



A novel healthcare resource allocation decision support tool: A forecasting-simulation-optimization approach

Muhammed Ordu, Eren Demir, Chris Tofallis & Murat M. Gunal

To cite this article: Muhammed Ordu, Eren Demir, Chris Tofallis & Murat M. Gunal (2021) A novel healthcare resource allocation decision support tool: A forecasting-simulation-optimization approach, Journal of the Operational Research Society, 72:3, 485-500, DOI: [10.1080/01605682.2019.1700186](https://doi.org/10.1080/01605682.2019.1700186)

To link to this article: <https://doi.org/10.1080/01605682.2019.1700186>



Published online: 03 Feb 2020.



Submit your article to this journal [↗](#)



Article views: 601



View related articles [↗](#)







View Crossmark data [↗](#)



Citing articles: 2 View citing articles [↗](#)

A novel healthcare resource allocation decision support tool: A forecasting-simulation-optimization approach

Muhammed Ordu^a , Eren Demir^b , Chris Tofallis^b  and Murat M. Gunal^c 

^aOsmaniye Korkut Ata University, Osmaniye, Turkey; ^bUniversity of Hertfordshire, Hertfordshire, UK; ^cNational Defence University, Istanbul, Turkey

ABSTRACT

The increasing pressures on the healthcare system in the UK are well documented. The solution lies in making best use of existing resources (e.g. beds), as additional funding is not available. Increasing demand and capacity shortages are experienced across all specialties and services in hospitals. Modelling at this level of detail is a necessity, as all the services are interconnected, and cannot be assumed to be independent of each other. Our review of the literature revealed two facts; First an entire hospital model is rare, and second, use of multiple OR techniques are applied more frequently in recent years. Hybrid models which combine forecasting, simulation and optimization are becoming more popular. We developed a model that linked each and every service and specialty including A&E, and outpatient and inpatient services, with the aim of, (1) forecasting demand for all the specialties, (2) capturing all the uncertainties of patient pathway within a hospital setting using discrete event simulation, and (3) developing a linear optimization model to estimate the required bed capacity and staff needs of a mid-size hospital in England (using essential outputs from simulation). These results will bring a different perspective to key decision makers with a decision support tool for short and long term strategic planning to make rational and realistic plans, and highlight the benefits of hybrid models.

ARTICLE HISTORY

Received 25 December 2018
Accepted 17 November 2019

KEYWORDS

Healthcare; decision support system; forecasting; discrete event simulation; integer linear programming

1. Introduction

Demand at National Health Service (NHS) hospitals in England has been increasing significantly over the past decade. The estimates show that there has been a 26% and 32% increase in accident & emergency (A&E) and inpatient hospital admissions from 2010/11 to 2017/18, respectively (National Health Services England, 2018a, 2018b). The increasing demand for services is closely linked to worsening prevailing conditions and an expanding elderly population, that often has multiple complex conditions, (such as diabetes and dementia), and which forms the highest demand for beds (The King's Fund, 2012). Advancements in technology and medicine have led to improvements in healthcare, greatly reducing length of stays in hospital and increasing the number of day-cases (or outpatient); however hospital beds remain fundamental resources for all health systems.

Despite a sharp growth in demand, the number of beds has continued to decline. In 2000 there were an average of 3.8 beds per 1,000 people, whereas this had dropped to 2.4 beds by 2015. Between 2006/07 and 2015/16 the number of overnight beds has decreased by over a fifth. As a result, the average bed occupancy rates have increased over time,

with rates for general and acute wards, and mental health, now peaking at over 91% (BMA, 2017). Hospitals are expected to aim for an 85% bed occupancy rate, whereas they are increasingly operating at very high levels of occupancy, particularly during the winter period.

The implications of high bed occupancy rates are widespread, and include, (1) it creates a backlog in emergency departments (Nuffield Trust, 2016), (2) patients can be placed on clinically inappropriate wards, which may affect the patient experience and the quality of care they receive (Goulding, Adamson, Watt, & Wright, 2015), and (3) evidence suggests that high occupancy rates increases the rate of hospital acquired infections, which may lead to temporary closure of beds or wards (Kaier, Mutters, & Frank, 2012).

Due to severe budget cuts in the NHS, hospitals do not have the necessary funding to increase capacity, either in the form of beds or staff. Therefore, hospital management needs to find efficient and effective ways of utilising existing resources. This may mean the management doing things differently, a shift from the conventional decision-making process to a more evidence-based approach (a behavioural change).

A hospital is a complex system made up of 25 or more specialties providing treatment within inpatient, outpatient and A&E services. There are numerous departments and wards within each specialty, with staff including consultants, nurses, healthcare assistances, and technicians. Therefore, determining the most effective use of resources (predominantly beds, consultants and nurses) is a major challenge.

The literature around developing models for healthcare providers is rich and vast. Many simulation-optimisation methods have been developed with the aim of determining optimal solutions for their decision variables (e.g. number of operating rooms and beds or staffing cost). Previous and current models in the literature typically maximized the number of admissions and financial outputs, or minimized length of stay, waiting time and costs in healthcare settings. In the majority of instances these models have focused on modelling a service, department or a specialty, however no models have tackled current and future bed occupancy (and other key metrics of interest) at the entire hospital level (details in the literature review below). A model for a single service (or a few) would not be adequate to determine the required capacity for all specialties within a hospital.

A comprehensive entire hospital modelling framework is necessary that combines all the specialties and services within a single decision support system (DSS). Such an integrated DSS should be able to: (1) forecast demand for all specialties within inpatient, outpatient and A&E, (2) capture the entire hospital patient pathway at a sufficient level of detail, and (3) optimise the required bed capacity and the required number of consultants and nurses.

Such a DSS is able to answer many key questions beyond capacity requirements. For example, a hospital may experience a sharp increase in activity. The forecasts will generate the expected activity to be integrated into the simulation model, whereas the simulation will capture all the uncertainties around the dynamics of the hospital, ranging from time related activities (e.g. length of stay, waiting times, and treatment duration) to hospital finances (revenue, cost and surplus), with the aim of testing a wide range of scenarios around impact of change. The simulation has limitations around establishing the optimal capacity requirements. This is where the optimisation becomes a valuable tool to estimate the exact bed requirements (along with consultant and nurse hours) subject to constraints (e.g. targeted bed occupancy rate).

The merger of these techniques, optimisation and simulation, creates more power in decision making. Optimisation techniques, such as mathematical

programming and heuristic algorithms, are able to provide an exact configuration for the better, and simulation techniques, such as Discrete Event Simulation (DES) and System Dynamics (SD), are able to tell possible outcomes of a scenario. Hybrid modelling approaches are becoming more popular in today's complex decision-making environment.

The core objectives of this study are as follows:

1. Develop an innovative approach combining DES and forecasting demand and capacity in healthcare settings. To our knowledge, the literature does not have an extensive study that forecasts demand for all types of attendances/admissions of each specialty which then integrates these demand inputs within an entire generic hospital simulation model.
2. Develop a linear optimisation algorithm to determine the required bed capacity and staff requirements to meet the needs of local populations. A number of essential outputs from the simulation model will be fed into the optimisation model.
3. Investigate the combined power of forecasting, simulation, and optimisation techniques in the hospital performance modelling domain. Loose or hard-coupled modelling techniques will bring a novel modelling approach.

The remaining sections of the paper are organized as follows: [Section 2](#) gives a review of the literature regarding simulation modelling and optimisation methods applied to healthcare settings. [Section 3](#) illustrates the proposed hybrid framework. [Section 4](#) and [5](#) presents the case study and results, respectively, and [Section 6](#) concludes the study.

2. Literature review

In some studies, optimization and simulation methods have been combined in solving capacity problems of healthcare services, where effective solutions have been investigated (see [Table 1](#)).

Kokangul (2008), Oddoye, Jones, Tamiz, and Schmidt (2009), Zhang, Puterman, Nelson, and Atkins (2012), Ma and Demeulemeester (2013) and Holm, Luras, and Dahl (2013) were interested in bed capacity problems, whereas Wang, Guinet, and Besombes (2009) conducted a study related to operating room capacity. Kokangul (2008) determined a maximum patient arrival value by means of simulation and assumed this value as required maximum bed capacity in paediatric intensive care. The simulation model is re-run using the maximum bed capacity and it is tested that neither rejection nor transfer is needed for any patient. The optimum bed

Table 1. Detailed information about studies related to simulation-optimization modelling in healthcare settings. ABS: Agent based simulation, BDU: Bed daily utilization, DES: Discrete event simulation, GP: Goal programming, ILP: Integer linear programming, INLP: Integer nonlinear programming, IP: Integer programming, Max: Maximization, Min: Minimization, MIP: Mixed integer programming, NG: Not given, NSGA II: Non-dominated sorting genetic algorithm II, O: Optimization, OR: Operating room, PIOA: Prevalence and incidence optimization algorithm, S: Simulation, SSA: Simultaneous search algorithm, SBSA: Sequential bisection search algorithm, VBA: Visual basic for application.

