Garg. Predictive Analytics: A Review of Trends and Techniques. *International Journal of Computer Applications*, 182(1):31–37, 2018.

[216] M. Li, P. Vanberkel, and X. Zhong. Predicting ambulance offload delay using a hybrid decision tree model. *Socio-Economic Planning Sciences*, 80:101146, 2022.

[217] R. Healey, C. Naugler, L. de Koning, and J. L. Patel. *Leukemia & Lymphoma, title = A classification tree approach for improving the utilization of flow cytometry testing of blood specimens for B-cell non-Hodgkin lymphoproliferative disorders*, 56(9):2619–2624, 2015.

[218] C. Salazar-Concha and P. Ramírez-Correa. Predicting the Intention to Donate Blood among Blood Donors Using a Decision Tree Algorithm. *Symmetry*, 13(8):1460, 2021.

[219] K. Shah, R. Punjabi, P. Shah, and M. Rao. Real Time Diabetes Predictionusing Naïve Bayes Classifier on BigData of Healthcare. *International Research Journal of Engineering and Technology*, 7(5):102–107, 2021.

[220] G. Subbalakshmi, K. Ramesh, and M. Chinna Rao. Decision Support in Heart Disease Prediction System using Naive Bayes. *Indian Journal of Computer Science and Engineering*, 2(2):170–176, 2011.

[221] K. Vembandasamy, R. Sasipriya, and E. Deepa. Heart Diseases Detection Using Naive Bayes Algorithm. *International Journal of Innovative Science, Engineering Technology*, 2(9):441–444, 2015.

[222] H. A. Hanson, C. Martin, B. O'Neil, C. L. Leiser, E. N. Mayer, K. R. Smith, and W. T. Lowrance. The Relative Importance of Race Compared to Health Care and Social Factors in Predicting Prostate Cancer Mortality: A Random Forest Approach. *Journal of Urology*, 202(6):1209–1216, 2019.

[223] S. Benbelkacem and B. Atmani. Random Forests for Diabetes Diagnosis. In *2019 International Conference on Computer and Information Sciences (ICCIS)*, 2019.

[224] S. Rallapalli and T. Suryakanthi. Predicting the risk of diabetes in big data electronic health Records by using scalable random forest classification algorithm. In *2016 International Conference on Advances in Computing and Communication Engineering (ICACCE)*, 2016.

[225] S. K. Lakshmanaprabu, S. N. Mohanty, S. S. Rani, S. Krishnamoorthy, J. Uthayakumar, and K. Shankar. Online clinical decision support system using optimal deep neural networks. *Applied Soft Computing*, 81:105487, 2019.

[226] D. Lin, A. V. Vasilakos, Y. Tang, and Y. Yao. Neural networks for computer-aided diagnosis in medicine: A review. *Neurocomputing*, 216:700–708, 2016.

[227] L. C. Chen, J. T. Sheu, Y. J. Chuang, and Y. Tsao. Predicting the Travel Distance of Patients to Access Healthcare Using Deep Neural Networks. *Journal of Translational Engineering in Health and Medicine*, 10:1–11, 2022.

[228] G. Battineni, N. Chintalapudi, and F. Amenta. Machine learning in medicine: Performance calculation of dementia prediction by support vector machines (SVM). *Informatics in Medicine Unlocked*, 16:100200, 2019.

[229] N. D. Forkert, T. Verleger, B. Cheng, G. Thomalla, C. C. Hilgetag, and J. Fiehler. Multiclass Support Vector Machine-Based Lesion Mapping Predicts Functional Outcome in Ischemic Stroke Patients. *PLOS ONE*, 10(6):e0129569, 2015.

[230] J. Kamruzzaman and R. K. Begg. Support vector machines and other pattern recognition approaches to the diagnosis of cerebral palsy gait. *IEEE Transactions on Biomedical Engineering*, 53(12):2479–2490, 2006.

[231] E. Šabić, D. Keeley, B. Henderson, and S. Nannemann. Healthcare and anomaly detection: using machine learning to predict anomalies in heart rate data. *AI & SOCIETY*, 36(1):149–158, 2020.

