Using TeleTriage to Model the Risk of Hospital Admission at the Time of Registration in an Emergency Department

Siddharth Arora[a,b] and James W. Taylor[a,c]

[a]Saïd Business School, University of Oxford, Park End Street, Oxford, OX1 1HP, U.K

[b]Siddharth.Arora@sbs.ox.ac.uk (corresponding author)

[c]James.Taylor@sbs.ox.ac.uk

Accurate and timely stratification of patients at an emergency department (ED) is imperative for efficient hospital operations and improved patient care. Patient stratification at an ED typically relies on the availability of triage information, which assesses patient acuity and is performed by clinical staff. However, triaging takes place after patient registration at the ED, and is prone to delays and interruptions. Delays in administering triage are associated with poor patient care and outcomes, especially for high-acuity patients who need to be admitted from the ED to the hospital. This motivates us, in the paper, to predict the triage category in the pre-triage phase, at the time of registration when patients arrive at the ED. We refer to the predicted triage as *TeleTriage*, as it can be administered remotely. We then use *TeleTriage*, along with other relevant features, to model the probability of a patient needing admission from the ED to the hospital. Using machine learning, we focus on the estimation of this admission risk at the time of registration, to enable early identification of patients needing admission, and the start of downstream tasks sooner. We evaluate our modelling approach using internal and external validation schemes across patient conditions, and we accommodate the asymmetric costs of decision-making associated with patient admissions at the ED. We demonstrate that the proposed modelling framework can help reduce the time taken to decide if a patient needs admission, thereby reducing the length of stay for high-acuity patients and mitigating the impact of waiting time targets on admissions.

Key words: high-acuity; length of stay; machine learning; patient stratification.

1. Introduction

Emergency departments (EDs) have been referred to as a gateway for hospital admissions (Hogan and Bouknight 2002). In the USA, admissions from the ED account for nearly 50% of all hospital admissions (Pines et al. 2013) and around 10% of total annual healthcare spending (Galarraga and Pines 2016). In the UK, EDs account for around 35% of all hospital admissions (Blunt et al. 2010) and cost £17 billion annually (Steventon et al. 2018). In England, there was an increase of 42% in hospital admissions from the ED, rising from 4.25 million in 2006/07 to 6.02 million in 2017/18 (Steventon et al. 2018); the corresponding growth in the national population and attendance at EDs over the same period was only around 9% and 13%, respectively (ONS 2017). This staggering growth in admissions can partly be attributed to an increase in the frailty of an ageing population and changes in access to primary care. The growth in demand for access to emergency healthcare has tended to exceed the corresponding growth in capacity, which has made EDs prone to poor patient flow. The resulting *crowdedness* poses challenges for the effective and timely stratification of patients.

Patient stratification at an ED is based on triage, which is the first crucial step in patient assessment. The aim of triage at the ED is to assess patient condition and prioritize patient treatment. Specifically, triaging involves assigning a health risk status (using say a triage category, such as 'major injury', 'minor injury' and 'urgent care') to the patients, and using that risk status to categorize and prioritize patients for their treatment at the ED, so that high-acuity patients are seen before low-acuity patients. Effective and timely stratification of patients based on their need for care is critical for improved outcomes, facilitating optimal resource allocation, and streamlining patient flow. To ensure timely identification of high-acuity patients, it is recommended that patients are triaged within the first 15 minutes of their arrival at the ED (RCEM 2017). However, a survey study reported that up to 62.9% of triaging is interrupted multiple times, which, in addition to delaying triage, results in errors, lower patient satisfaction, and poor patient outcomes (Johnson et al. 2022). It is worth noting that wrongly prioritising a low-acuity patient over a high-acuity patient can substantially increase the waiting time for high-acuity patients (Luo et al. 2021). In this study, we predict the triage in the pre-clinical assessment phase, which we refer to as *TeleTriage*.

The advantages of early task initiation have been discussed in the literature, for example, Batt and Terwiesch (2017) conclude that ordering diagnostic tests during the triage process can reduce treatment time. In a similar vein, we demonstrate that using the risk of admission at the time of registration, EDs could make an early inpatient bed request to the hospital, which could potentially help reduce patient boarding times, and in turn, the length of stay (LoS). Lower LoS are attractive because, for critically ill patients, even an hour of delay has been shown to increase the risk of death by 1.5% (Cardoso et al. 2011). It is worth noting that, for EDs in England, a record number of patients are waiting for long hours to be treated, discharged or admitted (King's Fund 2023). Moreover, providing ED patients with estimates of their risk (or probability) of hospital admission can help manage expectations, as communicating likely outcomes can improve patient satisfaction. It has been reported that clinical triage by a nurse is inadequate to expedite the admission of patients in an ED, and paramedics have very limited ability to predict whether a patient on their way to the hospital will need admission (Afnan et al. 2021; Levine et al. 2006). Crucially, it has been shown that ED crowdedness translates into a higher risk of a patient being admitted and an increased likelihood of the patient being triaged as high-acuity (Chen et al. 2020). These challenges make predictive tools, based on feature-rich modelling, relevant for the task of forecasting the risk of hospital admission for patients at an ED. Using a detailed patient-level dataset, we extract a range of *features* (or predictor variables) for modelling admission risk using machine learning. Previous studies have typically focused on estimating the patient's risk of admission after the triage information becomes available. We propose that the triage category, and the probability of hospital admission, are estimated at the ED front door to enable healthcare providers to make targeted and early interventions in the pre-clinical stage, and thus improve ED operations.

There has been a growing interest in employing predictive modelling in emergency healthcare, with numerous studies focusing on estimating the risk of emergency admission. In a retrospective study, Cameron et al. (2015) estimated the probability of hospital admission from the ED using triage as a model feature. In another retrospective study of electronic health records of ED patients, Parker et al. (2019) predicted the probability of hospital admission and reported that age, triage acuity, and mode of arrival were the most salient features. Sun et al. (2011) modelled the risk of hospital admission at the time of triage for all-cause ED patients and categorized patients with a predicted probability of 0.7 and above as high-risk patients who needed immediate hospital admission. A similar point was discussed by Cameron et al. (2015), who highlighted that the probability of prediction could be categorized (using appropriate cutoffs) as 'admission likely', 'admission unlikely', 'admission very unlikely' etc., and it could help improve patient streaming, bed management and decision making. Using data for nonobstetric ED presentations, Zlotnik et al. (2016) predicted the risk of hospital admission using information available at the end of the triage process. Chen et al. (2021) used retrospective observational data to predict the risk of hospital admission. Using estimates of admission probability at the time of triage, Chen et al. (2022) determined the benefits of requesting a bed early, while also considering the implications of the decision-making at the patient and the system level. For a systematic review of the literature on predicting hospital admission at the ED, see Brink et al. (2021).

We would like to highlight two points. Firstly, numerous studies on ED modelling tend to employ retrospective data (Brink et al. 2021), which can be time and cost-consuming to collect, such as, patient vital signs and lab test results. However, EDs may not have access to such detailed patient information to perform risk stratification in near real-time. Thus, we do not use any retrospective data in this study.

Secondly, using the triage information can help predict the risk of hospital admission (Parker et al. 2019). It is for this reason that studies have typically focused on estimating the risk of admission at the time of triage (e.g., Cameron et al. 2015, Chen et al. 2022). However, triage is performed face-to-face by clinical staff during the initial assessment, which incurs logistical and time costs, especially under resource constraint settings or during ED crowdedness. Thus, there is time-cost associated with the modelling, whereby predicting the risk of admission at: (1) the time of registration can result in poor prediction accuracy (as triage information is not available at registration), and (2) the time of triage can cause delays in making an inpatient bed request. This is problematic, as long wait times from bed request to hospital transfer are associated with increased odds of death at hospital discharge (Paton et al. 2019). Moreover, relying on clinician-administered triage can be limiting, as it negates the possibility of prioritizing patients prior to their clinical assessment.

To find a trade-off between prediction accuracy and the time-cost of including the triage information, Somanchi et al. (2022) propose a two-stage modelling framework, whereby in the first stage, they identify a subset of patients for whom triage has relatively little benefit for predicting the admission risk. In the second stage, for this subset of patients, the risk of admission is estimated using only the information available at the time of registration; while for the remaining patients, admission risk is estimated at the time of initial assessment after their triage information becomes available. In this paper, we build on this by: (1) predicting the triage category (which we refer to as *TeleTriage*) using only the information available at the time of registration, and (2) aiming to identify all high-acuity patients that need admission, and not just a subset of patients (using TeleTriage as one of the predictors). Specifically, we make the following three contributions to the literature that have operational implications for the EDs:

(1) **Modelling the risk of hospital admission at the time of registration –** For each patient, we generate the probability of hospital admission using readily available features at the time of registration to allow for early identification of high-acuity patients at the ED front door. We evaluate probability forecast accuracy across different illnesses, using internal and external datasets. Our internal validation examines the accuracy of a model trained on one ED site for use at that site. We also perform external validation to assess the scalability of models to other datasets and populations. Our external validation involves ED data from one hospital used for model training, and ED data from another hospital employed for evaluation. The advantage of comparing forecast cost densities for informed decision-making has been demonstrated for a different application (Arora and Taylor 2016). In contrast to previous studies, we thus also evaluate the model accuracy using differing costs associated with wrongly admitting a low admission risk patient (false positive) and failing to admit a high admission risk patient (false negative). We acknowledge that the cost of a false negative is much higher than the cost of a false positive, and while we do accommodate asymmetric costs of decision-making in this study, our model cannot account for the costs associated with a wrong decision in terms of patient outcome measures. Our empirical study contributes to the growing literature, at the intersection of healthcare operations research and analytics, aiming to improve patient care and ED efficiency (see, Ang et al. 2016; Arora et al. 2023; Leo et al. 2016; Rostami-Tabar and Ziel 2022).

(2) **Assessing the usefulness of triage category and forecasts of triage category for predicting admission risk –** The importance of moving away from the traditional nurse triaging and waiting room model, and placing the clinician at the point of entry of the ED has been discussed by Natsui et al. (2020). The EDs considered in our study had to adopt a similar strategy to meet the challenges of the pandemic,

whereby triage was moved to the entry point of the ED. To evaluate a permanent reorganisation of this type, in our modelling, for all patients in this study, the time of triage was artificially brought forward to match the time of registration. This allowed us to estimate admission risk for patients at their time of registration using the clinician-administered triage category. We compared the accuracy of this with estimation based on no triage information. We then included in the comparison admission risk forecasts based on predicted triage category, or *TeleTriage*. This approach does not rely on clinical input, and the triage process can even be administered remotely in the pre-clinical assessment phase, using, say, an electronic questionnaire. This differs from previous studies on forecasting admission risk in that our forecasts are probabilistic, personalized, and remote. Moreover, to help improve patient flow, we focus on the transparency and accountability of our modelling framework, by providing insight into key predictors of the triage category using explainable machine learning based on the SHapley Additive exPlanations (SHAP values, see Lundberg and Lee, 2017). In doing so, we contribute to the literature on patient flow and prioritization (Ding et al. 2019; Duma et al. 2023; Fabbri et al. 2024; Kamali et al. 2019; Long and Mathews 2018). In addition, building on the work of Gerlee et al. (2021), we investigate if population-level data can improve admission risk forecasting at a patient level.

(3) **Evaluating the potential of admission risk forecasts to improve patient length of stay (LOS) and estimating bed demand –** Using the probability of admission at the time of triage for decision-making has been associated with improving patient LOS at an ED (Chen et al. 2022). In this study, we employ numerical simulations to investigate if utilising forecasts of admission probability at the time of registration, based on predictions of triage category, can help reduce LOS for ED patients. We also investigate the impact of waiting time targets on the distribution of LOS and show that identification of high-acuity patients, based on admission risk at the time of registration, can help reduce LOS. Moreover, we estimate the demand for hospital beds from the ED, by aggregating the patient-level risk of admission.

The paper proceeds, in Section 2, with a description of the triage patient assessment protocol at the hospitals that provided our dataset. In the same section, we describe the dataset, which includes patient arrivals before and during the pandemic, prompting us to evaluate the impact of the pandemic on ED patient flow. Section 3 uses machine learning to predict the admission risk. Section 4 evaluates the accuracy of these predictions using internal and external validation, and includes the costs associated with decision-making in the ED during forecast evaluation. Section 5 proposes an alternative approach to forecasting admission risk, whereby we first model the triage category using historical data, and then use the predicted triage category, along with other features, to estimate admission risk. In addition, using numerical simulations, we investigate the effect on LOS of using admission risk forecasts for patient prioritization, and evaluate our modelling framework as a bed management tool. Section 6 discusses the contributions and operations insights of this study and summarizes the findings.