Authors and Year	Methods	Objective Function	Main constraints	Inputs (from to)
Kokangul (2008)	S: DES O: INLP	Max: Number of admissions	Service level, Occupancy level	S to O
Wang et al. (2009)	S: DES O: MIP	Min: Length of stay	OR capacity, emergency demand to be met, postponed demand, unmet demand	O to S
Oddoye et al. (2009)	S: DES O: GP	Min: Deviations from queues, waiting times, beds	Queue length, waiting time, number of beds	S to O
Ahmed and Alkhamis (2009)	S: DES O: IP	Max: Throughput Min: Cost	Budget, average waiting time, staffing level	S to O
Lamiri et al. (2009)	S: MCS O: MIP	Min: Overtime cost, Patients' related costs	Assignment of elective case once	O to S
Cabrera et al. (2011)	S: ABS O: O	Min: Waiting time	Cost of staff configuration	S to O
Zhang et al. (2012)	S: DES O: SSA, SBSA	-	-	S to O
Cabrera et al. (2012)	S: ABS O: O	Min: Length of stay	Cost of staff configuration	S to O
Ma and Demeulemeester (2013)	S: DES O: ILP, MIP	Max: Total financial contributions	Bed shortage, BDU, OR blocks, total surgery time, admission volume bound	O to S
Holm et al. (2013)	S: DES O: PIOA	-	-	S to O
Cappanera, Visintin, and Banditori (2014)	S: DES O: MIP	1st Min: Max ORs and BDU 2nd Min: Gaps between max and min ORs and BDU 3rd Min: Sum of quadratic positive deviations of ORs and BDU from a fixed threshold	Daily utilization of ORs, maximum BDU	O to S
Ghanes et al. (2015)	S: DES O: O	Min: Length of stay	Staffing budget, Door-to-door time	S to O
Saadouli, Jerbi, Dammak, Masmoudi, and Bouaziz (2015)	S: DES O: MIP	Min: Maximum completion time and total waiting time of operations	Completion times, waiting times, assignment of operations,	O to S
Uriarte et al. (2017)	S: DES O: NSGA II	-	-	S to O

capacity is found by taking into account parameter values (i.e. maximum bed capacity) using a mathematical model which targets maximum patient admission and has service and occupancy levels constraints. Oddoye et al. (2009) simulates a medical assessment unit in order to get five performance criteria values coupled with lengths of queues, numbers of beds and waiting times. The researchers address the balance between resources of the unit by using a goal programming approach which has the results from the simulation as input. Wang et al. (2009) used mixed integer programming in distributing operating room hours to departments, and simulate the room in order to specify the lengths of stay. In their mathematical model there are a number of assumptions, such as fixed patient demands for every week, fixed numbers of staff, and considering only weekdays although patients can also stay during a weekend.

On the other hand, Ma and Demeulemeester (2013) aim to effectively allocate the available beds. To do this the researchers conduct a study which consists of three steps. In the first step of the study an integer linear programming model is developed to efficiently use resources and maximize the financial situation of the hospital. Moreover, constraints such as bed capacity and bed utilization are taken into account. In the second step, the objective function is to allocate the beds effectively, for which a

mixed integer linear programming model is developed. In the final step, performance measurement is carried out by using the results obtained from the mathematical models through a DES model.

Simulation-optimization studies have been utilised in resource optimization as well as bed capacity problems. In this context human resources and hospital rooms have been evaluated in the scope of these problems. Ahmed and Alkhamis (2009), Cabrera, Taboada, Iglesias, Epelde, and Luque (2011), Cabrera, Taboada, Iglesias, Epelde, and Luque (2012), Ghanes et al. (2015) and Uriarte, Zuniga, Moris, and Ng (2017) investigated human resource needs of healthcare services. Ahmed and Alkhamis (2009) determined the staffing level from an optimization model by considering budget constraint, patient arrivals and waiting times. These optimum staff numbers are used as inputs in the simulation model and thereby measure the system performance.

Scheduling problems were solved using simulation-optimization approach by Lamiri, Grimaud, and Xie (2009), Cappanera et al. (2014), Saadouli et al. (2015). Lamiri et al. (2009) developed a simulation-optimization method to plan elective surgery cases since operating rooms are used by both elective and non-elective patients. A mixed integer programming model was developed with objective functions minimizing overtime costs and patient

related costs. A Monte Carlo simulation was developed to find a reasonable solution for elective surgery cases and compared the developed simulation-optimization method with a number of heuristic and meta-heuristic methods (i.e. sequential improvement, local optimization, and taboo search heuristics). Cappanera et al. (2014) compared different scheduling policies by developing a mixed integer programming model and then, combined this with a DES model. They tested the schedules generated by this optimization method in the stochastic setting of the hospital provided by the simulation model with the parameters (i.e. surgical time and length of stay) which have the features of variability.

Although our literature review and the summary in Table 1, includes the intersection of simulation and optimisation in hospitals, a few more words about simulation and hospitals is worth mentioning. First, whole hospital simulation models are rare in the literature and one of the pioneering studies were conducted by Günel and Pidd (2011). In their paper, they discussed the challenges of whole hospital simulation studies. Furthermore, Günel (2012) sets the scene for building hospital simulation models by using different simulation techniques. The use of simulation in healthcare, however, is not limited with optimisation or capacity. Health economics also benefits from simulation (Marshall et al., 2015) in terms of understanding the relationship between cost and benefit. In almost all of these studies, a natural conclusion can be drawn: to tackle the complexity in healthcare it is better to use multiple techniques. Hybrid modelling approaches, therefore, are emerging in healthcare. For example, Mielczarek and Zabawa (2016) used forecasting techniques to feed a DES and SD simulation models and studied a health condition in the country. Likewise, Harper, Mustafee, and Feeney (2017) presents a hybrid modelling methodology which includes forecasting and DES and studied endoscopy services in a hospital.

3. The proposed hybrid framework: FSO approach

In this study, the following three methodologies are combined to develop a forecasting-simulation-optimization (FSO) approach (see Figure 1) for the purpose of optimizing the level of resources of an NHS Trust: forecasting, DES and integer linear programming. The first component of the proposed methodology is to establish a decision support system (DSS) for comparative forecasting to select the best forecasting method to predict demand for the entire hospital, including inpatient specialties, outpatient specialties and A&E. The second component of the proposed methodology is to develop a generic

hospital simulation model that integrates all specialties (i.e. A&E, outpatient, and inpatient services) and interactions to capture the stochastic behaviour of the hospital. Finally, the third component of the proposed methodology is to develop an integer linear programming model to reallocate the available number of beds and optimize the staffing levels of inpatient services of a mid-size hospital. Thus, the FSO approach is able to capture the realistic (or stochastic) behaviour of the hospital. A number of outputs generated from the simulation model are then fed into the optimization model. The hospital is modelled using an integer linear programme and the desired decision variables are obtained by maximizing the number of discharged patients from each specialty under a number of constraints.

The main contribution of this study is to develop a hybrid framework integrating three distinct methodologies. The forecasting techniques used in our hybrid framework enable conversion from generic data sources to useful information. In today's world, and particularly on the verge of a new industrial revolution, so called Industry 4.0, we are rich in terms of data collection, however intelligence is required to convert data into information. Our first contribution to knowledge in this study, and in the proposed framework, is to give some guidance on how normally collected hospital data can be analysed to extract useful information. The "useful information" is not only the findings of the past but is also the key to the predictions of the future. Simulating hospital operations enables predictions by way of running "what-if" scenarios. Another benefit of simulation is to explore variation in patient demand and services provided in various service points in hospitals. Simulation is a realistic way to convince decision makers. Our second contribution is the scale of the simulation model we developed, as it depicts complex hospital care processes in the A&E, inpatient and outpatient facilities. Whole hospital level simulation models are rare in the literature. A simulation model helps us find the effects of variation and what-if scenarios, however it does not tell us the optimum configuration of hospital resources. An optimisation model is attached to our framework as it finds the optimum number of beds, doctors, and nurses for specialties. This model uses some of the outputs generated by the simulation model. Co-working of the three models, forecasting, simulation, and optimisation, is a novel approach and contributes to the healthcare domain.

Activity related data at PAH (i.e. the number of admissions to inpatient, outpatient and A&E services) is derived from the national Hospital Episode Statistics (HES) dataset after an extensive data preparation process. HES has all the information about

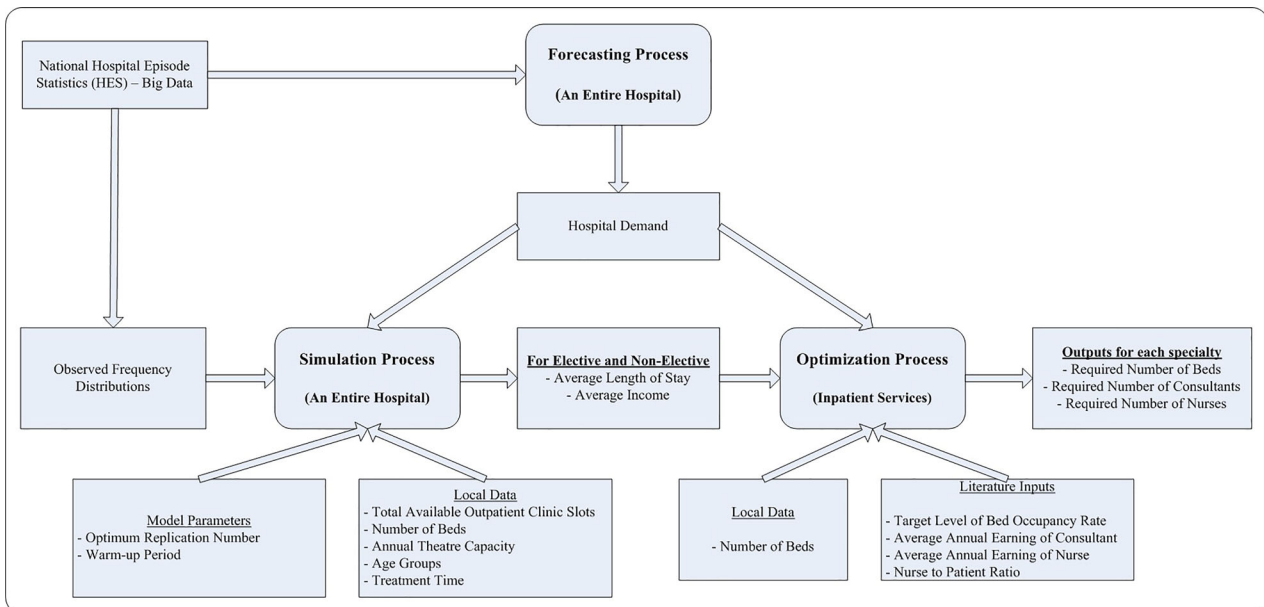


Figure 1. The structure of the FSO approach along with the relationships of inputs-outputs.

patients admitted and treated in NHS hospitals in England, capturing administrative, personal and medical details. PAH is a general hospital north of London (Harlow, Essex) with 557 beds serving a population of 350,000, providing services across 27 distinct specialties (e.g. cardiology, ophthalmology, trauma and orthopaedics, etc.) The extracted data period is from 01/04/10 to 31/03/13 (three financial years). During the data period, PAH had 248,910 A&E arrivals, 996,134 outpatient attendances and 191,462 inpatient admissions across the 27 specialties.