[232] S. Chauhan and L. Vig. Anomaly detection in ECG time signals via deep long short-term memory networks. In *2015 IEEE International Conference on Data Science and Advanced Analytics (DSAA)*. IEEE, oct 2015.

[233] Y. Kukita, J. Uchida, S. Oba, K. Nishino, T. Kumagai, K. Taniguchi, T. Okuyama, F. Imamura, and K. Kato. Quantitative Identification of Mutant Alleles Derived from Lung Cancer in Plasma Cell-Free DNA via Anomaly Detection Using Deep Sequencing Data. *PLoS ONE*, 8(11):e81468, 2013.

[234] S. Belciug, A. Salem, F. Gorunescu, and M. Gorunescu. Clustering-based approach for detecting breast cancer recurrence. In *2010 10th International Conference on Intelligent Systems Design and Applications*, 2010.

[235] M. Shouman, T. Turner, and R. Stocker. Applying k-Nearest Neighbour in Diagnosing Heart Disease Patients. *International Journal of Information and Education Technology*, pages 220–223, 2012.

[236] D.-Y. Liu, H.-L. Chen, B. Yang, X.-E. Lv, L.-N. Li, and J. Liu. Design of an Enhanced Fuzzy k-nearest Neighbor Classifier Based Computer Aided Diagnostic System for Thyroid Disease. *Journal of Medical Systems*, 36(5):3243–3254, 2011.

[237] M. Balamurugan, A. Nancy, and S. Vijaykumar. Alzheimer's Disease Diagnosis by using Dimensionality Reduction Based on Knn Classifier. *Biomedical and Pharmacology Journal*, 10(4):1823–1830, 2017.

[238] T. R. Gadekallu, N. Khare, S. Bhattacharya, S. Singh, P. K. R. Maddikunta, I.-H. Ra, and M. Alazab. Early Detection of Diabetic Retinopathy Using PCA-Firefly Based Deep Learning Model. *Electronics*, 9(2):274, 2020.

[239] Susmita Ray. A Quick Review of Machine Learning Algorithms. In *2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon)*. IEEE, 2019.

[240] I. Wickramasinghe and H. Kalutarage. Naive Bayes: applications, variations and vulnerabilities: a review of literature with code snippets for implementation. *Soft Computing*, 25(3):2277–2293, 2020.

[241] J. V. Tu. Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes. *Journal of Clinical Epidemiology*, 49(11):1225–1231, 1996.

[242] D. W. Hosmer and S. Lemeshow. *Applied Logistic Regression.* New York: John Wiley & Sons, Inc, 1989.

[243] L. Wasserman. *All of statistics: A concise course in statistical inference*, chapter 13. New York: Springer, 2004.

[244] B. L. Bowerman, R. T. O'Connell, and A. B. Koehler. *Forecasting, Time Series and Regression.* California: Thomson Brooks/Cole, 2015.

[245] P. K. Ozili. The acceptable r-square in empirical modelling for social science research. *Social Research Methodology and Publishing Results*, pages 1–9, 2023.

[246] A. Baratloo, M. Hosseini, A. Negida, and G. El Ashal. Part 1: Simple Definition and Calculation of Accuracy, Sensitivity and Specificity. *Emergency (Tehran)*, 3(2):48–49, 2015.

[247] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[248] D. S. K. Karunasingha. Root mean square error or mean absolute error? Use their ratio as well. *Information Sciences*, 585:609–629, 2022.

[249] B. K. Das and H. S. Dutta. GFNB: Gini index–based fuzzy naive bayes and blast cell segmentation for leukemia detection using multi-cell blood smear images. *Medical & Biological Engineering & Computing*, 58(11):2789–2803, 2020.

[250] M. Schonlau and R. Y. Zou. The random forest algorithm for statistical learning. *The Stata Journal: Promoting communications on statistics and Stata*, 20(1):3–29, 2020.

[251] D. Chicco, M. J. Warrens, and G. Jurman. The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. *PeerJ Computer Science*, 7:e623, 2021.

[252] W. McKinney. Data Structures for Statistical Computing in Python. In *Proceedings of the 9th Python in Science Conference*, pages 56 – 61, 2010.