2. ED Assessment and Patient Flow

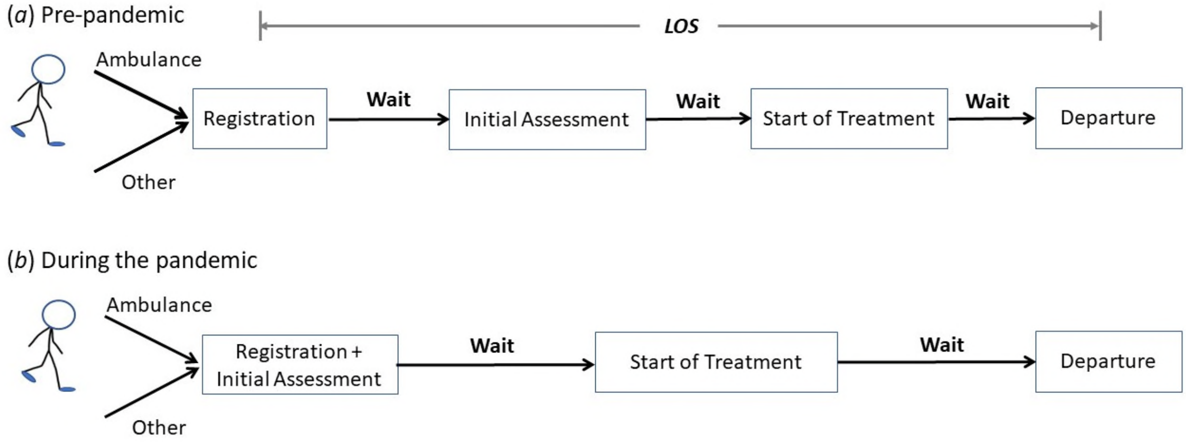
This section proceeds, in Section 2.1, by describing the ED patient assessment protocol at the hospitals that provided the data used in this paper. Section 2.2 describes the potential benefits to the ED and patient of providing a forecast of the probability of a patient being admitted. Sections 2.3 and 2.4 introduce our dataset, describing how it spans both pre-pandemic and pandemic periods.

2.1 ED Assessment Protocol

This study employs anonymised patient-level data from the John Radcliffe Hospital (JR) and Horton General Hospital (HG) situated in Oxfordshire, UK. These EDs fall within the remit of the UK's National Health Service (NHS). Figure 1a presents a schematic diagram of a typical patient flow at the ED before the pandemic, which comprises: (1) Registration – First step after arriving at the ED (at the reception desk, the patient provides their name, address, date of birth, etc.). (2) Initial assessment – Patients are prioritised typically by a nurse based on their need for care using: (a) triage category: 'major injury', 'minor injury', and 'urgent care', (b) patient group number: code to identify the reason for ED episode. The following eight patient group numbers are used at the JR and Horton hospitals (which are as per the guidelines of the NHS), 10: road traffic accident, 20: assault, 30: deliberate self-harm, 40: sports injury, 50: firework injury, 60: other accident, 70: brought in dead, and 80: other than above), and (c) chief complaint: primary reason for attendance, such as chest pain, abdominal pain, and dyspnoea. Only one chief complaint was assigned to each patient, which was the primary reason for their attendance to the ED. We coded the chief complaint as a categorical variable during the modelling. Supplement Figure A1 provides a word cloud for the 10 most common chief complaints for the JR hospital during the training period. (3) Start of treatment – A doctor assesses patients, and at this stage, the patient may undergo scans, blood tests, etc. (4) Departure – Patients leave the ED (including admission to the hospital). Figure 1b presents a schematic diagram of the redesigned patient flow during the pandemic, where the process of triage was performed, typically by a nurse, at registration.

We investigate how performing the initial assessment at registration can be beneficial for modelling admission risk by comparing the following three modelling approaches: (1) we incorporate clinical triage in the modelling (Sections 3 and 4), (2) we include no triage information (Sections 3 and 4), and (3) we include a predicted triage category, *TeleTriage* (Section 5). Before the pandemic, triage was performed at the time of initial assessment, and so, for that period, we have no historical records of patients for which triage took place at registration. To overcome this issue for modelling approach (1), we adapted our modelling to accommodate the reorganisation of the ED assessment protocol by bringing forward the time of initial assessment to the time of registration for all patient records, so that assessment is assumed to have occurred at registration (in Sections 3 and 4).

Figure 1. Schematic diagram of patient flow in the ED: (a) pre-pandemic, and (b) during the pandemic.



While model evaluation in Sections 3 and 4 assume the triage information is available at the time of registration, we do acknowledge that this might not be the case in practice for most EDs. Thus, in Section 5, our modelling assumes that no clinician-administered triage information is available at the time of registration, and so we use TeleTriage to predict the risk of admission. This allows the proposed modelling to be used in the pre-clinical phase for both triaging and identifying patients needing admission from the ED to the hospital.

2.2 Illustration of the Usefulness of Providing at Registration Probability Forecasts of Hospital

The importance of generating a probabilistic risk score for each patient has been highlighted by Cameron et al. (2015). The risk of admission can be communicated to patients (in a simplified form, such as 'likely', 'unlikely' or 'very unlikely'), and can be used for assigning patients to work-streams (for example, 'fast-track', 'rapid discharge' or 'senior review'). Such work-streams are associated with lower waiting times and admission rates, and fewer inappropriate discharges (Kelly et al. 2007).

Figure 2 also shows the estimated admission risk. Both patients were triaged as 'major injury' and assigned the same patient group number (60: 'other accident') by the clinical staff. It is worth noting that the original time stamps of initial clinical assessment for the younger and older patients were 20:26 and 19:30, respectively. Encouragingly, the model was able to predict the triage category for both patients correctly. The older patient was admitted to the hospital, while the younger patient was not. Based on information available at the time of registration, and using the TeleTriage as a feature, the older patient's admission risk is predicted as 88.5%, while the probability for the younger patient was 35.9%.

The illustration in Figure 2 prompts us to make three important points:

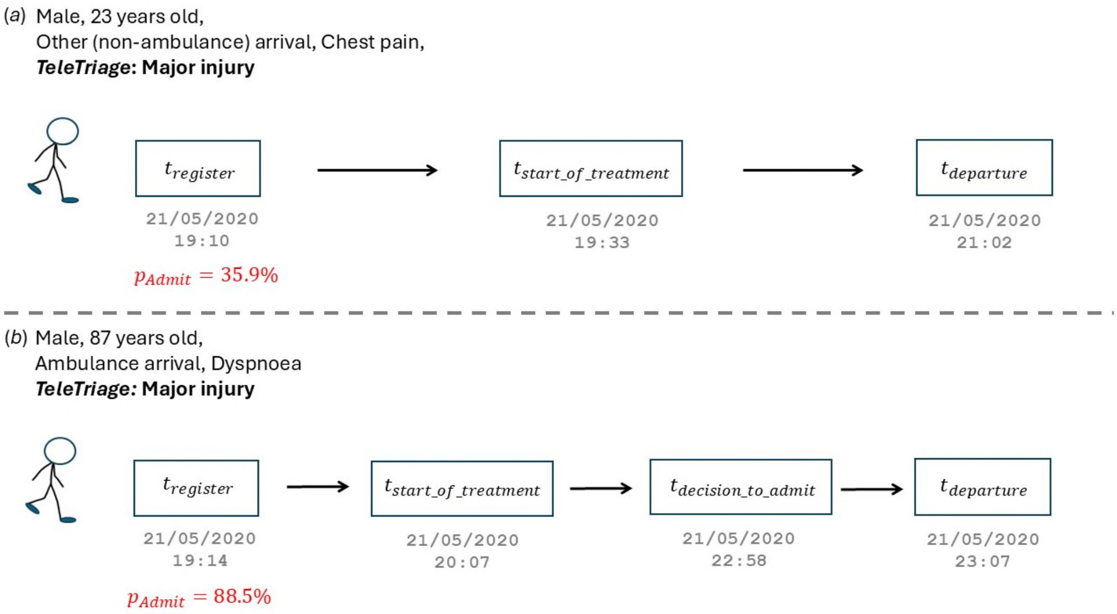
(1) Our model correctly predicted the triage category for both patients at their registration time, which can help patient prioritization in the pre-clinical phase. Moreover, at the time of registration, the model provides the clinical staff with a prediction for the risk of patient needing admission from the ED to the

hospital. It is worth noting that ED triage and modelling the risk of admission are two separate processes having different aims, and a patient with a high risk of admission may not necessarily need to be prioritized over a patient requiring immediate intervention. However, communicating the admission risk, in addition to the triage category, can provide healthcare providers with a better view of a patient's needs for essential care, and facilitate the timely prioritization of patients (using TeleTriage) and allow ED to make early requests to the hospital for inpatient beds (using the risk of admission).

(2) Patients are usually triaged first by a nurse, then assessed by an ED doctor, and then, if needed, the patient is referred to a specialist clinician. Instead, our suggestion is that at the time of registration, based on an estimate of the risk of admission, the clinical staff could decide if a patient should be referred to an acute medical team without necessarily being assessed by an ED doctor. In the NHS, acute medical units provide the entry point to patients referred by the general practitioner (GP) or the ED and have full access to investigative tests. This could help optimise the allocation of ED resources. We do acknowledge that the benefits of such an approach from a logistical perspective will, however, need to be evaluated.

(3) Using the risk of admission at the time of registration could help lower the total length of stay by reducing the boarding time from the ED to the hospital, especially in times of overcrowding. In the example of Figure 2, from the time of registration, the older patient had to wait 3 hours and 44 minutes before a request for hospital admission was made. Our modelling approach could serve as a support tool for bed management for unplanned first attendances that need admission

Figure 2. Schematic diagram illustrating patient flow. Panel (a) presents the actual data for a 23-year-old male, non-ambulance arrival, triaged as 'major injury', who presented with chest pain, and did not need admission. Panel (b) shows the next arrival at the ED as being an 87-year-old male, ambulance arrival, triaged as 'major injury', who presented with dyspnoea, and was ultimately admitted to the hospital.