3.1. Demand forecasting

In this study, a decision support system (DSS) is developed to identify better forecasting methods and time periods for each specialty of the hospital. The forecasted demand is used as an input in the simulation and optimization models, instead of using approximate demand, such as Wang, Li, Tussey, & Ross (2012) and Demir, Gunal, & Southern (2017). All required hospital data is derived from the National Hospital Episode Statistics (HES) dataset after an extensive data preparation process. The time series (i.e. daily, weekly and monthly) are established after extracting the required data. The data is divided into two sets: training set (the first 24-months of data) and validation set (the last 12-months of data). The best model parameters for each forecasting method are determined, i.e. ARIMA, stepwise linear regression (SLR), exponential smoothing (ES) and seasonal and trend decomposition using loess forecasting (STLF).

The goodness of fit (for in-sample) and the forecast accuracy (for out-of-sample) are obtained. The forecasting models are assessed according to the

results of the goodness of fit, whereas the best forecasting method and prediction frequency are selected by taking into account the results of the forecast accuracy. Finally, the demands of each specialty are estimated by using the forecasting results which was given in Ordu, Demir, and Tofallis (2019) and these demand inputs are embedded into the generic hospital simulation model.

A total of 760 forecasting models are developed for outpatient services (including first and follow up referrals) and inpatient services (elective and non-elective services) along with the A&E department. The 760 models are made up of the following:

- 19 outpatient specialties × 2 (first and follow up referrals separately) × 3 periods (daily, weekly and monthly) × 4 forecasting methods, which is 456 models for outpatients.
- 16 inpatient specialties (for elective admissions) × 3 periods (daily, weekly and monthly) × 4 forecasting methods, which is 192 models for inpatients.
- 9 inpatient specialties (for non-elective admissions) × 3 periods (daily, weekly and monthly) × 4 forecasting methods, which is 108 models for inpatients.
- A&E is forecasted daily only, thus 1 × 4 forecasting methods.

To give an example, Table 2 shows the forecasting results in terms of both training and validation sets for one inpatient service (the general surgery elective and non-elective specialties).

We have chosen the mean absolute scaled error (MASE) to measure the goodness-of-fit and the forecast accuracy since the MASE can be applied to compare the forecasting models carried out under

Table 2. Forecast accuracy values (mean absolute scaled error) for the general surgery inpatient elective and non-elective speciality. ARIMA: Autoregressive integrated moving average, ES: Exponential smoothing, SLR: Stepwise Linear Regression, STLF: Seasonal and trend decomposition by loess forecasting, TS: Training Set, VS: Validation Set.

Speciality	Forecasting Models	Daily			Weekly			Monthly		
		Parameters	TS	VS	Parameters	TS	VS	Parameters	TS	VS
Inpatient (Elective)	SLR	SLR	0.45	0.53	SLR	0.83	1.10	SLR	0.68	1.43
	ARIMA	(2,1,5)	0.77	0.76	(0,1,1)	0.83	0.81	(1,0,0)	0.95	0.71
	ES	(A,N,N)	0.87	0.76	(M,A _d ,N)	0.80	0.79	(M,A,N)	0.83	1.70
	STLF	(A,A _d ,N)	0.94	0.86	(M,A,N)	0.74	6.99	(M,N,N)	0.87	1.41
Inpatient (Non-elective)	SLR	SLR	0.83	0.88	SLR	1.17	1.32	SLR	0.88	1.71
	ARIMA	(0,1,1)	0.75	0.73	(0,1,1)	0.90	0.90	(1,0,0)	1.02	1.65
	ES	(A,N,N)	0.75	0.83	(A,A _d ,N)	0.93	0.90	(A,N,N)	0.95	0.91
	STLF	(A,N,N)	0.67	0.98	(A,N,N)	0.64	1.45	(M,N,N)	0.96	2.05

Table 3. Best performing Forecasting methods for the inpatient specialities; ARIMA: Autoregressive integrated moving average, ES: Exponential smoothing, SLR: Stepwise linear regression, STLF: Seasonal and trend decomposition by loess forecasting.

Specialities	Elective		Non-elective	
	Forecasting model	Forecasting period	Forecasting model	Forecasting period
General Surgery	SLR	Daily	ARIMA (0,1,1)	Daily
Trauma & Orthopaedics	ARIMA (1–3)	Daily	ARIMA (0,1,1)	Weekly
General Medicine	ES: ETS(M,A,N)	Monthly	STLF: STL + ETS(A,N,N)	Monthly
Cardiology	SLR	Daily	ES: ETS(A,N,N)	Monthly
Paediatrics	ARIMA (0,1,3)	Weekly	ARIMA (1)	Daily
Geriatric Medicine	–	–	ARIMA (0,1,2)	Daily
Obstetrics	–	–	ARIMA (0,1,1)	Daily
Gynaecology	ARIMA (1,0,0)	Monthly	SLR	Monthly
Others	ES: ETS(M,A _d ,N)	Weekly	ARIMA (0,1,1)	Daily

different periods (Hyndman and Koehler, 2006). The MASE divides the mean absolute error of the forecasting method by the mean absolute error of the naïve method (Hyndman and Koehler, 2006). Using MASE, 64 best forecasting models are selected out of 760 models, comprising 38 for outpatient specialties demand, 25 for inpatient specialties, and 1 for A&E. Table 2 illustrates the entire process for the general surgery specialty and the remaining specialties are shown in Table 3.

The best forecasting result for the inpatient elective specialty is daily stepwise linear regression with the lowest MASE value 0.53. In addition, the best forecasting result for the general surgery non-elective inpatient specialty is daily ARIMA (0,1,1) with the lowest MASE value 0.73. The best forecasting method and period for the general surgery inpatient specialty are highlighted in Table 2.

Table 3 illustrates the best forecasting methods and periods for each inpatient specialty. As seen from the table, there is no single outperforming forecasting method, thus the best technique should be determined accordingly (rather than relying on one specific method).

3.2. Generic hospital simulation modelling

The DES model developed here is simulating patient pathways in a general hospital with multiple specialties. Patients compete for scarce hospital resources in their journeys and therefore waiting is likely to

occur. The cascading services and shared resources create a complex structure in hospital services.

We used a variety of inputs in our study, including primary and secondary datasets. The following inputs were collected from the hospital as primary data and verified by a number of experts (i.e. directors, consultants and nurses): A&E treatment time, outpatient consultation time for first and follow up attendances, pre-assessment time for A&E, outpatient and inpatient services, A&E laboratory process for each type of laboratory test, theatre time, and day-case procedures. In addition, several other secondary data sets were provided by the hospital, e.g. number of beds, financial tariffs, outpatient clinic slots for each specialty, theatre capacity for each specialty, and the number of physicians and nurses.

In this study, a DES model is used in a decision support system (DSS) for demand and capacity planning in the hospital. For this, the predicted demand is obtained from forecasting techniques (instead of using presumptive demand) to embed as input in the simulation model. Based on historical data (i.e. HES dataset), transfers from A&E to outpatient and inpatient non-elective specialties is around 5% and 69%, respectively. Outpatient and inpatient non-elective specialties also have transfers from other sources too, for example general practitioners and community dental services. We independently forecasted the number of patients for outpatient first attendances and inpatient non-elective admissions excluding those transfers from

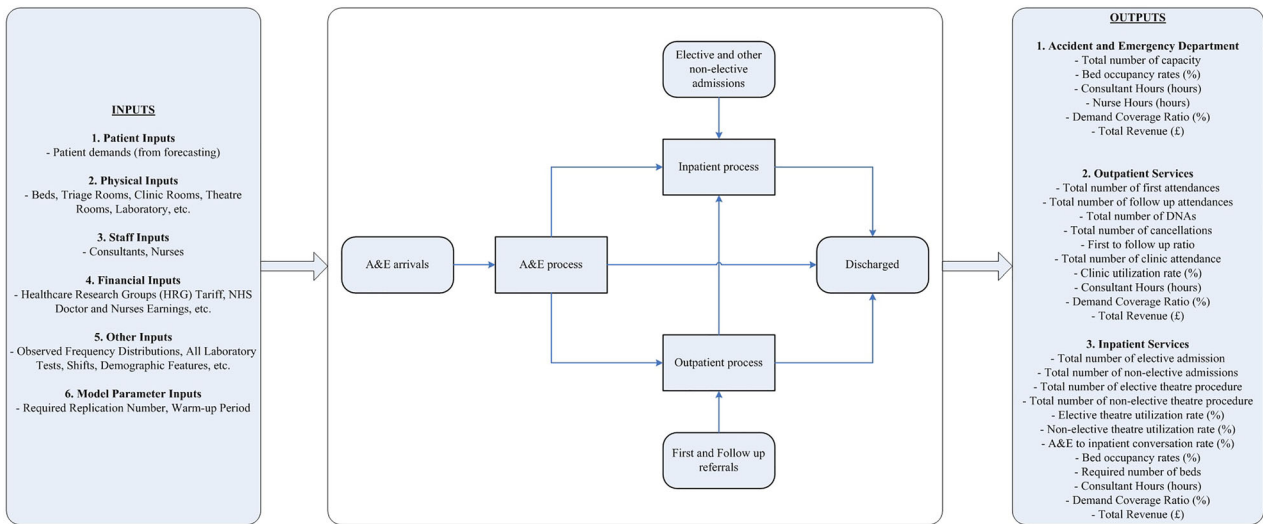


Figure 2. Outline illustration of the whole hospital conceptualized model.