[253] J. Perktold, S. Seabold, J. Taylor, and statsmodels developers. Statsmodels. Available online at: “`https://www.statsmodels.org/stable/index.html`”. Last Accessed 22 March 2023, 2022.

[254] M. Islam, M. Hasan, X. Wang, H. Germack, and M. Noor-E-Alam. A Systematic Review on Healthcare Analytics: Application and Theoretical Perspective of Data Mining. *Healthcare*, 6(2):54, 2018.

[255] J. Lopes, T. Guimarães, and M. F. Santos. Predictive and Prescriptive Analytics in Healthcare: A Survey. *Procedia Computer Science*, 170:1029–1034, 2020.

[256] K. Lepenioti, A. Bousdekis, D. Apostolou, and G. Mentzas. Prescriptive Analytics: A Survey of Approaches and Methods. In *Business Information Systems Workshops*, pages 449–460. Springer International Publishing, 2019.

[257] F. Maggioni and S. W. Wallace. Analyzing the quality of the expected value solution in stochastic programming. *Annals of Operations Research*, 200(1):37–54, 2010.

[258] J. R. Birge and F. Louveaux. *Introduction to stochastic programming*, chapter 2. New York, NY: Springer New York, 2011.

[259] A. M. Mestre, M. D. Oliveira, and A. P. Barbosa-Póvoa. Location-allocation approaches for hospital network planning under uncertainty. *European Journal of Operational Research*, 240(3):791–806, 2015.

[260] A. Prékopa. *Stochastic Programming*, chapter 12. Dordrecht: Kluwer Academic Publishers, 1995.

[261] A. P. Ruszczyński and A. Shapiro. *Stochastic programming*. Amsterdam: Elsevier, 2003.

[262] NHS. Agenda for change - pay rates. Available online at: "`https://www.healthcareers.nhs.uk/working-health/working-nhs/nhs-pay-and-benefits/agenda-change-pay-rates/agenda-change-pay-rates`". Last Accessed 22 March 2023, 2022.

[263] Wales Audit Office. Informatics systems in NHS Wales. Available Online at: "`https://www.audit.wales/publication/informatics-systems-nhs-wales`". Last Accessed 22 March 2023, 2018.

[264] Organisation for Economic Co-operation and Development. Elderly Population. Available online at: "`https://data.oecd.org/pop/elderly-population.htm`". Accessed 28 April 2020, 2018.

[265] National Health Service. NHS Data Dictionary - Admission Method. Available online at: "`https://www.datadictionary.nhs.uk/attributes/admission_method.html?hl=admission%2Cmethod"`". Accessed 28 July 2023, 2023.

[266] National Health Service. NHS Data Dictionary - Admission Source (Hospital Provider Spell). Available online at: "`https://www.datadictionary.nhs.uk/data_elements/admission_source__hospital_provider_spell_.html?hl=admission%2Csource`". Accessed 28 July 2023, 2023.

[267] National Health Service. NHS Data Dictionary - Main Specialty Code. Available online at: "`https://www.datadictionary.nhs.uk/attributes/main_specialty_code.html`". Accessed 28 July 2023, 2023.

[268] R. E. G. Upshur, R. Moineddin, E. Crighton, L. Kiefer, and M. Mamdani. Simplicity within complexity: Seasonality and predictability of hospital admissions in the province of Ontario 1988–2001, a population-based analysis. *BMC Health Services Research*, 5(1):13, 2005.

[269] P. Venkatesan. The changing demographics of COVID-19. *The Lancet Respiratory Medicine*, 8(12):e95, 2020.

[270] S. K. Pillai, M. M. Raghuwanshi, and M. Gaikwad. Hyperparameter Tuning and Optimization in Machine Learning for Species Identification System. In *Proceedings of International Conference on IoT Inclusive Life (ICIIL 2019), NITTTR Chandigarh, India*, pages 235–241. 2020.

[271] P. R. Harper. A review and comparison of classification algorithms for medical decision making. *Health Policy*, 71(3):315–331, 2005.

[272] P. R. Harper. Combining data mining tools with health care models for improved understanding of health processes and resource utilisation. *Clinical and Investigative Medicine*, 28(6):338–341, 2005.