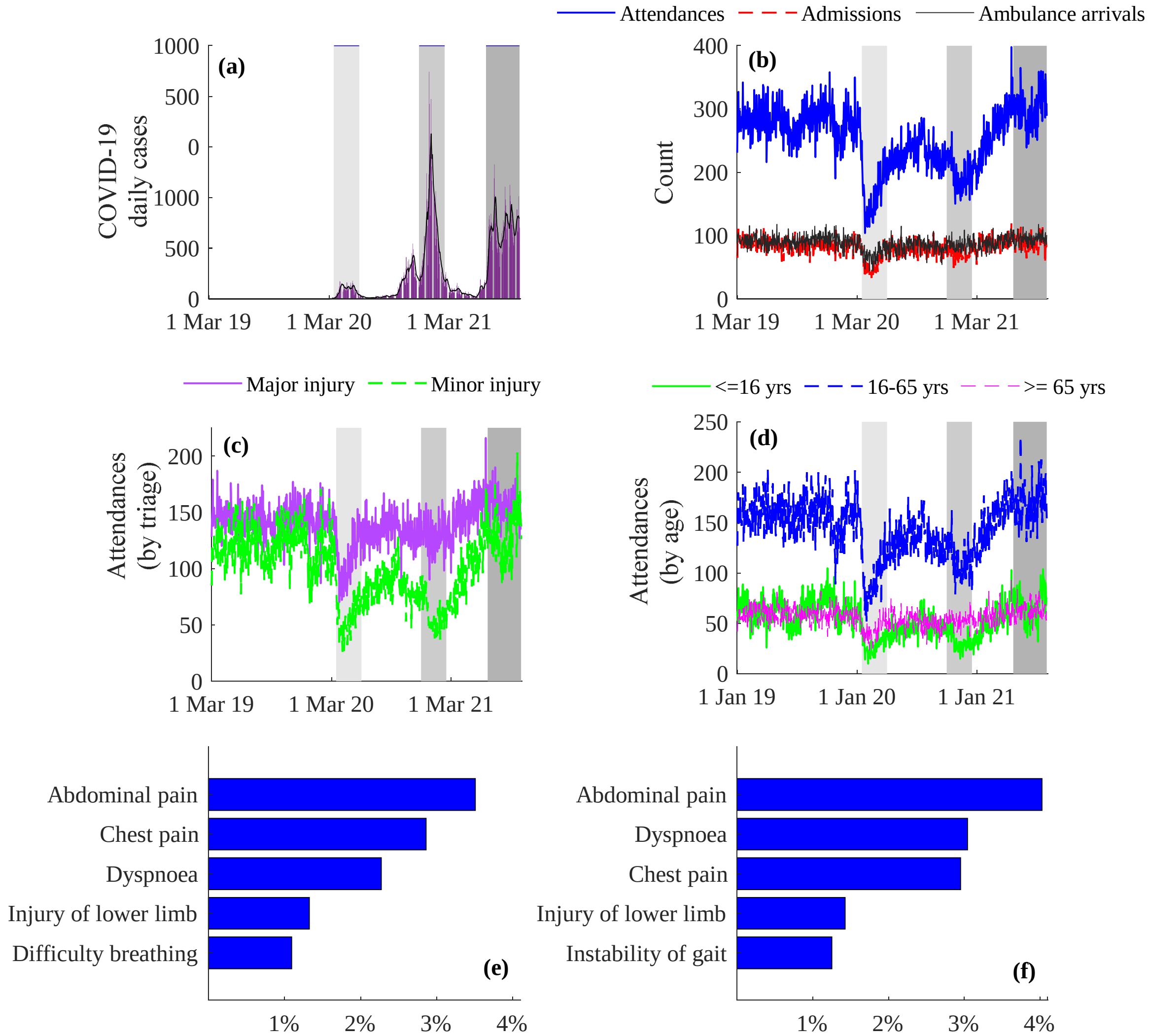
2.3 ED Data

The JR and HG are the only EDs in Oxfordshire, with the JR being the larger ED site. In view of this, and for conciseness, rather than describing the data for both of our ED sites, we describe the data for only the JR. We employ patient-level data for 1 March 2019 to 30 September 2021. During this period, 239,067 patients visited the ED at the JR. The ED staff work in an environment that is often intense, during which they need to make multiple data entries at different stages of the patient flow while also trying to deliver care to patients. This can result in incomplete or wrong data entries. We discarded entries with null entries, negative waiting times, and impossible or highly improbable values (negative waiting times, waiting time from arrival to decision to admit > 14 hours, total length of stay at the ED > 18 hours, and age > 110 years). From the 239,067 patient-level records, we identified 216,995 complete entries. Out of these, 32.4% were admitted from the ED to the hospital, it was the 'first attendance' for 98.7%, 49.6% were female, 34.4% were ambulance arrivals (55.2% of these needed admission), the average age was 40.2 years (standard deviation 26.8 years), 21.5% were minors (age ≤ 16 years), 22.2% were senior citizens (age ≥ 65 years), 55.8% were triaged as 'major injury', 38.9% were 'minor injury', and 5.3% needed 'urgent care'. To model only unplanned and unanticipated risk of admission, we predict and evaluate the risk of admission for patients categorized as first attendances. We did not have access to the source of referral or post-hospital discharge data from previous visit/s for follow-up patients. We thus used the records for n = 214,239 patients. We did not have access to COVID test outcomes for all patients, so we omitted this information from our study. As predictors, we employed only information readily available at the time of registration, and ignored any retrospective diagnosis based on, say, PCR test results and lung scans.

2.4 ED Patient Flow and the Impact of the Pandemic

As the pandemic spread, EDs faced unprecedented challenges; although the total ED attendances decreased, the proportion of patients visiting the ED with more acute symptoms increased (Jeffery et al. 2020). Figure 3 shows plots for the following key outcomes of interest: attendances, ambulance arrivals, hospital admissions for the JR, and the reason for admission. Figure 3a presents the number of daily new cases of COVID-19 reported in Oxfordshire, UK, for the period of our study, which covers the first and second pandemic waves, and part of the third wave. Figure 3b shows a sharp decline in ED attendances during each wave of the pandemic. Some days witnessed a decrease of up to around 60% in attendances compared to the same day from the previous year. Due to limited testing at the start of the pandemic, the number of confirmed COVID-19 cases was significantly underestimated. There was also a fall in the number of admissions from the ED to the hospital, and ambulance arrivals in the first wave (Figure 3b). The number of daily admissions between different waves of the pandemic were comparable with the pre-pandemic period (Figure 3b). Figure 3c shows a decrease in the number of patients triaged as 'major injury' and 'minor injury' during the pandemic. Further analysis revealed that, compared to the same period of the year, prior to the pandemic, there was a greater decrease in the proportion of patients with 'minor injury' compared to those with 'major injury'. Figure 3d shows a drop in attendances across all age groups, with the impact of the second and third waves least noticeable among elderly patients (age >= 65 years). Figures 3e and 3f present the five most common complaints associated with hospital admissions from the ED during the pre-pandemic and pandemic periods, respectively.

Figure 3. Daily COVID-19 cases, patient flow, and main reasons for admission. Panel **3a**: daily COVID-19 cases for Oxfordshire with grey vertical regions for the three waves, **3b**: daily counts of total ED attendances, admissions to the hospital from the ED, and ambulance arrivals at the ED, **3c**: attendances by triage categories 'major' and 'minor injury', **3d**: daily attendances by age group (<=16 years, between 16 and 65 years, and >=65 years), **3e**: top 5 reasons for hospital admissions from the ED during the pre-pandemic period (1 Mar 2019 to 29 Feb 2020) along with the percentage of total patients who presented with that condition (x-axis), and **3f**: top 5 reasons for admission to the hospital during the pandemic period (1 Mar 2020 to 30 Sep 2021). Plots are generated using data from the JR.



During the pandemic, as expected, there was an increase in the proportion of patients with dyspnoea, a symptom associated with COVID-19. Using admission rates, ambulance arrivals, clinical triage, and age as proxies for patient acuity, we found that, although there was a decrease in patient volumes, a larger percentage of patients arriving at the EDs had higher acuity overall. These findings are in broad agreement with previous studies that have assessed the impact of the pandemic on patient flow (Reschen et al. 2021). To deal with the evolving nature of the patient flow during the pandemic, we re-estimated the model hyperparameters using a moving window, as explained in the next section.

3. Modelling the Risk of Hospital Admission using Random Forests

In this paper, we use random forests (RF) to model the risk of hospital admission. In Section 3.1, we describe the feature extraction process, while Section 3.2 provides details regarding our use of RFs.

3.1 Feature Extraction

Using a detailed patient-level dataset, we extracted the following categories of features:

Category 1 Features: Patient demographics, calendar effects, and measures of ED workload – For each patient that attended the ED, we extracted information about demographics (age/sex), calendar effects (hour/day of arrival), mode of arrival, measures of ED workload, and staff count (estimated using the unique staff code of the staff member responsible for discharging the patient). We use ED workload measures as features because higher ED occupancy is associated with an increased risk of a patient being admitted (Chen et al. 2020). Although we generate and evaluate estimates of admission risk for 'first attendance' patients, we use data from 'follow-up' patients (both planned and unplanned repeat attendances) for feature engineering to accommodate a complete representation of the ED workload during the modelling. We extracted a total of 21 features of this type, and collected them together in a matrix that we refer to as $\boldsymbol{X}_1$. Table 1 summarizes the features.

Category 2 Features: Patient triage category, group number, and chief complaint – Since the severity of the patient's condition can help predict the risk of admission, we extracted features relating to patient triage category, patient group number, and chief complain. This information is available at the time of assessment. We use $\boldsymbol{X}_2$ to refer to a matrix containing the three features.

Category 3 Features: Population-level information – We considered whether population-level data could help inform the risk of admission for an individual. We included human mobility data for Oxfordshire from Google, and 7-day moving averages of the stringency index, which measures the strictness of lockdown policies (Hale et al. 2021). In addition, we included total COVID-19 cases in Oxfordshire, total COVID-19 hospitalisations, number of patients on mechanical ventilators, and cumulative vaccine uptake. To model the risk of admission for a patient on a given day, we used the values of the population-level variables from the previous day. We extracted 11 features of this type and placed them in a matrix that we refer to as $\boldsymbol{X}_3$. Supplement Figure A2 provides details for these features.

Table 1. Category, names, and brief descriptions of the features used to model risk of hospital admission.

Feature category and name	**Brief description**
Category 1 features ($\boldsymbol{X}_1$): Patient demographics, calendar effects, and measures of ED workload	
Age	Patient age (0 to 110 years)
Sex	Patient sex (male, female, not specified/unknown)
Calendar effects	Arrival hour of day, hour of week, day of week, and month of year
Mode of arrival	Ambulance or self-arrival
Staff count	Total hourly staff count (inferred via unique staff codes)
ED workload	Number of patients in the ED (total, ambulance arrivals, other mode of arrival, less than 16 years, between 16 and 65 years, more than 65 years, triaged as minor/major/urgent/resus, admitted in the last 4 hours, in the ED queue for whom a decision to admit has been taken but have not yet departed from ED)
4-hour breach	Number of patients who waited more than 4 hours
Category 2 features ($\boldsymbol{X}_2$): Patient triage category, group number, and chief complaint	
Triage category	Category to determine patient's priority for treatment
Group number	Code to identify the reason for ED episode (defined by the NHS as, 10: road traffic accident, 20: assault, 30: deliberate self-harm, 40: sports injury, 50: firework injury, 60: other accident, 70: brought in dead, and 80: other than above)
Chief complaint	Main symptom/reason for presentation at the ED (e.g., dyspnoea, chest pain)
Category 3 features ($\boldsymbol{X}_3$): Population-level information	
Mobility data[1]	Google mobility data, which captures changes in human movements over time across (a) retail and recreation, (b) groceries and pharmacies, (c) recreational parks, (d) transit stations, and (e) workplaces.
Stringency index[2]	A measure that aims to quantify the strictness of lockdown-style policies.
COVID related data[3]	Total number of COVID-19 cases, hospital admissions, and patients on mechanical ventilators in Oxfordshire. Total cumulative update of the first and second doses of the vaccine.

[1] https://www.google.com/covid19/mobility/
[2] https://covidtracker.bsg.ox.ac.uk/
[3] https://coronavirus.data.gov.uk/

3.2 Random Forests (RFs)

Modelling the risk of hospital admission is a binary classification problem comprising two classes: ‘admit’ and ‘not admit’. In this study, we use a RF classifier (Breiman 2001) to predict the probability (which we also refer to as ‘risk’) of admission, conditional on the feature vectors described in Section 3.1.

We used 12 month-periods of data for model training. To develop a modelling framework that can potentially adapt to the evolving nature of patient flow for different periods of the pandemic (during and between the pandemic waves), we re-estimated the model hyperparameters and re-trained the entire model at the start of each month in the out-of-sample period, using the most recent 12-month period. The model was then used for each patient arriving at the ED in the next month. As our dataset covered a period of 31 months, this led to out-of-sample forecasts for a 19-month period, which consisted of 125,002 patient records. Given the high number of patient-level records and different feature categories considered in this study, we chose RF for classification due to its overall computational advantages and ease of calculating feature importance compared to other more sophisticated methods.

In this study, around one-third (32.4%) of the attendances were admitted from the ED to the hospital, which is broadly in keeping with the national levels (Blunt et al. 2010). This, however, leads to an imbalanced dataset, i.e., a different number of training observations in the two classes. Imbalanced data can be an issue as the classifier tends to over-classify the overrepresented (or majority) class (He and Garcia 2009). A classifier biased towards identifying patients that do not need admission (majority class) would be suboptimal in classifying patients that need admission (minority class). This is particularly concerning because failing to admit a high-acuity patient is more serious than admitting a low-acuity patient. To deal with class imbalance, we considered the following three common approaches: (1) undersampling the majority class with RUSBoost (Seiffert et al. 2010), (2) oversampling the minority class with ADASYN observations (He et al. 2008), and (3) using prior probabilities (Breiman 2001). The out-of-sample performance of RF with priors was slightly better than RF with ADASYN. Out of the three approaches, RUSBoost was the least competitive. For conciseness, the discussion and results in the paper all relate to RF with priors, while results for the other methods are presented in the supplement.

Using RF with priors, we prevent the classifier from being biased towards the majority class by using prior probabilities that were proportional to the class relative frequency in the response variable during the modelling. The class prior probabilities were calculated using only the training data and recalculated each time we retrained the model. This approach required the estimation of the number of trees, the minimum size of the leaf node, and the number of features for split-point selection.