A&E. Our simulation model captures both the patients from A&E (forecasting was independently carried out for the A&E demand) using transfers as stated above and the independent forecasts for the number of attendances/admissions from other sources (forecasting was conducted for the outpatient first attendances and inpatient admissions). In addition, the forecasts of the A&E arrivals and inpatient-elective admissions are completely independent forecasts.

Some of the required data are extracted from the National Hospital Episodes Statistics (HES) dataset and from local hospital data over the study period. The data is used for both demand forecasting and parameter estimation of the statistical distributions for the DES model. These inputs along with model parameters, financial inputs and local data provided by the hospital, are embedded into the generic hospital simulation model. The model then generates current and future levels of key output metrics for each specialty as seen in Figure 2. Simulation outputs for outpatient specialties and A&E are provided as supplementary materials whereas outputs for inpatient services are discussed in Section 4.

There are three main sources of patient arrivals in a hospital. Emergency patients arrive at an A&E department and after treatment some of these patients are admitted to hospital as inpatients. Although most A&E departments are seen isolated from hospitals, they still require hospital resources such as radiology and biochemistry services and are therefore linked with main hospital resources. Furthermore, A&E departments are linked with upstream bed resources, for example a patient who needs to stay in hospital must be accommodated on a hospital ward bed. In fact, some emergency patients might require surgery, and, in this case, they also consume operating theatre capacity, and pre-empt planned surgeries.

Elective patients arrive at hospitals by way of referrals, mostly made by General Practitioners (GP). This type of demand can be controlled in a way since they are mostly non-urgent cases whose admissions can be delayed, or programmed, based on the status of the resources, such as consultants, operating rooms, and elective beds. Elective patients, or inpatient, processes generally require scheduling, admission and bed management. In the model, although these processes are shown as a single box in Figure 2, they are modelled in detail to be able to read the inputs and generate the desired performance outputs.

Outpatients arrive at hospitals on a daily basis and are referred by themselves or by GPs. An outpatient episode might end the same day, by a consultant’s decision, or may need a follow up. The number of follow ups might increase for patients who have chronic diseases. The existence of feedback in outpatients creates complications in the management of outpatient processes. The DES model includes a booking mechanism which mimics the outpatient first and follow up procedures.

Our generic hospital simulation model generates many key outputs as seen in Figure 2, for example, bed occupancy rates, staffing hours, total revenue, clinic utilization, theatre utilization, demand coverage ratio and many other outputs. However, we focussed on capturing the stochastic nature of the hospital into the future. It is important for our study in terms of how long a patient stays in a bed depending on age group, diagnostics and specialties. Similarly, we need to better understand how much revenue the hospital expects in the future. All these aspects are crucial for this study to develop a robust model which plans bed capacity and optimizes staffing levels. All these reasons highlight the importance of capturing the expected stochastic behaviour of the hospital in future. The main output variables

from the simulation model are therefore the average length of stay and average revenue for each elective and non-elective inpatient specialty.

The revenue based on the length of stay for each patient staying in a bed is calculated using the formula we developed, see Eq. (1). A long stay payment for days exceeding the trimpoint is applied by considering the trimpoint determined by the NHS. Trimpoint is a threshold for the length of stay for patients (NHS Digital, 2018). For example, the non-elective long stay trimpoint is 5 days for an HRG Code EB01Z (Non-Interventional Acquired Cardiac Conditions) in the general surgery inpatient specialty and £211 (£243 with the Market Forces Factor (MFF)) per day for exceeding the trimpoint is charged. The non-elective spell tariff is £585 (£675 with the MFF). The hospital reimburses £1161 (MFF × £585 + MFF × 2 days × £211) if a non-elective patient with the HRG code “EB01Z” stays in a bed for 7 days. The Market Forces Factor (MFF) is used as a multiplier in the calculation of the revenue, and is a reflection of service cost, which might depend on the location of each hospital in the country (Department of Health and Social Care, 2014).

$$\begin{aligned} \text{Total Revenue } (\pounds) = & \left(\sum_{i=1}^{s_e} \sum_{j=1}^{e_j} (TE + TAE)_{ij} \times MFF \right. \\ & \left. + \sum_{i=1}^{s_{ne}} \sum_{j=1}^{ne_j} (TNE + TANE)_{ij} \times MFF \right) \end{aligned} \quad (1)$$

where s_e : Total number of inpatient elective specialties; s_{ne} : Total number of inpatient non-elective specialties; e_j : Total number of admissions at inpatient elective specialty j ; ne_j : Total number of admissions at inpatient non-elective specialty j ; TE_{ij} : i tariff at inpatient elective specialty j ; TAE_{ij} : i tariff adjustment at inpatient elective specialty j due to exceeding the trimpoint; TNE_{ij} : i tariff at inpatient non-elective specialty j ; $TANE_{ij}$: i tariff adjustment at inpatient non-elective specialty j due to exceeding the trimpoint; MFF : Market forces factor.

Black-box and white-box validations were used to validate our DES model. We collaborated with several key stakeholders at PAH to help us with the verification (checking for face validity) and validation stages of the simulation model. The model was continually improved according to the feedbacks received. Each model unit was then rigorously tested under extreme conditions, enabling us to pass white-box validation tests. In the final stage of model development the key stakeholders were convinced that the model is appropriate for testing scenarios and interventions of interest.

The next phase of validation is the Black-box approach, where the model is tested to examine if it behaves like the real system by comparing

simulation results with the observed ones. Several key metrics were used for this purpose, including occupancy rates and total revenue. We confirm that the differences between real world observations and the model outputs were within the 95% confidence interval range, which then deemed the model to be validated, hence suitable to evaluate a wide range of alternative policy scenarios.

3.3. Integer linear programming integrated with forecasting and DES

We developed a decision support system (DSS), named as the FSO approach, integrating three distinct methodologies to determine the required level of resources of a mid-size hospital inpatient services. Demand is forecasted for each specialty to be used as input for both the simulation and optimization models. A generic hospital simulation model is developed to capture the future key performance metrics (i.e. average length of stay and revenue) by modelling the stochastic behaviour of the hospital (including patients, human resources, beds, treatment procedures, etc.). An integer linear model is integrated with forecasting and DES to reallocate the existing number of beds and optimize the staffing levels. The structure of the DSS is presented as a flow diagram which shows how three distinct techniques are combined in Figure 1. The number of beds for each inpatient specialty is provided by the hospital. The FSO approach then generates future levels of key output metrics (i.e. the required number of beds, consultants and nurses working with full time equivalent).

3.3.1. Parameters and decision variables

The set of inpatient specialties that serve wards for patients consists of general surgery, trauma & orthopaedics, general medicine, cardiology, paediatrics, geriatric medicine, obstetrics, gynaecology, and “others” which includes all other specialities with less than 1% patients of total patient activity in the hospital.

The parameters of the integer linear programming are defined as follows:

- NDP_s**: Number of discharged patients at specialty s ,
- BOR**: Bed occupancy rate (assumed to be annual bed occupancy rate of the hospital)
- TARGET**: Target level of bed occupancy rate (assumed to be 85% according to the literature),
- NB_s**: The number of available beds at specialty s ,
- BEDS**: The total number of available beds (assumed to be available bed capacity of the hospital),
- NDEP_s**: Number of discharged elective patients at specialty s ,

NAEP_s: Number of forecasted admitted elective patients at specialty *s* (assumed to be forecasted bed demand of the hospital),

NE_s: Number of non-elective patients at specialty *s*,

NONELECTIVE_s: Number of forecasted non-elective patients at specialty *s* (assumed to be forecasted bed demand of the hospital for non-elective patients),

INCOME: Total income from patient care,

COST: Total staffing costs,

NPR_s: Nurse to patient ratio at specialty *s*,

CH_s: Consultant hours at specialty *s* (assumed to be the consultation time served by total number of consultants),

CT_s: Consultation times at specialty *s* (assumed to be the consultation time required by patients).