[273] StatsWales. NHS beds by organisation and year, 2009-10 onwards. Available online at: "`https://statswales.gov.wales/Catalogue/Health-and-Social-Care/NHS-Hospital-Activity/NHS-Beds/nhsbeds-by-organisation-site`". Last Accessed 22 March 2023, 2021.

[274] A. Street, N. Gutacker, C. Bojke, N. Devlin, and S. Daidone. Variations in outcome and costs among NHS providers for common surgical procedures: econometric analyses of routinely collected data. *Health Services and Delivery Research*, 2(1):1–90, 2014.

[275] A. Jones, T. Powell, S. Vougioukalou, M. Lynch, and D. Kelly. Research into nurse staffing levels in Wales. Technical Report 27/2015, Cardiff University, Wales, 2015.

[276] University Hospital Southampton NHS Foundation Trust. Ward Staffing Nursing Establishment Review August 2022 – October 2022. Available online at: "`https://www.uhs.nhs.uk/Media/UHS-website-2019/Docs/About-the-trust/performance/TB-6-monthly-staffing-review-report.pdf`". Last Accessed 1 August 2023, 2022.

[277] S. Kean and M. Mitchell. How do intensive care nurses perceive families in intensive care? Insights from the United Kingdom and Australia. *Journal of Clinical Nursing*, 23(5-6):663–672, 2013.

[278] L. East, K. Knowles, M. Pettman, and L. Fisher. Advanced level nursing in England: organisational challenges and opportunities. *Journal of Nursing Management*, 23(8):1011–1019, 2014.

[279] StatsWales. Nursing, midwifery and health visiting staff, by grade and area of work. Available online at: "`https://statswales.gov.wales/Catalogue/Health-and-Social-Care/NHS-Staff/Non-Medical-Staff/nursingmidwiferyandhealthvisitingstaff-by-grade-areaofwork-year`". Last Accessed 22 March 2023, 2022.

[280] T. M. Lal, T. Roh, and T. Huschka. Simulation based optimization: Applications in healthcare. In *2015 Winter Simulation Conference (WSC)*, 2015.

[281] A. Respicio, M. Moz, M. V. Pato, R. Somensi, and C. D. Flores. A computational application for multi-skill nurse staffing in hospital units. *BMC Medical Informatics and Decision Making*, 18(1), 2018.

[282] D. Ranum, B. Miller, J. Zelle, and M. Guzdial. Successful approaches to teaching introductory computer science courses with python. In *Proceedings of the 37th SIGCSE technical symposium on Computer science education*, 2006.

[283] COIN-OR Foundation. OPEN SOURCE FOR THE OPERATIONS RESEARCH COMMUNITY. Available online at: "`https://www.coin-or.org/`". Last Accessed 22 March 2023, 2016.

[284] R. Lougee-Heimer. The Common Optimization INterface for Operations Research: Promoting open-source software in the operations research community. *IBM Journal of Research and Development*, 47(1):57 – 66, 2003.

[285] IBM. IBM CPLEX Optimizer. Available online at "https://www.ibm.com/uk-en/analytics/cplex-optimizer". Last Accessed 22 March 2023.

[286] Gurobi Optimization. Gurobi Optimizer: The world's Fastest Solver. Available online at: "https://www.gurobi.com/solutions/gurobi-optimizer/?campaignid=2027425879&adgroupid=77414946451&creative=602222804300&keyword=gurobi&matchtype=e&gclid=CjwKCAiAxP2eBhBiEiwA5puhNfYpFYlEtlzvzmBPHQdeOHshJHby_gn70yXeKeoeAc1g9rgSbV0ZkBoC1gkQAvD_BwE". Last Accessed 22 March 2023.

[287] E. Williams. Bed and staff capacities planning models. Available online at: "https://github.com/Williamsem/Bed-and-Staff-Capacities-Planning-Models"., 2023.

[288] NHS Wales. The Grange University Hospital. Available online at: "https://abuhb.nhs.wales/healthcare-services/the-grange-university-hospital/". Last Accessed 22 March 2023.