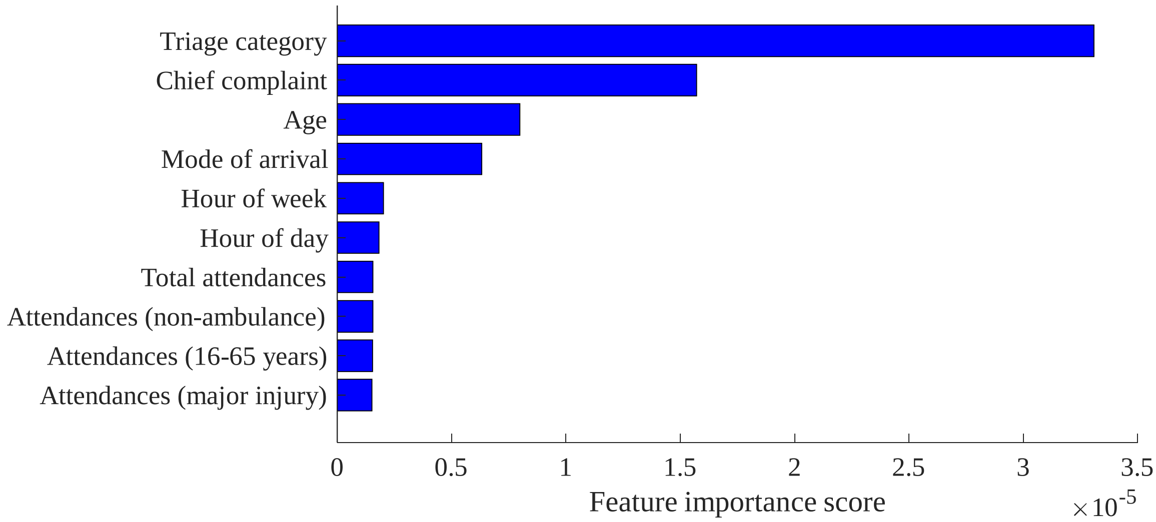
4 Evaluating Forecasts of the Risk of Hospital Admission from the ED

This section proceeds, in Section 4.1, with rankings of the most salient features in predicting the risk of admission. A statistical evaluation of whether it is beneficial to incorporate information regarding patient assessment and population-level features in a model are also presented in this section. Section 4.2 evaluates accuracy using external validation to investigate whether a model trained on one ED site can be used reliably without any adaptation at another ED site. We also evaluate whether combining patient-level datasets from multiple sites results in greater accuracy than a model built separately for each site. Section 4.3 then evaluates the model accuracy in terms of cost. This section also evaluates the benefit of a probabilistic admission forecast, in comparison with a binary classification of whether or not to admit.

4.1. Feature Importance and Statistical Evaluation of Probability Forecasts of Admission using Internal Validation

To uncover the causal relationships between the features and admission risk, for a given tree, we calculated feature importance by aggregating changes in the node risk (node impurity weighted by the node probability) from splits on each feature. These estimates were summed over all classification trees in the ensemble. Node impurity was quantified using the Gini index. As explained in Section 3.2, we produced a separate model to predict the risk of admission for each of the 19 months in the out-of-sample data. Each model provided importance scores for all features, resulting in 19 feature rankings. We averaged the feature importance scores across the 19 models to obtain a single feature ranking.

Figure 4. Feature importance scores for modelling the risk of hospital admission from the ED.



Feature rankings are presented in Figure 4 for the features found to be most salient using RF with priors. At the top of Figure 4 are two Category 2 features, triage category and chief complaint, which quantify symptom severity. The other features in the figure are from Category 1, with patient age and the mode of arrival being the most important, followed by calendar effects and ED workload features at the time of patient registration (total attendances, non-ambulance arrivals, attendances between the ages of 16 and 65 years, and patients with major injury). Population-level features (from Category 3) do not appear in Figure 4. In this section, we also evaluate out-of-sample forecasts of the probability of a patient being admitted, using the area under the receiver operating characteristic curve (AUC), Brier score, and calibration plots. For internal validation, we used data from the JR, which, as we explained in Sections 2.3 and 3.2, consisted of 214,239 patient-level records, of which 125,002 were used for out-of-sample forecast evaluation.

Table 2 presents the mean AUC (along with corresponding 95% confidence intervals), estimated using the three different choices of features indicated in each column heading. The AUC can be interpreted as the probability a classifier will rank a randomly chosen positive instance ('admit') higher than a randomly chosen negative instance ('not admit'). The AUC thus lies in the range of 0 to 1, with 1 corresponding to a perfect forecast. In Table 2a, the model performance was evaluated using all attendances in the out-of-sample data. In terms of the choice of features, the AUC using only demographic and ED workload features (X_1) was notably improved by also including features recorded at the initial assessment (X_1 U X_2). This underscores the importance of employing features that characterise patient symptoms and triage during the modelling. The resulting AUC values are comparable with the recent study of Somanchi et al. (2022) on modelling admission risk. Table 2 shows that including population-level features (X_1 U X_2 U X_3) did not noticeably improve forecast accuracy.

The utilisation of hospital resources depends on a patient's underlying condition. Thus, for this modelling framework to be used in practise, it is worth investigating its performance for different conditions. Table 2b presents the AUC for each of the five most common chief complaints associated with hospital admission from the ED at the JR. Model performance was best for patients with lower limb injuries and patients with dyspnoea, and poorest for patients with abdominal pain. Incorporating features recorded at the initial assessment (X_1 U X_2) led to the greatest improvement in accuracy for patients with dyspnoea, and the least improvement for patients with instability of gait. These findings could help EDs identify chief complaints for which additional scans, vital signs, and tests may be required to decide if a patient needs admission. Model rankings for the three different RF-based approaches to dealing with class imbalance (RUSBoost, RF with ADASYN, and RF with priors) across all attendances are provided in Table A1, while model rankings for the most common chief complaints are provided in Table A2. We also evaluated accuracy using the Brier score, which is a strictly proper scoring rule commonly used to assess probability forecasts (Brier 1950). The model rankings obtained using this score were consistent with those based on the AUC. We present the Brier score results in Supplement Table A3.

Table 2. Out-of-sample mean AUC (and 95% confidence intervals) for predicting the risk of hospital admission from the ED at the JR, for (**a**) all attendances, and (**b**) across the five most common chief complaints associated with admission, listed in descending order of frequency. RF with priors was used, with X_1, X_1 U X_2, and X_1 U X_2 U X_3.

(a) Internal Validation Across All Attendances: The JR Hospital

Method	X_1	X_1 U X_2	X_1 U X_2 U X_3
RF with priors	0.728 (0.725-0.731)	0.881 (0.879-0.883)	0.882 (0.880-0.884)

(b) Internal Validation Across Common Chief Complaints: The JR Hospital

Chief Complaint	X_1	X_1 U X_2	X_1 U X_2 U X_3
1. Abdominal Pain	0.621 (0.612-0.632)	0.697 (0.686-0.708)	0.697 (0.688-0.709)
2. Dyspnoea	0.743 (0.731-0.756)	0.851 (0.841-0.860)	0.851 (0.842-0.860)
3. Chest Pain	0.643 (0.633-0.653)	0.715 (0.705-0.724)	0.723 (0.713-0.733)
4. Injury of lower limb	0.878 (0.867-0.887)	0.961 (0.958-0.064)	0.961 (0.957-0.965)
5. Instability of gait	0.637 (0.617-0.661)	0.676 (0.653-0.700)	0.681 (0.661-0.705)

Note: Higher values of the AUC are better.

To provide further insight into probability forecast accuracy, we used calibration curves (Steyerberg et al. 2010) to convey the extent to which the predicted risk of admission matches with the corresponding actual risk. The ideal calibration curve is $y = x$, as it plots the true proportion of positive cases (i.e., admitted patients) against the mean predicted probability value. Calibration curves, estimated with feature matrix set as X_1 and as X_1 U X_2, are presented in Figure A3. The calibration for RF with priors is close to perfect when patient assessment information is included alongside demographics and ED workload measures (X_1 U X_2). Using only demographic and ED workload features (X_1) produced poor calibration.

4.2 Evaluation of Probability Forecasts using External Validation and Integrated Datasets

The models proposed in this study rely on access to large patient-level datasets, for which the EDs need to have adequate IT infrastructure and staff. It can, however, be envisaged that not all EDs would have access to sufficient datasets. Thus, to address this challenge associated with data availability, and also to gauge the generalizability of the proposed modelling framework, we perform external validation. Specifically, we assess whether a model trained on one ED site could be deployed at another ED site for which there were no reliable historical patient-level records, which could be due to a lack of IT infrastructure, poor data quality, managerial reasons, or if an ED site was newly operational.

Our external validation involved data from both hospitals involved in our study, the JR and HG. Models were trained on data from the JR, for the same period as in Section 4.1 (1 March 2019 to 29 February 2020), and then used to predict the probability of admission for patients attending the HG for the same out-of-sample period used in that section (1 March 2020 to 30 September 2021). We also evaluated models trained on data from the HG for producing probability forecasts for the JR.

In addition, we investigated whether forecast accuracy for the JR and HG could be improved by using models estimated using an integrated dataset consisting of patient records from both hospitals. The importance of integrating healthcare data from different sources for devising better treatment strategies has been emphasized by Dash et al. (2019). Recently, the benefit of integrating patient-level datasets to model patient waiting times at EDs has been investigated by Arora et al. (2023). In the context of this study, developing a centralized modelling framework by harmonizing and integrating datasets from multiple ED sites is attractive as it could facilitate interoperability and cooperation between EDs, help better accommodate the heterogeneity of patient conditions, increase the statistical power of the analysis, and aid in efficient data management and utilization of computational resources. The integrated data from the JR and HG comprised 315,860 patient-level records. The data from HG consisted of 101,621 patient-level records, of which 58,962 were out-of-sample records. In the RF modelling, an indicator variable was included as a feature to identify the ED site. In Table 3, the out-of-sample AUC values are reported for the two hospital sites, using three different validation schemes.

Table 3. Out-of-sample mean AUC values (and 95% CI) for predicting the risk of hospital admission from the EDs at the JR and HG.

JR Hospital	X_1	X_1 U X_2	X_1 U X_2 U X_3
Internal Validation	0.728 (0.725-0.731)	0.881 (0.879-0.883)	0.882 (0.880-0.884)
External Validation	0.672 (0.669-0.675)	0.850 (0.848-0.852)	0.853 (0.851-0.855)
Integrated Dataset	0.728 (0.726-0.731)	0.880 (0.879-0.882)	0.881 (0.880-0.883)

HG Hospital	X_1	X_1 U X_2	X_1 U X_2 U X_3
Internal Validation	0.785 (0.780-0.790)	0.913 (0.910-0.915)	0.911 (0.909-0.913)
External Validation	0.771 (0.767-0.776)	0.893 (0.891-0.896)	0.892 (0.889-0.895)
Integrated Dataset	0.789 (0.784-0.793)	0.913 (0.911-0.916)	0.912 (0.911-0.914)

In Table 3, the benchmark is essentially internal validation, which uses a model trained on data from one ED site for use at that site. Although the AUC values are highest for internal validation, the AUC values for external validation are comparable with the internal validation for both ED sites.

4.3 Cost Evaluation of Probability Forecasts of Admission using the Integrated Dataset

The costs of decision-making in a healthcare setting are often asymmetric. In the context of this study, poor decisions for patient prioritization can incur costs for the ED in terms of medical ethics and liability, poor clinical outcomes, and higher staffing costs. Thus, in addition to evaluation metrics, such as the

AUC and Brier score, model performance needs to be assessed using the costs associated with decision-making. This can enable healthcare managers, nurses, and emergency physicians, to better assess how digital technologies can affect operational processes and decisions. This is the focus of this section.