The bed occupancy rates of hospitals are measured by the NHS Trusts in the UK on a regular basis as a key output metric of their inpatient services. For this, NHS Trusts use Eq. (2), a ratio that divides the number of hospital beds occupied by the total number of available hospital beds in a period (Harper and Shahani, 2002).

$$\text{Bed Occupancy Rate (\%)} = 100 \times \left(\frac{\text{The number of occupied bed days}}{\text{Total number of beds} \times \text{Number of days in the period}} \right) \quad (2)$$

In the UK a full time equivalent (FTE) is 37.5 h per week, which equates to 1950 h in a year (NHS improvement, 2017). Consultant hours are calculated by multiplying annual working time based on 1.0 FTE with the total number of consultants at the related specialty. According to experts, a consultant on average spends around 20 min per day per patient. Consultation times are calculated by multiplying the time for care (i.e. 20 min) with total length of stay (as specified in Eq. (3)).

$$\text{Consultation time}_j = T \times \left(\sum_{i=1}^{e_j} (\text{NAEP}_{ij} \times \text{EALoS}_{ij}) + \sum_{i=1}^{ne_j} (\text{NANEP}_{ij} \times \text{NEALoS}_{ij}) \right) \quad (3)$$

where *e_j*: Total number of admissions at inpatient elective specialty *j*; *EALoS_{ij}*: Length of stay of elective patient *i* at specialty *j*; *NAEP_{ij}*: Number of admitted elective patient *i* at specialty *j*; *NANEP_{ij}*: Number of non-elective patient *i* at specialty *j*; *ne_j*: Total number of admissions at inpatient non-elective specialty *j*; *NEALoS_{ij}*: Length of stay of non-elective patient *i* at specialty *j*; *T*: Average time for patient care by consultant.

According to the report published by the National Institute for Health and Care Excellence (NICE) (2014), a nurse should not be liable for

more than 8 patients, otherwise the risk of harm is increased for nurses (NICE, 2014). The quality of service significantly decreases when a nurse is responsible for more than 8 patients (Griffiths, Dall’Ora, & Ball, 2017). Therefore, nurse to patient ratio is considered as a nurse care with a maximum of 8 patients per shift in a day.

Total revenue from an inpatient specialty is calculated by multiplying the number of patients with the average revenue per patient (an output from the simulation model). The formula is shown in Eq. (4).

$$\text{Total revenue} = \sum_{j=1}^{s_e} (\text{EAR}_j \times \text{ENP}_j) + \sum_{j=1}^{s_{ne}} (\text{NEAR}_j \times \text{NENP}_j) \quad (4)$$

where *EAR_j*: average revenue at elective specialty *j*; *ENP_j*: Number of patients at elective specialty *j*; *NEAR_j*: average revenue at non-elective specialty *j*; *NENP_j*: number of patients at non-elective specialty *j*; *s_e*: Total number of inpatient elective specialties; *s_{ne}*: Total number of inpatient non-elective specialties.

Total costs are the staffing costs of all related inpatient specialties as shown in Eq. (5). The average annual earnings of consultants and nurses are taken into account in the calculation of staffing costs.

$$\text{Total cost} = \text{AEC} \times \sum_{j=1}^s \text{NC}_j + \text{AEN} \times \sum_{j=1}^s \text{NN}_j \quad (5)$$

where *AEC*: Annual average earnings of consultants; *AEN*: Annual average earnings of nurses; *NC_j*: number of consultants at specialty *j*; *NN_j*: Number of nurses at specialty *j*; *s*: Total number of inpatient specialties.

3.3.2. Decision variables

The decision variables of the integer linear programming are defined as follows.

DE_s: Number of discharged elective patients at specialty *s*,

DNE_s: Number of discharged non-elective patients at specialty *s*,

NB_s: Required number of beds at specialty *s*,

NN_s: Required number of nurses at specialty *s*,

NC_s: Required number of consultants at specialty *s*.

3.3.3. Objective function and constraints

The optimisation model is as follows.

$$\text{Max} \sum_{i=1}^s \text{NDP}_s, \quad \forall s \in S \quad (6)$$

Subject to:

$$\text{BOR}_s \leq \text{TARGET}, \quad \forall s \in S \quad (7)$$

Table 4. Input values of the FSO approach. ENB: Existing number of beds, ALoS: Average length of stay, NAP: Number of admitted patients, AR: Average revenue.

Code	Specialty Name	ENB	Elective			Non-elective		
			ALoS (day)	NAP	AR (£)	ALoS (day)	NAP	AR (£)
1	General surgery	85	1.04	3468	£1282	4.09	3660	£2080
2	Trauma & Orthopaedics	59	1.67	3276	£3034	5.90	1536	£3459
3	General Medicine	88	0.33	1469	£1486	3.90	9004	£2110
4	Cardiology	25	0.73	972	£2117	7.15	1224	£2383
5	Paediatrics	16	0.96	264	£1568	1.42	2196	£1038
6	Gynaecology	41	0.60	1553	£1224	1.64	2147	£1386
7	Others	91	0.77	372	£784	3.99	732	£1629
8	Geriatric Medicine	111	–	–	–	6.03	7692	£2280
9	Obstetrics	41	–	–	–	1.94	7320	£1636

$$\sum_{s \in S} NB_s \leq BEDS \quad (8)$$

$$NDEP_s \leq NAEP_s, \forall s \in S \quad (9)$$

$$NE_s = NONELECTIVE_s, \forall s \in S \quad (10)$$

$$\sum REVENUE \geq \sum COST \quad (11)$$

$$NPR_s \leq 8, \forall s \in S \quad (12)$$

$$CH_s \geq CT_s, \forall s \in S \quad (13)$$

$$DE_s, DNE_s, NB_s, N_s, C_s \in Z^+, \forall s \in S \quad (14)$$

The objective function (6) maximizes the number of discharged patients (throughput). Constraint (7) ensures that bed occupancy rate of each specialty does not exceed the target level of 85%. Constraint (8) allocates the bed capacity (total number of available beds). Constraint (9) ensures the number of discharged elective patients does not exceed the number of admitted elective patients. Constraint (10) ensures that each non-elective patient is admitted to the hospital and stays in a bed. Constraint (11) ensures that the total income from patient care must be more than or equal to the total costs. Constraint (12) indicates that each nurse must be responsible for no more than eight patients. Constraint (13) ensures that the total consultation time served by total number of consultants must be more than the total consultation time needed by patients. Constraint (14) denotes that all decision variables must be positive integers.

4. A case study: reallocating number of beds and optimizing staffing levels

We applied the FSO approach to PAH and provided inputs from four types of resources: Local data, forecasting, simulation and the literature (see Table 4 for a breakdown of beds by specialty). Financial inputs, target level of bed occupancy rate and nurse to patient ratio are obtained from the literature. The bed occupancy rate in the UK hospitals must not exceed the target level of 85% (Royal College of Physicians, 2015). In addition, the financial inputs (e.g. the average annual earnings of consultants and nurses) are obtained from the NHS Digital (2014). Another important input from the literature is the

nurse to patient ratio, thus a nurse to patient ratio of eight is included in the model as a constraint.

The demand for each specialty is estimated using the comparative forecasting methods. Average length of stay and average revenue are inputs generated by the generic hospital simulation model. In the simulation model, a diagnostics code is assigned to each patient, derived from the HES dataset using the observed frequency distributions. According to the diagnostic codes based on type of specialty and age groups, we considered different HRG tariffs depending on length of stay of patients. Thus, the simulation model calculates the average revenue for each specialty. The integer linear model is then embedded by using these stochastic inputs (i.e. average length of stay and revenue) to reflect the reality of the hospital. Values of all input parameters are given in Table 4.

We developed a hybrid framework integrating three distinct methodologies to determine the required level of resources of a mid-size hospital inpatient services in England. Demand is forecasted for each specialty to be used as input for both the simulation and optimization models. A generic hospital simulation model is developed to capture the key performance metrics (i.e. average length of stay and revenue) by tackling stochastic behaviour of the hospital (including patients, human resources, beds, treatment procedures, etc.). An integer linear model is integrated with forecasting and DES to reallocate the existing number of beds and optimize the staffing levels. The developed integer linear model is solved using LINGO 17.0 software.

4.1. Reallocated number of beds

Table 5 illustrates the number of discharged elective and non-elective patients, bed occupancy rate (%), the reallocated number of beds and the number of human resources required to meet all demands for each specialty. Bed occupancy rates of all the inpatient specialties are less than the target level of 85%. A total of 486 beds is adequate for the hospital to ensure the hospital is within the recommended bed occupancy rate.

Table 5. The results of the FSO approach for the base model. NDEP: Number of discharged elective patients, NDNEP: Number of discharged non-elective patients, BOR: Bed occupancy rate, NB: Number of beds, NC: Number of consultants, NN: Number of nurses.

Code	Name	NDEP	NDNEP	BOR (%)	NB	NC	NN
1	General surgery	3468	3660	84.82	60	4	20
2	Trauma & Orthopaedics	3276	1536	84.72	47	3	15
3	General Medicine	1469	9004	84.81	115	7	37
4	Cardiology	972	1224	83.62	31	2	10
5	Paediatrics	264	2196	83.98	11	1	4
6	Gynaecology	1553	2147	81.33	15	1	5
7	Others	372	732	79.88	11	1	4
8	Geriatric Medicine	-	7692	84.72	150	8	48
9	Obstetrics	-	7320	84.58	46	3	15
Total		11374	35511	-	486	30	158

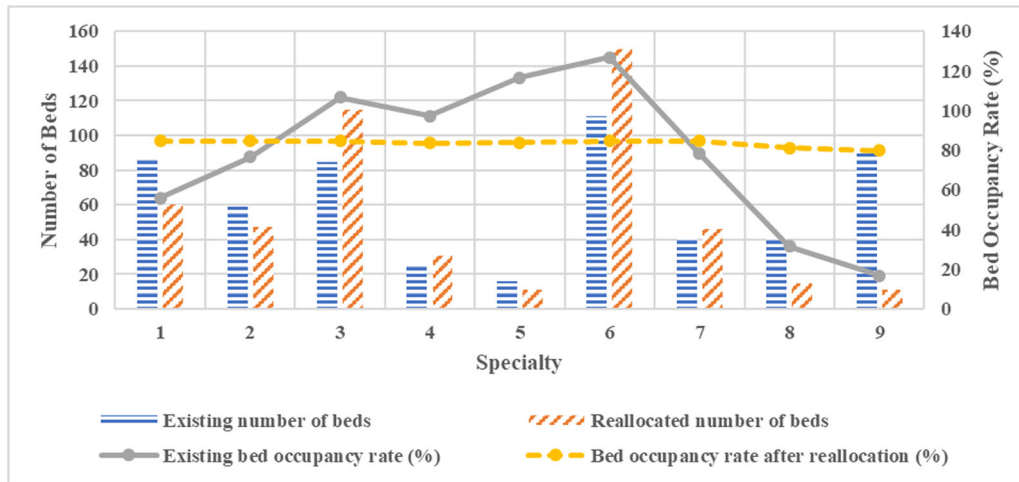


Figure 3. Comparison of results with the current situation. Specialties are (1) General surgery, (2) Trauma & orthopaedics, (3) General medicine, (4) Cardiology, (5) Paediatrics, (6) Geriatric medicine, (7) Obstetrics, (8) Gynaecology and (9) Others.