[289] Welsh Government. Statistics Wales Demography newsletter: August 2022. Available online at: "https://www.gov.wales/statistics-wales-demography-newsletter-august-2022-html". Last Accessed 22 March 2023, 2022.

[290] Auditor General for Wales. Tackling the Planned Care Backlog in Wales. Available online at: "https://audit.wales/sites/default/files/publications/Tackling_the_Planned_Care_Backlog_in_Wales.pdf". Last Accessed 22 March 2023, 2022.

[291] R. Luben, S. Hayat, A. Khawaja, N. Wareham, P. P. Pharoah, and K.-T. Khaw. Residential area deprivation and risk of subsequent hospital admission in a British population: the EPIC-Norfolk cohort. *BMJ Open*, 9(12), 2019.

[292] E. Devereux. Record number of nurses leaving the NHS in England. Available online at: "https://www.nursingtimes.net/news/workforce/record-number-of-nurses-leaving-the-nhs-in-england-30-09-2022/". Last Accessed 22 March 2023, 2022.

[293] Royal College of Nursing. NHS pay dispute: new February strike dates announced for England and Wales. Available online at: "https://www.rcn.org.uk/news-and-events/news/uk-nhs-pay-dispute-new-february-strike-dates-announced-for-england-and-wales-160123". Last Accessed 22 March 2023, 2023.

[294] Welsh Government. Sickness absence in the NHS: April to June 2022. Available online at: "https://www.gov.wales/sickness-absence-nhs-april-june-2022-html#:~:text=Long%20term%20trend,months%20ending%2030%20June%202022.". Last Accessed 22 March 2023, 2022.

[295] Calderdale and Huddersfield NHS Foundation Trust. There's no place like home for a quicker recovery and to help ease the pressure on your local hospital. Available online at: "`https://www.cht.nhs.uk/news/theres-no-place-like-home-for-a-quicker-recovery-and-to-help-ease-the-pressure-on-your-local-hospital`". Last Accessed 22 March 2023, 2022.

[296] J. Trueland. What are virtual wards – and how do they work? *Primary Health Care*, 33(1):18–19, 2023.

[297] Health Innovation Network South London. Rapid Evaluation of Croydon-Virtual Ward. Available online at: "`https://healthinnovationnetwork.com/wp-content/uploads/2022/01/Croydon-VW-Evaluation-Report-to-NHSX-v10.pdf`". Last Accessed 22 March 2023, 2021.

[298] British Medical Association. An NHS under pressure. Available online at: "`https://www.bma.org.uk/advice-and-support/nhs-delivery-and-workforce/pressures/an-nhs-under-pressure#:~:text=The%20NHS%20is%20experiencing%20some,pressure%20points%20in%20the%20NHS`". Last Accessed 22 March 2023, 2022.

[299] N. D. Clement, C. E. H. Scott, J. R. D. Murray, C. R. Howie, D. J. Deehan, and IMPACT-Restart Collaboration. The number of patients "worse than death"' while waiting for a hip or knee arthroplasty has nearly doubled during the COVID-19 pandemic. *The Bone & Joint Journal*, 103-B(4):672–680, 2021.

[300] M. R. Robling, R. M. Pill, K. Hood, and C. C. Butler. Time to talk? patient experiences of waiting for clinical management of knee injuries. *Quality and Safety in Health Care*, 18(2):141–146, 2009.

[301] S. Mitchell, A. Kean, A. Mason, M. O'Sullivan, A. Phillips, and F. Peschiera. Optimization with PuLP. Available online at: "`https://coin-or.github.io/pulp/`". Last Accessed 22 March 2023, 2019.

[302] The Health Foundation. ABCi Mathematical Modelling and Analytics Academy. Available online at: "`https://www.health.org.uk/improvement-projects/abci-mathematical-modelling-and-analytics-academy`". Accessed 5 August 2023, 2023.

[303] D. Gartner, I. Spernaes, T. J. England, D. A. Behrens, J. Buchanan, and P. R. Harper. Evaluation of Participants' Reaction and Learning in a Taught Analytics and Modelling Academy Program in U.K.'s National Health Service. In *2022 IEEE 10th International Conference on Healthcare Informatics (ICHI)*. IEEE, 2022.