Previous studies on modelling the risk of hospital admission from the ED have not accommodated asymmetric costs of decision-making. We do this following a similar framework to the one employed by Taylor and Jeon (2018) for a different application. Specifically, we compare probability forecasts using the following costs of making the wrong decision: (1) $C_{NotAdmitHAR}$ is the cost of failing to admit a high admission risk patient (HAR), and (2) $C_{AdmitLAR}$ is the cost of unnecessarily admitting a low admission risk patient (LAR). If these costs are not easily quantified, they can be chosen to reflect subjective managerial preference at an ED. We consider there to be no cost associated with admitting a HAR and not admitting a LAR. Note that a HAR is a patient who was admitted from the ED to the hospital, while a LAR is a patient who was not admitted. Using these costs, and p defined as the probability of a patient being a HAP, we calculate the expected cost of admitting a patient:

$$E_{Admit} = p \times 0 + (1 - p) \times C_{AdmitLAP}$$

Similarly, the expected cost of not admitting a patient is:

$$E_{NotAdmit} = p \times C_{NotAdmitHAP} + (1 - p) \times 0$$

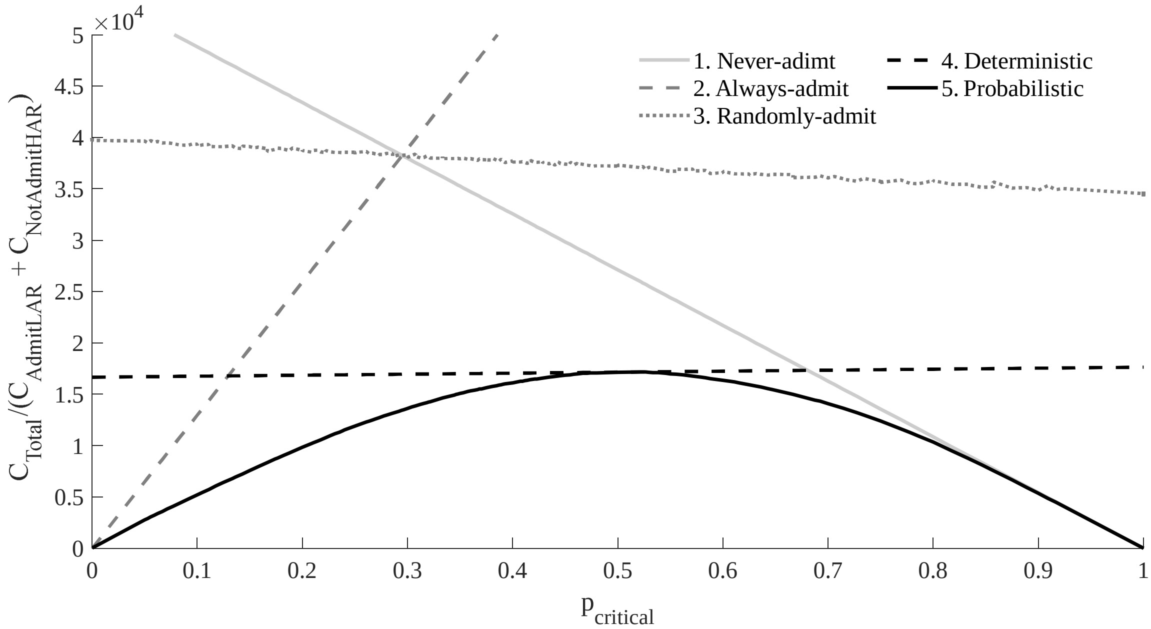
To minimise expected costs, the optimal decision would be to admit a patient if $E_{Admit} < E_{NotAdmit}$. This corresponds to p being greater than a critical value $p_{critical}$ given as:

$$p_{critical} = C_{AdmitLAP} / (C_{AdmitLAP} + C_{NotAdmitHAP})$$

To evaluate the cost associated with a probability forecast, we note that the decision would be to admit the patient if the forecast is greater than $p_{critical}$. If the cost of unnecessarily admitting a LAP is considerably lower than the cost of failing to admit a HAP, it would be optimal to admit the patient even if the predicted probability of the patient being a HAP is low. Using $C_{AdmitLAP}$, $C_{NotAdmitHAP}$, $p_{critical}$, and our model's probability forecast, we calculate the cost for each patient as: $C_{NotAdmitHAP}$ if the decision is not to admit but the patient turned out to be a HAP; $C_{AdmitLAP}$ if the decision is to admit and the patient turned out to be a LAP; and zero otherwise. We sum the costs for all patients in the out-of-sample period to calculate the total cost C_{Total}. In our empirical analysis, we computed C_{Total} for different values of $p_{critical}$. As $p_{critical}$ is a function of $C_{NotAdmitHAP}$ and $C_{AdmitLAP}$, the only way to vary $p_{critical}$ is to vary either or both of these costs, which implies a change in C_{Total}. In view of this, and following Taylor and Jeon (2018), we consider how different values for $p_{critical}$ impact the ratio of C_{Total} to the sum of $C_{AdmitLAP}$ and $C_{NotAdmitHAP}$, with lower values of this ratio being preferable.

Figure 5 plots the ratio for different values of $p_{critical}$, with the decision of whether to admit or not based on five different approaches. Approach 1 is never admit a patient. Approach 2 is always admit a patient. Approach 3 is randomly admit a patient (based on weighted random sampling, where the weights for the two classes, admit and no admit, are proportional to the percentage of patients in the training data that were admitted and not admitted). Approach 4 is *deterministic* because it uses the RF purely as a classifier, with the classification being to admit if the probability forecast of admission is more than 0.5. Our proposal is to use Approach 5, which is probabilistic. Specifically, the proposal is to admit a patient if the probability forecast from the RF exceeds $p_{critical}$. Approaches 4 and 5 use the same feature matrix, X_1 U X_2, that was most successful for the AUC, Brier score and calibration.

Figure 5. Comparison of costs associated with decision-making based on five approaches: (1) never admit a patient, (2) always admit a patient, (3) randomly admit a patient, (4) deterministic approach that uses only a binary classification output, and (5) probabilistic approach that uses the risk of admission for decision-making.



Note: For each value of $p_{critical}$, lower cost ratio values are better. Deterministic and probabilistic approaches employ predictions using the same feature matrix, X_1 U X_2.

Importantly, Figure 5 shows that each of the probabilistic approaches is preferable to the deterministic approach that used the same features (e.g., Approach 5 is superior to Approach 4), and this superiority can be seen for all values of $p_{critical}$ (except 0.5 for which the deterministic and probabilistic approaches are the same, as would be expected). This result highlights the importance of using probability forecasts of admission, rather than a binary classifier. Figure 5 employs the integrated dataset from the JR and HG, as discussed in Section 4.2.

Note that in Table 3 we found that validation using the integrated dataset delivered similar accuracy to internal validation. To corroborate the findings in Table 3, Figure A4 presents calibration curves and costs associated with decision-making, as used for internal and external validation in Sections 4.1 and 4.2, respectively. Figures A4a and A4b show that model calibration was close to perfect using internal validation and the integrated dataset, but poorer with external validation. For costs, the results are similar for internal validation and the integrated dataset (Figures A4c and A4d).

In this section, we evaluated probabilistic forecasts using the AUC and the asymmetric costs of decision-making. Next, we investigate the effect of using the risk of admission at registration on the distribution of LOS using a simulation study. Moreover, we evaluate our modelling framework to estimate the demand for hospital beds from the ED, by aggregating the patient-level risk of admission.

5 Telemonitoring of Triage Category as the basis for Forecasting Admission Risk

In the previous sections, to model admission risk at the time of registration, we employ the clinician-administered triage category as a feature in the modelling. Note that the triage category was identified as the most salient predictor of admission risk. However, the proposed modelling framework cannot be used for EDs that operate under resource-constrained settings and are unable to triage patients at the time of registration. We overcome this limitation by extending the applicability of our modelling framework to EDs that do not perform initial assessment at the time of patient registration. The aim of our modelling is not to replace the nurse but to develop a decision-support tool that can assist EDs in making early and informed patient stratification, especially in times of crowding and under resource-constrained settings.

Section 5.1 presents the motivation for modelling and predicting the clinician-administered triage category using historical data. We refer to the predicted triage as *TeleTriage*, as it can be administered remotely. Section 5.2 proposes a framework for performing TeleTriage in the pre-assessment stage in the ED queue. Section 5.3 predicts the risk of admission using the TeleTriage category. Section 5.4 evaluates the benefit of using the risk of admission for patient stratification at the ED. Section 5.5 estimates the demand for inpatient hospital bed by aggregating patient-level risk of admission.

5.1 Motivation for Our Telemonitoring Approach

To expedite patient triaging and meet waiting time targets while tackling the surge in volumes, EDs have recently investigated the adoption of TeleTriage at the entry point (Robert 2022). Encouragingly, the deployment of a TeleTriage system, which employs questionnaires about symptoms and medical history, has been shown to help eliminate front door queues, improve the visibility of high-acuity patients, improve staff efficiency, and make triage more streamlined (Robert 2022).

We aim to contribute towards digital innovations in OR at EDs by proposing a modelling framework to perform TeleTriage at registration, and generating probabilistic and personalized estimates of admission risk for automated stratification of patients in the pre-clinical assessment phase at the ED. We propose a two-stage modelling framework. In the first stage, we model the triage category of patients using readily available features that could be incorporated into a digital questionnaire. This delivers a triage forecast. In the second stage, we predict the risk of admission by including in the model the prediction of triage generated from the first stage. We envisage that these digital estimates of triage and admission risk would be validated later in the patient flow during face-to-face assessment by the clinical staff. Out approach has practical implications for:

(1) *EDs* – Estimates of TeleTriage and admission risk can assist clinical staff in making timely and cost-effective decisions for patient stratification. Moreover, such estimates can help reduce the time to decide if a patient needs admission or should bypass the queue and be seen sooner by a specialist. This could help optimize resource allocation and reduce queue lengths. This is attractive as abandonment in EDs is influenced by the queue length and the observable queue flows (Batt and Terwiesch 2015). Also, timely detection of high-acuity patients in the front door queue is relevant for, say, patients with heart attack symptoms, so they are seen within the 'golden hour'. This translates into better outcomes.

(2) *Ambulance staff* – Paramedic staff deal with complex patient conditions and provide patient care during transportation, which necessities preliminary diagnosis. It has, however, been shown that paramedics can predict the requirement for hospital admission with an accuracy of around 70% (Cummins et al. 2013), while having limited ability to predict the probability of admission and the level of required care (Levine et al. 2006). Digital triage of patients could assist paramedics in better gauging the level of patient acuity. Studies have aimed at reducing the ambulance offload delays, i.e., the time taken to transfer the patient from the ambulance to the ED by considering an optimal ambulance destination policy (Li et al. 2021). Our study has implications in reducing ambulance offload delays for high-acuity patients. Specifically, during ambulance queuing at the EDs, digital estimates of triage category and admission risk could enhance cooperation and communication between paramedics and EDs to identify high-risk patients that need to be moved from the ambulance to the ED as a priority.

(3) *Hospital speciality ward* – This study has implications for the downstream stage, i.e., the hospital ward where the patient is transferred from the ED. Using estimates of admission risk, EDs could notify the hospital ward of the likelihood that a patient may need an inpatient bed and expedite the order of diagnostic tests and medication that may be required downstream by the patient. Initiating downstream tasks earlier has been shown to improve the efficiency of care (Batt and Terwiesch 2017).

(4) *Patients* – Accurate and timely estimates of triage category and admission risk can improve patient outcomes. Moreover, it has been reported that 92% of patients feel that TeleTriage systems enhance privacy, while 82% of staff felt that they could carry out their role more effectively (Robert 2022).

5.2 TeleTriaging Patients in the Pre-clinical Assessment Stage at the ED

The objective of our TeleTriage is to use a model to replicate the clinical triage, typically performed by a nurse, as accurately as possible using historical data. This involves triaging patients as either: ‘major injury’, ‘minor injury’, or ‘urgent care’. We thus use the clinically assessed triage category as the “gold standard” (or the label). To predict the triage category, we employed an integrated dataset from the JR and HG, comprising 315,860 patient-level records. For predicting the triage category, the following five patient-specific features were used: (1) age, (2) sex, (3) chief complaint, (4) patient group number, and (5) mode of arrival. These features could readily be incorporated into a digital questionnaire to perform TeleTriage at registration.