Figure 3 illustrates a comparison of FSO results with the current situations. According to the results, there are significant differences between the proposed number of beds and current number of beds in a few specialties. Unfortunately, many beds in a number of specialties are idle while some specialties run in overcapacity. For example, the general surgery inpatient specialty requires only 60 beds, however it currently has 88 beds. A large number of idle beds were found to be available in the other specialties. On the other hand, the geriatric medicine inpatient specialty works under severe demand pressure in terms of available beds, and the current bed occupancy rate is 127%. This specialty has been struggling against the overcapacity demand by referring and admitting the patients in beds of other wards. The specialty decreases the bed occupancy rate from 127% to 85% on condition that the required number of beds is increased by 39 beds. In summary, the following specialties have unused beds: general surgery (with 28 beds), trauma & orthopaedics (with 12 beds), paediatrics (with 5 beds), gynaecology (with 26 beds) and others (with 80 beds). The remaining specialties require additional beds to cope with the overcapacity running, for example, general medicine (by 30 beds),

cardiology (by 6 beds), geriatric medicine (by 39 beds) and obstetrics (by 5 beds).

A sensitivity analysis was also carried out, increasing the forecasted demand (i.e. above the base model). The sensitivity analysis consists of 20 experiments where demand is cumulatively increased by 1% steps, for example, the demand increased by 1% of the forecasted demand in the first experiment, and 2%-increase in demand is used as input in the second experiment.

A new metric known as Demand Coverage Ratio (DCR) is developed for the purpose of measuring the percentage of patients admitted to the hospital and discharged using the available resources of each specialty. The DCR formula is given in Eq. (15). This output illustrates whether the hospital is able to cope with the demand. For example, 100% DCR means that all patient demands are met with the available resources, whereas a DCR of 85% means that the specialty is unable to meet all demand, and thus requires additional resources to admit and discharge the remaining 15%. Hospitals are able to better understand their ability using this output metric. The DCR of the hospital is calculated by the formula and used in our FSO model.

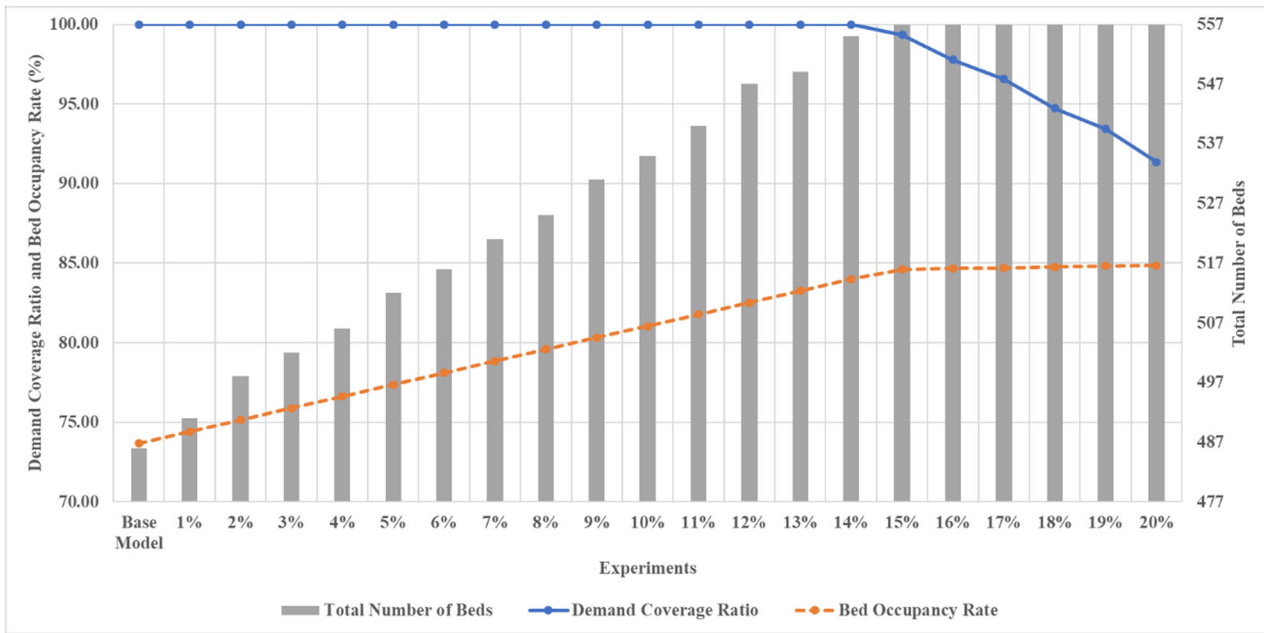


Figure 4. Analysis results with total number of beds, DCR and BOR.

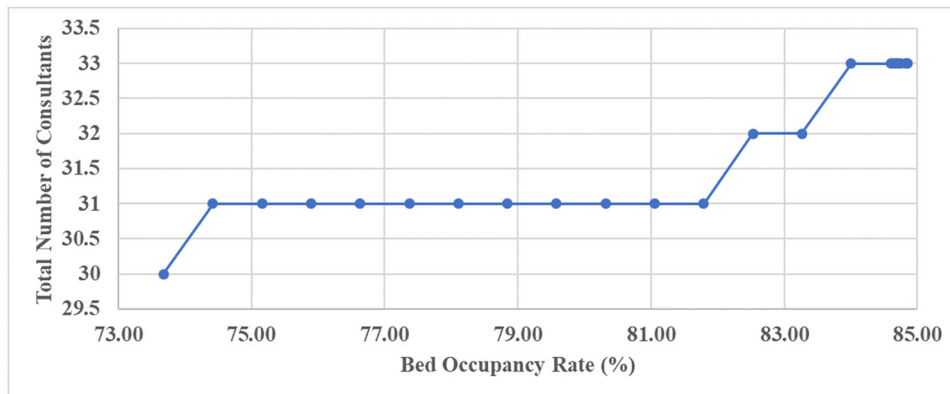


Figure 5. Relationship between BOR and total number of consultants.

$$\text{Demand Coverage Ratio (\%)} = 100 \times \left(\frac{\sum_{j=1}^{s_i} NPDI_j}{\sum_{j=1}^{s_i} NPAI_j} \right) \tag{15}$$

where $NPDI_j$ is the number of patients who are discharged using available resources from the inpatient specialty j ; $NPAI_j$ is the number of patients who are admitted to inpatient specialty j ; s_i is the total number of inpatient specialties.

We tested the sensitivity of the developed model against unexpected demand increases. The results show that the hospital wards will be able to cope with up to 14% demand increase within the forecasted year (see Figure 4). That is why DCR rates (i.e. Experiments 1 to 14) are 100%, meaning that the related inpatient specialties are able to discharge all patient demands with their available resources. In addition, BOR (bed occupancy rate) is less than 85%, meaning the wards are able to operate within

the recommended level. An increase in demand exceeding 15% will force the hospital to reject or transfer patients to other hospitals, whereas the bed occupancy rate will remain around the desired level. The demand coverage ratio reduces with increasing demand. For example, in Experiment 20, DCR is below 100% (i.e. approximately 91%). This means that the inpatient specialties were unable to cope with all demand, thus need additional resources to treat the remaining 9%.

4.2. Optimized number of staff

Figures 5 and 6 show the relationship between bed occupancy rate and total number of consultants, and bed occupancy rate and total number of nurses, respectively. According to the graphs, a minimum of 30 consultants and 158 nurses are needed to be able to operate effectively. These estimates are based on the assumption that a consultant or a nurse

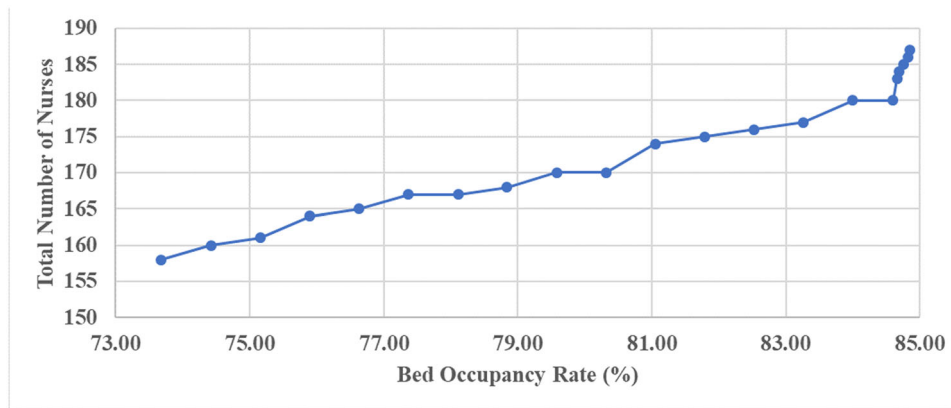


Figure 6. Relationship between BOR and total number of nurses.

spends the entire workload treating patients (typically 1950 h per year). This is a strong assumption and not realistic.