To predict the triage category, we implemented a single classification tree, random forests, artificial neural networks (ANNs), and logistic regression, to deal with this three-class classification problem. We first estimated model hyperparameters using the last month of the training data as the cross-validation hold-out sample. For random forests, we estimated the leaf size, the number of features to use for split point selection, and the number of trees in the ensemble. To find a suitable ANN architecture, we tried up to two hidden layers, with up to five nodes within each hidden layer. Moreover, we used the rectified linear unit (ReLU) activation function for the hidden layer and the softmax activation function for the output layer. The performance of a single classification tree with default parameters was competitive compared with the other more sophisticated methods. The encouraging performance of a parsimonious method (e.g., classification tree) could be attributed to the low dimensionality of the feature matrix that comprises only five features, some of which may be correlated. Model hyperparameters and out-of-sample triage prediction accuracies are presented in the Supplement. We used prior weights during modelling to deal with class imbalance (i.e., the difference in the proportion of patients across the three triage categories), which, as we discussed briefly in Section 3.2, was the best of the approaches to class imbalance that we considered for modelling the risk of hospital admission.

To evaluate the out-of-sample accuracy, we use a confusion matrix, as it is commonly used to evaluate multi-class classification accuracy, which we provide in Supplement Table A4. It shows that we were able to predict the clinical triage assessment with an overall out-of-sample accuracy of 76.3%. The classification accuracy was highest at 80% for patients triaged as ‘major injury’ and 77.1% for those with ‘minor injury’. However, the accuracy was poor at 5.4% for patients triaged as ‘urgent care’, which could be because patients triaged as ‘urgent care’ comprised only a small proportion of the dataset. Specifically, patients with ‘major injury’, ‘minor injury’, and ‘urgent care’ comprised 55.5%, 41.1%, and 3.4% of visits in the out-of-sample period, respectively.

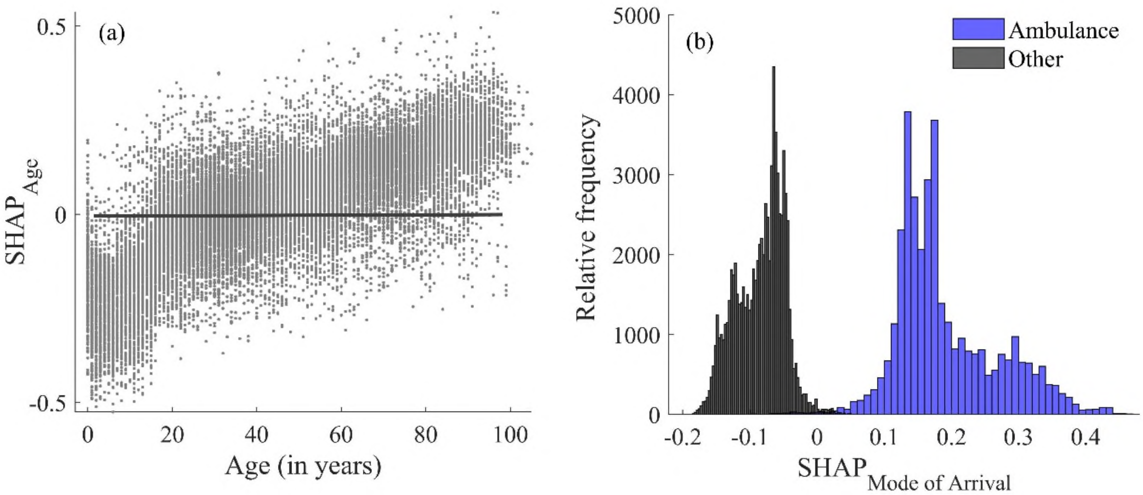
This modelling framework necessitates that the chief complaint and patient group number are assessed reliably by a trained non-clinical staff (e.g., a receptionist) at the ED registration desk, a paramedic if the patient is brought to the ED in an ambulance, or by the patient themselves. It can be foreseen that there may be discrepancies in the assessment of chief complaint between clinical and non-clinical staff. While we acknowledge this as a limitation of our study, we envisage that the digital estimates of triage and admission risk would be validated later in the patient flow during face-to-face assessment by the clinical staff at the ED.

To explain model predictions and gain insights into the triage process, we use SHAP values, which are based on game theory, where features are treated as players in a game, and the contribution of each feature to the predicted probability of an outcome measure is used to derive feature importance (Lundberg and Lee 2017). In Section 4.1, we used node impurity to derive global feature rankings for predicting risk of admission, because the computational cost associated with SHAP values is high if the number of features is large. Since only five features were used for modelling the triage category, we use SHAP values to derive the local feature importance scores in this section. In the context of this study, SHAP values are attractive as they provide feature importance at the patient level. Insights into the key factors that influence the triage process at the patient level could enhance model transparency and accountability, improve future triages via an improved understanding of previous triages, and build trust with decision-makers based on the explanations derived from prediction models.

Given the poor performance of the classifier across patients triaged as 'urgent care', in Figure 6 we present SHAP values calculated using a classification tree that was trained only on those patients that were triaged as 'major injury' or 'minor injury'. It is worth noting that patients triaged as 'urgent care' were excluded only for the calculation of SHAP value, and not from any of the subsequent analysis in this section. Specifically, the feature rankings for predicting the triage category were calculated using the mean absolute SHAP values across all patients triaged as 'major injury' or 'minor injury' in the training dataset. The mean absolute SHAP value for a given feature denotes the contribution of that feature towards prediction accuracy. Chief patient complaint, patient age, and mode of arrival were identified as the most salient features in predicting the triage category. There was no notable impact of patient sex on triage based on SHAP values, which is intuitive. Patient group number was ranked lowest; however, it can be surmised that the information in the patient group number (comprising eight codes) is encapsulated in the chief patient complaint that has a higher level of granularity (141 unique complaints). Moreover, a vast majority of the attendances were assigned the same patient group number (60: 'other accident'), for details, see Supplement Table A5. The chief complaint thus provides a more exhaustive representation of patient acuity than the patient group number.

In Figure 6a, we show the SHAP values for each patient across different age groups (each dot corresponds to the SHAP value for a given patient). For a given feature, positive SHAP values are associated with a higher chance of a patient being triaged as 'major injury'. The magnitude and direction of the SHAP values provide insight into the relationship between the feature and the prediction. As expected, an increase in age is associated with a higher positive SHAP value, and hence a higher chance of being triaged as ‘major injury’. Figure 6b shows that ambulance arrivals have a higher chance of being triaged as ‘major injury’ compared to other arrivals.

Figure 6. SHAP values for predicting the triage category as either ‘major injury’ or ‘minor injury’.



Note: positive SHAP values are associated with a higher chance of a patient being triaged as ‘major injury’.

5.3 Modelling the Risk of Admission Using Forecasts of Triage from TeleTriage

Using RF trained on the integrated dataset from the JR and HG, Table 4 compares the AUC values for predicting the risk of admission using the triage category that was: (1) clinician-administered, and (2) digitally estimated (TeleTriage). In Table 4, for the clinically assessed results, we used the actual triage category (the “gold standard”, as assessed by a clinician) along with the other features (X_1 U X_2) specified in Table 1. For the TeleTriage results, the only difference was that instead of using the actual triage category (i.e., clinician-administered triage), we used the out-of-sample predictions of the triage category when producing out-of-sample forecasts of admission risk. The objective of this analysis was to investigate if estimates of admission risk could be generated accurately without relying on clinical staff. Although the prediction accuracies of admission risk using the TeleTriage were around 8-10% lower compared to clinical triage, an attractive aspect of the TeleTriage category is that it enables timely identification of high-acuity patients without relying on clinical staff to perform triage at registration.

Moreover, to investigate if the proposed modelling framework could adequately adapt to the evolving nature of patient flow during and between the different pandemic waves, in Table 4, we present the AUC values using out-of-sample data covering the: (1) entire test period (1 March 2020 to 30 September 2021), (2) the pandemic wave period (first wave: 15 March 2020 to 31 May 2020; second wave: 29 Nov 2020 to

14 Feb 2021; and partial third wave: 20 June 2021 to 30 September 2021), and (3) period between the pandemic waves (entire test period minus the pandemic wave period). Encouragingly, as evident from Table 4, the model performance was competitive for different periods of the pandemic in the out-of-sample data.

Table 4. Out-of-sample mean AUC values (and 95% CI) for predicting the risk of hospital admission from the EDs at the JR and HG using triage category that is clinician-administered and digitally assessed for: (a) entire test period, (b) during the pandemic waves, and (c) between the pandemic waves.

Triage Category	Clinician-administered	Digitally assessed (TeleTriage)
a. Entire test period		
Integrated dataset	0.897 (0.896-0.898)	0.803 (0.800-0.805)
JR dataset	0.880 (0.879-0.882)	0.779 (0.776-0.782)
HG dataset	0.913 (0.911-0.916)	0.835 (0.831-0.839)
b. During pandemic waves		
Integrated dataset	0.897 (0.896-0.899)	0.807 (0.803-0.811)
JR dataset	0.883 (0.880-0.886)	0.784 (0.780-0.788)
HG dataset	0.914 (0.910-0.917)	0.841 (0.835-0.847)
c. Between pandemic waves		
Integrated dataset	0.897 (0.894-0.899)	0.801 (0.799-0.804)
JR dataset	0.880 (0.877-0.882)	0.777 (0.773-0.780)
HG dataset	0.913 (0.910-0.916)	0.833 (0.827-0.838)

5.4 A Simulation Study to Measure the Benefits of Estimating Admission Risk at Registration

In this section, we use simulation to investigate the potential benefits of stratifying patients based on their probability of admission, estimated in the pre-clinical assessment stage at the ED. We address the following two questions having operational implications for the ED:

(1) *Can estimates of admission risk, generated at the time of registration, help mitigate the impact of waiting time targets on the admission decision-making and length of stay?* ED physicians have been shown to admit multiple patients in a relatively short period of time, typically at the end of a physician's work shift. This behaviour has been referred to as *admission batching*. Previous studies have noted that while batching admissions reduces the physician's throughput time, it translates into long delays in the downstream queue for batched patients (Feizi et al. 2022). In a similar vein, we note that physicians tend to admit a noticeably higher proportion of patients in a short period leading up to the waiting time targets. We thus investigate the impact of waiting time targets on the ED admission decision-making process. Our contribution lies in analysing the effect, on the distribution of LOS, of using predictions of the probability of admission for patient stratification at registration, where the LOS is the time between registration and patient's departure from the ED.

(2) *Can estimates of admission risk, generated at the time of registration, help reduce the patient's length of stay?* Providing estimates of admission risk at the time of registration could potentially help physicians make an early request for hospital admission. Our modelling scheme is aimed at facilitating active involvement from the clinical staff to make proactive and timely requests regarding admitting a patient, based on the TeleTriage and admission risk forecast. To simulate patient flow at the ED, we implemented the following steps:

Step 1: For patient i, that presents at the ED, we calculate their feature vector ($X_1^{(i)}$ U $X_2^{(i)}$). Note that $X_2^{(i)}$ includes a digital estimate of the triage category, and not the clinician-administered triage. Thus, this simulation aims to emulate patient stratification in the pre-clinical phase.

Step 2: Using the feature vector ($X_1^{(i)}$ U $X_2^{(i)}$), at their time of registration, we predict the probability of admission (p_i) for patient i. We employ the integrated model trained using the JR and HG datasets.

Step 3: We convert the probability of admission into a binary decision, $\hat{y}_i = \mathbf{1}(p_i > p_{critical})$.

Step 4: If the model prediction matched the clinical decision to admit a patient (i.e., true positive, y_i and $\hat{y}_i = 1$), we recalculated the length of stay as $LOS_{Simulate}^{(i)} = LOS_{Original}^{(i)} - TTRA^{(i)}$, where $TTRA^{(i)}$ refers to the time taken to request admission by the clinical staff for patient i (from the time of registration), and $LOS_{Original}^{(i)}$ refers to the time taken from registration to departure from the ED.

Step 5: When the model prediction did not match with the clinical judgment, i.e., the case of a false positive or a false negative, we penalized the model by increasing the LOS for all patients waiting in the ED queue at the time of decision-making: $LOS_{Simulate}^{(j)} = LOS_{Original}^{(j)} + \delta_{FP} \times \mathbf{1}_{FP} + \delta_{FN} \times \mathbf{1}_{FN}$, where δ_{FP} and δ_{FN} denote the penalty for a false positive and a false negative, $\mathbf{1}_{FP}$ and $\mathbf{1}_{FN}$ are the corresponding indicator variables, and $j \in J$ (for $j \neq i$), where J represents all patients in the ED queue (i.e., patients that have registered at the ED, but have not yet been discharged) when a request to admit the patient i was made.

Step 6: Following the decision to admit (or not admit) patient i, we updated the ED workload features by incorporating $\hat{y}_i$ and $LOS_{Simulate}^{(k)}$ in the calculation of $X_1^{(i+1)}$ for the next arrival, $k \in K$ (for $k \neq i + 1$), where K represents all patients in the queue when patient $i + 1$ registers at the ED. We used the empirical arrival rate, i.e., the same patient registration timestamps as provided in the original dataset.

We would like to highlight a few points regarding this simulation. In Step 1, the registration desk is treated as a multi-server queue where patients can register and undergo TeleTriage in parallel. This is consistent with the clinician-administered triage process, which is also a multi-server service. In Step 3, for a risk-neutral strategy, $p_{critical}$ would equal 0.5, while for a risk-averse strategy, $p_{critical}$ would be less than 0.5. We consider different values of $p_{critical}$ for the simulation. In Step 4, if the decision to admit a patient is taken at registration, the $LOS_{Simulate}^{(i)}$ would be: (1) smaller than the corresponding original LOS, and (2) the same as the boarding time (time between the decision to admit a patient and their departure from the ED). A patient typically needs to wait in the ED after a decision to admit them has been made, until an inpatient bed becomes available.

We infer a false positive or a false negative as wrongly identifying a patient in the ED queue for admission, and its impact on a patient's LOS is accommodated via the penalty terms δ_{FP} and δ_{FN}, respectively. Specifically, in step 5, each time a decision to wrongly admit a low-acuity patient was made based on the model predictions, we penalized the LOS for all patients waiting in the ED queue at the time of decision-making. To select the value of the penalty, we rely on a study by Luo et al. (2021), which showed that one additional low-acuity patient arrival in an 8-hour interval increased the expected waiting time per high-acuity patient by around 0.3 to 1.2 minutes. In the absence of a counterfactual to quantify the impact of a wrong decision, i.e., the impact of a false positive and false negative on the LOS for other patients in the ED queue, we ran numerical simulations for different values of the penalty δ_{FN} ranging from 1 minute to ten minutes (in increments of 1.5 minutes). We considered three scenarios for penalizing the model for making a wrong prediction: (1) $\delta_{FP} = 0.5 \times \delta_{FN}$, (2) $\delta_{FP} = \delta_{FN}$, and (3) $\delta_{FP} = 1.5 \times \delta_{FN}$.

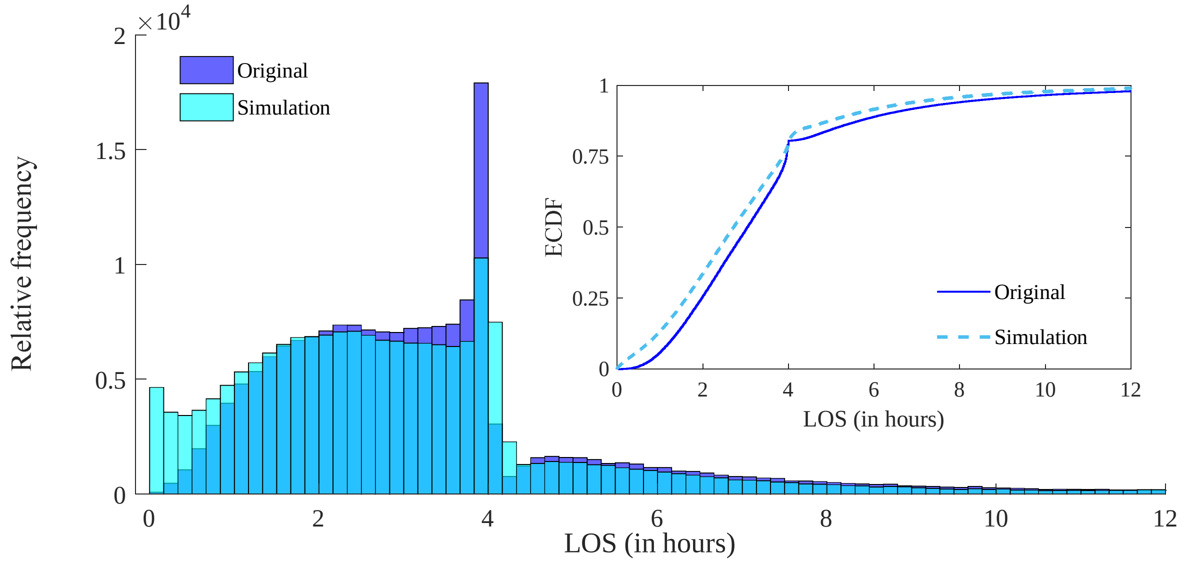
Allocating a hospital bed to low admission risk patients (false positives) can translate into long boarding times for high admission risk patients. Furthermore, it can be envisaged that not admitting a high admission risk patient (or severe delays in transferring a patient to the hospital) can result in an increased likelihood of the patient presenting with more serious condition/s and need for care by the time they see an ED physician. This can exert additional strain on the limited ED resources, especially during overcrowding. It is for this reason that we penalized both false positive and false negative predictions in our simulation study. If the model prediction matched the clinical decision not to admit a low admission risk patient (true negative), the simulated LOS was the same as the actual LOS (i.e., no penalty).

It can be surmised that a timely decision on whether a patient needs admission will affect the patient flow and waiting times for subsequent ED attendances. Thus, in step 6, the features characterizing ED workload were calculated iteratively for each patient, while also considering the accuracy of decisions to admit previous patients in the ED queue. Specifically, if the model predictions matched clinical judgement, i.e., if the decision to admit/not admit patient i was correct, it would be reflected in the calculation of ED workload features (Category 1 features) for subsequent patients in the queue who are behind patient i ($X_1^{(i+1)}$). This simulation thus allows for the LOS of future attendances, i.e., those who are yet to arrive at the ED, to be penalized multiple times, for each wrong admission decision.

Figure 7 presents the unconditional distribution of the actual LOS for the out-of-sample attendances at the JR and HG. We note that the time the decision to discharge or admit a patient was made was influenced by waiting time targets. In England, the EDs are directed to treat, admit or discharge at least

95% of the patients within four hours (NHS 2019). Thus, in Figure 7, we see that a significant proportion of the patients depart from the ED around the 4-hour mark. We also noticed that waiting time targets impacted the admission decision-making. Specifically, for patients who were admitted within four hours of their arrival, around 10% of the decisions to admit a patient were taken during the last 10 minutes of their four-hour wait. Figure 7 also presents the LOS from our simulation. Encouragingly, the results of the simulation show that facilitating early intervention and decision-making using an estimate of the probability of admission at the time of registration can potentially help mitigate the impact of waiting time targets on the distribution of LOS. In Figure 7, the simulated LOS shows that about 10,000 patients exceed the 4-hour mark by less than 20 minutes. It can thus be argued that these patients might be rushed to meet the wait time targets, in which case, the proposed modelling may not have a noticeable effect on mitigating wait time targets.

Figure 7. Relative frequencies of the total length of stay (with corresponding ECDFs in the inset) for the original dataset and simulation.



Note: Simulation results are presented for a risk-neutral strategy ($p_{critical}$ = 0.5) and a penalty (δ) of 2.5 minutes.

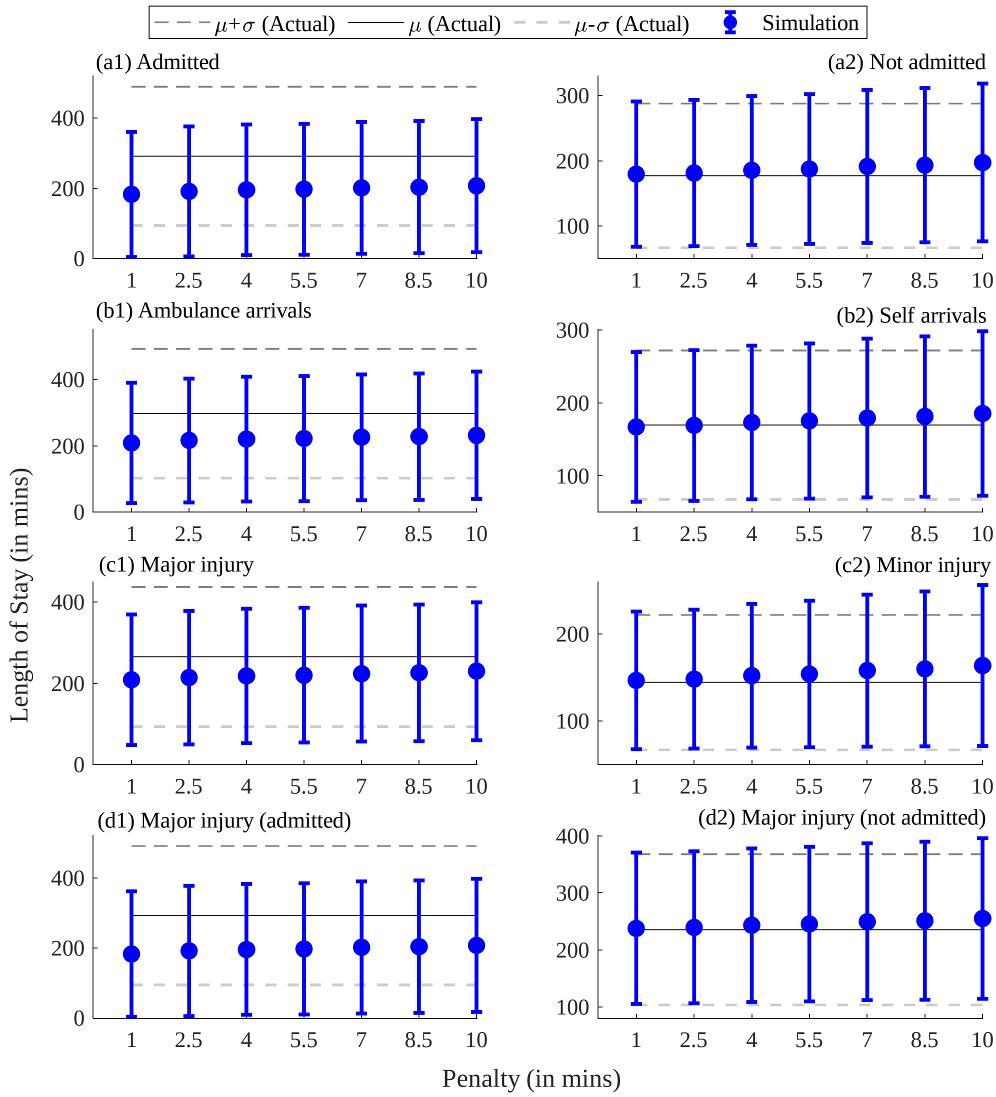
In Figure 7, compared to the original LOS distribution, the simulated LOS is lower, and the distribution is smoother, with a higher proportion of patients being admitted around the time of registration. The inset in Figure 7 presents the empirical cumulative distribution function (ECDF) of the original and simulated LOS, showing that the stratification of patients using admission risk at registration would be the preferred option based on a comparison of the two ECDFs (i.e., using first order stochastic dominance). Figure 7 presents simulation results for a risk-neutral strategy ($p_{critical}$ = 0.5). Supplement Figure A5 presents LOS for a risk-averse strategy ($p_{critical}$ = 0.4) and a less risk-averse strategy ($p_{critical}$ = 0.6), whereby a lower value of $p_{critical}$ would encourage EDs to admit more patients at the time of registration. For illustration in Figure 7, we chose a penalty $\delta_{FP} = \delta_{FN}$ = 2.5 minutes. In Supplement Figures A6 and A7, we present the LOS distribution for other penalties. We find that a lower penalty (δ = 1 minute) results in a considerable reduction in the simulated LOS, while for a large penalty (δ = 10 minutes), the simulated LOS was not preferred over the original LOS.

In Figure 8, we present the original and simulated LOS for different penalties associated with a wrong decision to admit/not admit a patient (using $\delta_{FP} = \delta_{FN}$). The figure presents four rows, each with a pair of plots. Each pair aims to compare LOS for patients with high and low acuity. In panels a1 and a2, we show LOS for patients based on their admission status. Using the probability of admission at registration can help reduce the time taken to decide if a patient needs admission, which translates into lower LOS, as shown in panel a1. However, the reduction in LOS for patients needing admission comes at the cost of slightly longer waiting times for patients that do not require admission, as shown in panel a2. As expected, a higher penalty translates into a longer LOS. Panels b1 and b2 show that, for acuity based on the mode of arrival, stratification using admission risk helps reduce waiting times for ambulance arrivals, at the cost of longer waiting times for non-ambulance arrivals. We see a similar picture when acuity is defined based on the triage category in panels c1 and c2, where the LOS for patients triaged as 'major injury' is reduced at the cost of an increase in LOS for patients triaged as 'minor injury'. Patients needing admission to the hospital are typically triaged as 'major injury'. Encouragingly, in panels d1 and d2, we see that using the probability of admission allows for a more granular level of patient stratification, whereby despite patients having the same triage category ('major injury'), patients with a high risk of admission are prioritized over those who do not need admission. These findings were consistent for patients triaged as 'urgent care', as presented in Supplement Figure A8.