In the NHS, a full-time equivalent (FTE) is 37.5 h per week (around 1950 h per year) (NHS improvement, 2017). As seen in Table 5, the required number of consultants and nurses in all inpatient services, are determined by assuming that all staff are in full time employment (i.e. 1 FTE). However, consultants or nurses may choose to have a lower FTE rate instead of full time contracts, or the hospital management may employ on a different FTE rate. More importantly, staff do not spend all of 1950 h per year treating patients.

Therefore, the estimates in Table 5 (number of consultants and nurses) needs to be adjusted, 1) according to the percentage of their time spent on treating patients, and 2) various FTE employment contracts (for example, a hospital may have a combination of consultants, some working full time, some 0.5 FTE, etc.). In addition, consultants are typically employed in an inpatient service for a certain rate of their working hours and the remaining part is consumed by consulting in an outpatient clinic, or vice versa. Considering these reasons, the inpatient specialties will require more than the number of consultants or nurses stated in Table 5. Therefore, we specified the number of staff working with different FTE rates (i.e. 1.0, 0.8, 0.5, and a case mix). A case mix includes 50% of consultants work with 1.0 FTE, 30% with 0.8 FTE, and the remaining with 0.5 FTE. We also considered the percentage of time a typical consultant or a nurse treats patients within their contractual hours, i.e. 20%, 40% and 60%. For example, 20% means that a consultant on average spends 20% of his/her time treating patients, and the remaining time is allocated for other activities (i.e. writing reports, meetings, training, etc.). Let's assume that consultants at PAH work with a case mix of FTE's, and on average they spend 40% of their time treating patients. The inpatient services will need 109 consultants for the 8 specialties plus a

group of "Other" specialties (see Table 6). Similar interpretations can be made for nurses in Table 7.

5. Conclusion and future work

The increasing pressures on the healthcare system in the UK and other parts of the world are well documented. Widespread coverage in the media backed with findings from healthcare professionals and academics (using real life data) provides evidence of such strains (the NHS is no different). Problems surrounding these pressures include budgeting constraints, increasing demand, disjointed care and lack of workforce. These pressures will not be resolved anytime soon. Austerity in the UK will continue beyond 2020, and the impact of Brexit is likely to cause further disruptions to the UK economy, with a knock-on effect on all public services (including healthcare).

In these circumstances 'business as usual' in the NHS is not an option. This may mean a major shift in the decision-making process in relation to capacity (resource) requirements. The experiences of the authors suggest that key decision makers in most NHS hospitals rely heavily on basic statistical analysis (e.g. average number of admissions, average length of stay, average bed occupancy rates). A simple analysis of activity around averages (or any other statistical measures) within a specialty or a service will not be adequate for an effective solution of the problem. A holistic approach that integrates the entire hospital's specialties is needed, as all the services are interconnected, and cannot be assumed to be independent of each other. There is an increasing demand, and capacity shortages across all the services in the NHS, thus modelling at this level of detail is a necessity.

After an exhaustive review of the literature we notice that whole hospital models are rare, however hybrid methodologies which combine forecasting, simulation and optimization techniques are now appearing. We developed a model that linked each

Table 6. Number of consultants depending on different FTE ratios for the base model. FTE: Full time equivalence, Case Mix is 50% with 1.0 FTE, 30% with 0.8 FTE, 20% with 0.5 FTE.

Code	Name	1.0 FTE			0.8 FTE			0.5 FTE			Case Mix		
		20%	40%	60%	20%	40%	60%	20%	40%	60%	20%	40%	60%
1	General surgery	20	10	7	25	13	8	40	20	13	30	15	10
2	Trauma & Orthopaedics	15	8	5	20	10	7	30	15	10	20	10	7
3	General Medicine	35	18	12	45	23	15	70	35	23	45	23	15
4	Cardiology	10	5	3	15	8	5	20	10	7	15	8	5
5	Paediatrics	5	3	2	10	5	3	10	5	3	10	5	3
6	Gynaecology	5	3	2	10	5	3	10	5	3	10	5	3
7	Others	5	3	2	10	5	3	10	5	3	10	5	3
8	Geriatric Medicine	40	20	13	50	25	17	80	40	27	55	28	18
9	Obstetrics	15	8	5	20	10	7	30	15	10	20	10	7
Total		150	78	51	205	104	68	300	150	99	215	109	71

Table 7. Number of nurses depending on different FTE ratios for the base model. FTE: Full time equivalence, Case Mix is 50% with 1.0 FTE, 30% with 0.8 FTE, 20% with 0.5 FTE.

Code	Name	1.0 FTE			0.8 FTE			0.5 FTE			Case Mix		
		20%	40%	60%	20%	40%	60%	20%	40%	60%	20%	40%	60%
1	General surgery	100	50	33	125	63	42	200	100	67	130	65	43
2	Trauma & Orthopaedics	75	38	25	95	48	32	150	75	50	100	50	33
3	General Medicine	185	93	62	235	118	78	370	185	123	240	120	80
4	Cardiology	50	25	17	65	33	22	100	50	33	65	33	22
5	Paediatrics	20	10	7	25	13	8	40	20	13	30	15	10
6	Gynaecology	25	13	8	35	18	12	50	25	17	35	18	12
7	Others	20	10	7	25	13	8	40	20	13	30	15	10
8	Geriatric Medicine	240	120	80	300	150	100	480	240	160	310	155	103
9	Obstetrics	75	38	25	95	48	32	150	75	50	100	50	33
Total		790	397	264	1000	504	334	1580	790	526	1040	521	346

and every service and specialty within A&E, outpatient and inpatient services, with the aim of, (1) forecasting demand for all the specialties (including first/follow up outpatient attendances, elective/non-elective inpatient admissions, and A&E admissions), (2) capturing all the uncertainties of patient pathway within a hospital setting to facilitate the testing of wide range of scenarios in the safety of a validated simulation model, (3) provide a precise estimate of the required bed capacity (and staff) needs. In terms of whole hospital modelling, our model differs from previously published examples in two ways; first, our model is more complex and therefore can be used for operational level decisions, and second, the data analysis part of our model is more elaborate than the other whole hospital models and therefore can generate more insight, which is required for better customization.

The authors had the opportunity of presenting to the boards of directors of many NHS Trusts in England, and time and time again were confronted with the following type of question: "According to your simulation findings you clearly show the estimated average and the confidence intervals (CI); exactly how many beds should we have in our service?". From experience it can be difficult to get the service manager to understand CI and, for the right reasons, they are not entirely convinced with the estimated average. The optimization enables the analyst (developer) to provide a precise estimate. The entire hospital modelling framework will

therefore facilitate the service planning and decision making and, more importantly, speed up the pace of change in the specialty or service of interest.

Given the complexity of a hospital with services across 31 specialties (outpatient, inpatient and A&E), a very large-scale analysis of data was carried out using both local data and the hospital episodes statistics dataset. We developed in total 760 forecasting models to capture demand broken down by age group, and a further 600 frequency distributions were established, all to be integrated into the simulation model. A total of 85 constraints along with objective functions were developed for the optimization model. Possible risks of harm for nurses related to work overload in hospitals were taken into consideration in the optimization model. At each stage, the decision support tool was designed, verified and validated with specialists (consultants, service managers, and directors).

According to our findings a total of 486 beds is adequate for the hospital to maintain the 85% bed occupancy target. We also noticed that the beds were inappropriately distributed. For instance, at present general medicine has 88 beds, whereas the FSO results indicate that it should have around 115 beds in order to cope with existing demand. On the other hand, other specialties have 91 beds, but only need around 11. Inappropriate bed allocation is a major concern amongst many healthcare providers. Goulding et al. (2015) explored patient's perspectives of the quality and safety of the care received

during their inpatient stay on a clinically inappropriate hospital ward and found that patients have reported dissatisfaction in terms of preference and belonging.

The FSO approach also established the required number of consultants and nurses within inpatient services. In the majority of instances, the NHS has varying FTE contractual agreements, particularly amongst consultants, where they may have multiple roles, e.g. as a researcher, academic roles, private clinics, etc. The assumption that they work full time is wrong. Based on a mix of FTE contractual agreements, if a consultant spends 20% of their workload physically treating patients, then the inpatient services will require 215 consultants.

These results can be immensely useful for the management in a number of key areas: (1) capture demand for the entire hospital for each specialty, (2) observe the impact of change (such as change in resources) on key performance metrics before it is implemented in practice, and (3) determine the optimal resource requirements with confidence (namely staff and beds) in order to meet demand now and into the future. These results will bring a different perspective to key decision makers with a DSS for short and long term strategic planning to make rational and realistic plans.

The FSO technique has a number of limitations. Outpatient services play a crucial role within a hospital; a typical NHS Trust could deal with over 300,000 attendances per year. The forecasting techniques are able to capture outpatient demand for each specialty, broken down by age and first/follow up referrals. The simulation model captures the entire outpatient services pathway with the flexibility of testing wide range of scenarios, but the optimization model does not determine the required resources, such as consultation room requirements, clinic slot requirements, outpatient staff needs, etc. Likewise, the optimization should also focus on A&E departments and determine the required number of emergency beds in order to minimize patient waiting times (i.e. less than the four-hour target from admission to discharge set by the Department of Health in England). Future research can be directed towards integrating these issues into the optimisation model.

Disclosure statement

No potential conflict of interest was reported by the authors.

ORCID

Muhammed Ordu  <http://orcid.org/0000-0003-4764-9379>

Eren Demir  <http://orcid.org/0000-0002-9087-7169>

Chris Tofallis  <http://orcid.org/0000-0001-6150-0218>

Murat M. Gunal  <http://orcid.org/0000-0002-4127-1013>

References

- Ahmed, A., & Alkhamis, T. M. (2009). Simulation optimization for an emergency department healthcare unit in Kuwait. *European Journal of Operational Research*, 198 (3), 936–942. doi:10.1016/j.ejor.2008.10.025
- BMA. (2017, February 16). State of the health system – Beds in the UK. Retrieved from <https://www.bma.org.uk/collective-voice/policy-and-research/nhs-structure-and-delivery/monitoring-quality-in-the-nhs/beds-in-the-nhs>.
- Cabrera, E., Taboada, M., Iglesias, M. L., Epelde, F., & Luque, E. (2011). Optimization of healthcare emergency departments by agent-based simulation. *Procedia Computer Science*, 4, 1880–1889. doi:10.1016/j.procs.2011.04.204
- Cabrera, E., Taboada, M., Iglesias, M. L., Epelde, F., & Luque, E. (2012). Simulation optimization for healthcare emergency departments. *Procedia Computer Science*, 9, 1464–1473. doi:10.1016/j.procs.2012.04.161
- Cappanera, P., Visintin, F., & Banditori, C. (2014). Comparing resource balancing criteria in master surgical scheduling: A combined optimisation-simulation approach. *International Journal of Production Economics*, 158, 179–196. doi:10.1016/j.ijpe.2014.08.002
- Department of Health and Social Care. (2014). *Payment by results in the NHS: tariff for 2013 to 2014*. Available at: <https://www.gov.uk/government/publications/payment-by-results-pbr-operationalguidance-and-tariffs>. [Accessed 25th March 2018].
- Demir, E., Gunal, M. & Southern, D. (2017). Demand and capacity modelling for acute services using discrete event simulation. *Health Systems*, 6, 33–40. doi:10.1057/hs.2016.1
- Ghanes, K., Wargon, M., Jouini, O., Jemai, Z., Diakogiannis, A., Hellmann, R., ... Koole, G. (2015). Simulation-based optimization of staffing levels in an emergency department. *Simulation*, 91 (10), 942–953. doi:10.1177/0037549715606808
- Goulding, L., Adamson, J., Watt, I., & Wright, J. (2015). Lost in hospital: A qualitative interview study that explores the perception of NHS inpatients who spent time on clinically inappropriate hospital wards. *Health Expectations*, 18 (5), 982–994. doi:10.1111/hex.12071
- Griffiths, P., Dall’Ora, C., & Ball, J. (2017). *How many nurses: What does the evidence say? Evidence Brief*. Southampton, UK: University of Southampton.
- Gunal, M. M. (2012). A guide for building hospital simulation models. *Health Systems*, 1 (1), 17–25. doi:10.1057/hs.2012.8
- Gunal, M. M., & Pidd, M. (2011). DGHPsim: Generic simulation of hospital performance. *ACM Transactions on Modeling and Computer Simulation (TOMACS)*, 21(4), 1. doi:10.1145/2000494.2000496
- Harper, A., Mustafee, N., & Feeney, M. (2017). A hybrid approach using forecasting and discrete-event simulation for endoscopy services. In 2017 Winter Simulation Conference (WSC), pp. 1583–1594. IEEE,
- Harper, P. R., & Shahani, A. K. (2002). Modelling for the planning and management of bed capacities in hospitals. *Journal of the Operational Research Society*, 53 (1), 11–18. doi:10.1057/palgrave.jors.2601278

- Holm, L. B., Luras, H., & Dahl, F. A. (2013). Improving hospital bed utilisation through simulation and optimisation with application to a 40% increase in patient volume in a Norwegian general hospital. *International Journal of Medical Informatics*, 82 (2), 80–89. doi:10.1016/j.ijmedinf.2012.05.006
- Hyndman, R. J., & Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting*, 22, 679–688. doi:10.1016/j.ijforecast.2006.03.001
- Kaier, K., Mutters, N. T., & Frank, U. (2012). Bed occupancy rates and hospital-acquired infections - Should beds be kept empty? *Clinical Microbiology and Infection*, 18 (10), 941–945. doi:10.1111/j.1469-0691.2012.03956.x
- Kokangul, A. (2008). A combination of deterministic and stochastic approaches to optimize bed capacity in a hospital unit. *Computer Methods and Programs in Biomedicine*, 9, 56–65. doi:10.1016/j.cmpb.2008.01.001
- Lamiri, M., Grimaud, F., & Xie, X. (2009). Optimization methods for a stochastic surgery planning problem. *International Journal of Production Economics*, 120 (2), 400–410. doi:10.1016/j.ijpe.2008.11.021
- Ma, G., & Demeulemeester, E. (2013). A multilevel integrative approach to hospital case mix and capacity planning. *Computers & Operations Research*, 40, 2198–2207. doi:10.1016/j.cor.2012.01.013
- Marshall, D. A., Burgos-Liz, L., IJzerman, M. J., Crown, W., Padula, W. V., Wong, P. K. ... Osgood, N. D. (2015). Selecting a dynamic simulation modeling method for health care delivery research—Part 2: Report of the ISPOR dynamic simulation modeling emerging good practices task force. *Value in Health*, 18(2), 147–160. doi:10.1016/j.jval.2015.01.006
- Mielczarek, B., & Zabawa, J. (2016). Modeling healthcare demand using a hybrid simulation approach. In E. T. M. K. Roeder, P. I. Frazier, R. Szechtman, E. Zhou, T. Huschka, and S. E. Chick (Ed.), *Proceedings of the 2016 Winter Simulation Conference*.
- National Health Services England. (2018a). A&E attendances and emergency admissions and bed availability and occupancy. Retrieved from <https://www.england.nhs.uk/statistics/statistical-work-areas/ae-waiting-times-and-activity/>.
- National Health Services England. (2018b). NHS inpatient elective admission events and outpatient referrals and attendances. Retrieved from <https://www.england.nhs.uk/statistics/statistical-work-areas/hospital-activity/quarterly-hospital-activity/qar-data/>.
- National Institute for Health and Care Excellence (NICE). (2014). Safe staffing for nursing in adult inpatient wards in acute hospitals. Retrieved from [nice.org.uk/guidance/sg1](https://www.nice.org.uk/guidance/sg1).
- NHS Digital. (2014). NHS staff earnings estimates to June 2014 – provisional statistics. Retrieved from <https://content.digital.nhs.uk/catalogue/PUB14955>.
- NHS Digital. (2018). *Trimpoint methodology HRG4+ 2017/18 Reference Costs Grouper*. Available at: https://digital.nhs.uk/binaries/content/assets/legacy/pdf/o/7/hrg4_201718_reference_costs_grouper_trimpoint_methodologyv1.0.pdf. [Accessed 26th May 2018]
- NHS improvement. (2017). Equality for all: Delivering safe care – seven days a week. Retrieved from <https://www.england.nhs.uk/improvement-hub/wp-content/uploads/sites/44/2017/11/Equality-for-all-Delivering-safe-care-seven-days-a-week.pdf>.
- Nuffield Trust. (2016). Understanding patient flow in hospitals. Nuffield Trust. Retrieved from <https://www.nuffieldtrust.org.uk/resource/understanding-patient-flow-in-hospitals>.
- Oddoye, J. P., Jones, D. F., Tamiz, M., & Schmidt, P. (2009). Combining simulation and goal programming for healthcare planning in a medical assessment unit. *European Journal of Operational Research*, 193 (1), 250–261. doi:10.1016/j.ejor.2007.10.029
- Ordu, M., Demir, E., & Tofallis, C. (2019). A comprehensive modelling framework to forecast the demand for all hospital services. *The International Journal of Health Planning and Management*, 34 (2), e1257–e1271. doi:10.1002/hpm.2771
- Royal College of Physicians. (2015). Work and wellbeing in the NHS: Why staff health matters to patient care. Retrieved from <https://www.rcpsych.ac.uk/pdf/RCP-%20WorkWellbeingNHS.pdf>.
- Saadouli, H., Jerbi, B., Dammak, A., Masmoudi, L., & Bouaziz, A. (2015). A stochastic optimization and simulation approach for scheduling operating rooms and recovery beds in an orthopedic surgery department. *Computers & Industrial Engineering*, 80, 72–79. doi:10.1016/j.cie.2014.11.021
- The King's Fund. (2012). Older people and emergency bed use. The King's Fund. Retrieved from <https://www.kingsfund.org.uk/publications/older-people-and-emergency-bed-use>.
- Uriarte, A. G., Zuniga, E. R., Moris, M. U., & Ng, A. H. C. (2017). How can decision makers be supported in the improvement of an emergency department? A simulation, optimization and data mining approach. *Operations Research for Health Care*, 15, 102–122. doi:10.1016/j.orhc.2017.10.003
- Wang, T., Guinet, A., & Besombes, B. (2009). A sizing tool for allocation planning of hospital bed resources. In S. McClean, P. Millard, E. El-Darzi, & C. Nugent (eds.), *Intelligent Patient Management*. Berlin, Heidelberg: Springer.
- Wang, J., Li, J., Tussey, K. & Ross, K. (2012). Reducing length of stay in emergency department: A simulation Study at a community hospital. *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transaction*, 42, 1314–1322. doi:10.1109/TSMCA.2012.2210204
- Zhang, Y., Puterman, M. L., Nelson, M., & Atkins, D. (2012). A simulation optimization approach to long-term care capacity planning. *Operations Research*, 60(2), 249–261. doi:10.1287/opre.1110.1026