Interestingly, compared to the risk-neutral strategy presented in Figure 8, we find that a lower threshold for admitting a patient (risk-averse strategy, $p_{critical}$ = 0.4) translates into a further improvement in the LOS for patients that require admission, arrived by ambulance, were triaged as major injury, or were triaged as major injury and needed admission (for simplicity, we refer to these patients as high-acuity patients in this section). However, this comes at the cost of a higher increase in LOS for low-acuity patients. Similarly, using a higher threshold to admit a patient (less risk-averse strategy, $p_{critical}$ = 0.6) results in a smaller improvement in LOS for high-acuity patients.

Supplemental Figures A9 and A10 present original and simulated LOS for different values of $p_{critical}$. It can be surmised that there is a trade-off associated with the value of $p_{critical}$ used for decision-making, where a lower value of $p_{critical}$ incentivizes higher rates of admission, resulting in lower LOS for high-acuity patients, which is relevant for improving patient outcomes. This improvement comes at the cost of more false positives and longer LOS for low-acuity patients. The optimal value of $p_{critical}$ is contextdependent, and would depend on availability of resources, workload, and costs of false positives and false negatives at an ED.

Figure 8. Mean and standard deviation of LOS (in minutes) resulting from the simulation for patients in the out-of-sample period along with corresponding actual LOS. Different values considered (on the *x*-axes) for the penalty δ (in minutes) associated with a false positive and a false negative. Predictions were generated using the feature matrix, X_1 U X_2, using the integrated dataset from the two ED sites. The simulated LOS is presented using blue vertical lines, where the blue circles in the centre denote the mean and the horizontal bars at the end denote one standard deviation from the mean.



In this study, we also analysed the original and simulated LOS for the following three penalty scenarios: (1) $\delta_{FP} = 0.5 \times \delta_{FN}$, (2) $\delta_{FP} = \delta_{FN}$, and (3) $\delta_{FP} = 1.5 \times \delta_{FN}$, as presented in Supplemental Figure A11. Figure A11 presents the mean and standard deviation of the original and simulated LOS for patients with high and low acuity. For the ease of presentation, we present results for the highest penalty value, δ_{FN} equal to 10 minutes. Consistent with the findings in Figure 8, we see that using the probability of admission at registration helps reduce the LOS for high-acuity patients (left side panels in Figure A11), but this comes at a cost of slightly higher LOS for low-acuity patients. Moreover, penalty scenario (1) corresponds to the lowest penalty overall, while scenario (3) imposes the highest penalty for a false positive. This is reflected in the simulated LOS, where the mean of simulated LOS for scenario (1) is the lowest, followed by scenario (2), with mean LOS for scenario (3) being the highest.

We also should highlight the limitations of our simulation approach. Firstly, the estimates of wait times reported by Luo et al. (2021), which were based on data from five hospitals in the USA, may or may not hold true for our simulation. Thus, to try and mitigate this limitation, we ran simulations with a wider range of delay penalties. Moreover, for every misclassification, we used a constant penalty for patients in the ED, whereas in reality, the penalty added to LOS for each patient will be different depending on a range of parameters, including patient age, triage category, and staff level, to name a few. Furthermore, we did not have information about staff schedules or the level of staff experience and skills due to data protection issues. Using the unique identifier codes of the clinical staff, we could only calculate the total staff count at the time of each patient's registration for the simulation. We also acknowledge that for high admission risk patients that need stabilization at the ED, e.g., patients presenting with acute dyspnoea, they would need airway stabilization at the ED first, before being transferred to the hospital. We did not have access to such patient scenarios in the dataset, and thus, the LOS savings reported in this section may be overestimated for patients needing immediate treatment or stabilization at the ED. It is also worth noting that high admission risk patients identified using our modelling framework who need stabilization could bypass the typical ED protocol and be transferred directly to the acute medical unit, before being admitted to the hospital.

5.5 Estimating the Demand for Hospital Beds by Aggregating Patient-level Risk of Admission

In addition to evaluating our modelling framework to predict the risk of admission and triage category, we investigate if this study could be extended to estimate the number of bed requests made from the ED to the hospital. In this setting, we aggregate the binary predictions ('admit or no admit') at the individual patient-level to calculate the number of patients that need admission (or 'number of bed requests'). Estimating the number of admissions by aggregating patient-level predictions at the time of registration could allow for real-time estimation of bed demand from the ED to the hospital. This could, in turn, facilitate timely and informed decision-making for inpatient bed utilization and pooling, and help reduce patient boarding times.

To evaluate the accuracy of our modelling approach to estimate bed demand, we divide the entire out-of-sample period (comprising say, N hours) into w non-overlapping windows of h hours each, where w = 1, 2, …, $\left\lfloor \frac{N}{h} \right\rfloor$. For a patient i in a given window, we convert their probability of admission (p_i, where $i \in w$) into a binary prediction, $\hat{y}_i$ ($\hat{y}_i \in \{0,1\}$), such that $\hat{y}_i = \mathbf{1}(p_i > p_{critical})$. The total bed demand is calculated by summing all predictions ($\sum_{i,w} \hat{y}_i$) within a given window. This process is repeated by rolling the window forward until all predictions in the entire out-of-sample data are aggregated.

Aggregating patient-level predictions to estimate outcome measures at the system-level within a hospital setting has garnered attention (Bertsimas et al. 2022, Kim et al. 2024). However, it has been shown that aggregating patient-level predictions can result in biased estimates at the system-level. To deal with this issue, studies have either used a threshold-based approach (estimating the critical threshold, $p_{critical}$, for which the number of false positives and false negatives are equal) or employed a linear regression-based correction (see, for example, Bertsimas et al. 2022, Kim et al. 2024). In this study, we used both threshold- and linear regression-based bias correction approaches. Another important point to note is that we estimate the demand for a hospital bed at the time of registration. However, in practice, the ED makes a bed request to the hospital when the clinical staff makes a patient admission request. Due to this mismatch in timestamps, we cannot simply compare the model-based predictions with the actual clinical request. To deal with this issue, we compare the model predictions against an oracle, i.e., we assume that the actual clinical request for admission was made at the time of registration. The oracle is assumed to know the clinician's request for admitting a patient in advance, i.e., at the time of registration, as opposed to making the perfect decision of whether a patient needs admission or not. We use the oracle only to evaluate the bed request predictions and not for any model training or bias correction.

Using RF trained on the integrated dataset from the JR and HG, Table 5 compares the out-of-sample mean absolute error (MAE) and root mean square error (RMSE) values for predicting the number of beds requested from the ED to the hospital using the triage category that was: (1) clinician-administered, and (2) digitally estimated (TeleTriage). For conciseness, we present results for only the linear regression-based bias correction approach, as it gave slightly better results compared to using naïve aggregate estimates and threshold-based correction. Consistent with the findings in Table 4, the MAE and RMSE values for TeleTriage are slightly higher than the model that employs clinician-administered triage. Moreover, the model errors for the JR hospital are higher than the HG, which can partly be attributed to higher volumes of admissions at the JR. Model predictions in Table 5 were aggregated across the out-of-sample data using a 4-hour non-overlapping moving window. We present the MAE and RMSE values for window sizes of two- and three-hour in Supplement Tables A6 and A7.

Table 5. Out-of-sample MAE and RMSE for estimating the number of bed requests from ED to the hospital at the JR and HG using triage category that is clinician-administered and digitally assessed.

Triage Category	Clinician-administered		Digitally assessed (TeleTriage)	
	MAE	RMSE	MAE	RMSE
Integrated dataset	2.77	3.58	3.84	5.05
JR dataset	2.64	3.45	3.10	4.08
HG dataset	1.39	1.83	1.56	2.06

Note: These results were calculated using a 4-hour moving window across all periods in the out-of-sample data.

Table 5 shows that the model accuracy in predicting bed demand is higher when clinician-administered triage is used as a feature, as compared to using *TeleTriage*. However, clinician-administered triage is available only after patient assessment, while *TeleTriage* can be estimated sooner, i.e., at the time of registration, which can allow the ED to make an early request for a hospital bed. This trade-off between model accuracy and time savings thus needs to be considered by the ED in practice. It is worth noting here that the bed demand can broadly be estimated using a time series or machine learning approach (employing historical data for the number of beds requested) or using a bottom-up approach of aggregating patient-level predictions. Although beyond the scope of this study, an interesting line of future work would be to compare these different modelling approaches.

6 Discussion and Concluding Remarks

In this study, we propose a machine learning approach using random forests to estimate the probability of admission for patients attending an ED. We employ a large feature-rich patient-level dataset from two EDs. Feature rankings were obtained, with the three most salient predictors being triage category, chief patient complaint, and age. To deal with the issue of class imbalance during model training, we considered three schemes: undersampling with boosting, oversampling, and prior weights. Model estimation was performed at the start of each month in the out-of-sample period to adapt the modelling to changes in patient flow during different waves of the pandemic. A thorough evaluation of probability forecast accuracy was undertaken based on the AUC, Brier score, and calibration curves, with validation schemes involving internal, external, and integrated datasets. Models trained using one ED site performed well when applied to the data from another ED site data. This helps garner confidence in the reliability and scalability of the modelling framework. In addition, we employed a forecast evaluation scheme based on asymmetric false positive and false negative costs. Random forest with priors was the best-performing model. Crucially, we showed that in cases where initial assessment is not performed at the time of registration, reliable estimates of triage category and hence admission risk can be produced using RF models based on historical data. Our modelling framework, which is personalized and probabilistic, could be administered remotely for stratifying patients in the pre-clinical stage at an ED. Using a simulation study, we show that using forecasts of the probability of admission estimated in the pre-clinical assessment stage can help reduce the LOS for high-acuity patients, and mitigate the impact of waiting time targets on admission decision-making and LOS. To improve transparency in reporting for this study, in Supplement Table A8, we present the TRIPOD checklist (Collins et al. 2015). The findings of this study provide the following three key operational insights for EDs:

(1) We show that the triage category that is used for patient prioritization in EDs, is also an important predictor of the likelihood of a patient needing admission. Moreover, the chief patient complaint, patient age, and mode of arrival are the most salient features in predicting the triage category.

(2) From an operational perspective, *TeleTriage* could be used for patient prioritization at the ED and help expedite the start of treatment for high acuity patients. This, in turn, could help improve patient safety and optimize staff time by speeding up the triage process. Also, using the risk of admission, EDs could make an early request for an inpatient hospital bed and reduce the boarding time for high admission risk patients. Moreover, demand for hospital beds can be estimated by aggregating patient-level risk of admission.

(3) We find that the estimates of admission risk generated during registration can facilitate timely identification and prioritization of high-acuity patients at the front door queue and help decide if a patient can bypass the usual ED protocol. Previous studies have focussed on optimizing staff recruitment considering uncertain demand conditions in the hospital (Malaki et al. 2023). For predetermined staffing levels, our study could assist EDs optimize their staffing resources by determining if it would be more beneficial for a patient to be assessed by the acute medical team, based on the admission risk estimates.

We should acknowledge the limitations of our study. Firstly, we assumed that the penalty associated with a false positive and false negative prediction would be the same in our numerical simulation. Also, we quantified the impact of a decision to admit a patient only in terms of their length of stay, and not patient-related outcomes, such as quality of life, risk of readmission, mortality, etc. It can be surmised that these patient-related outcome measures are more important than the LOS. We do, however, envisage that the predictions generated from our model in the preclinical assessment phase would be validated during the face-to-face triage by clinical staff at the time of initial assessment. We also assumed that only the admission risk estimates would be used in the pre-clinical assessment phase, but in practice, such estimates may be used along with scans and test results, although these would typically not be available at registration. Another limitation of this study was poor *TeleTriage* prediction accuracy for patients needing ‘urgent care’, which could partly be attributed to the small proportion of these patients in the dataset. Previous studies have used vital signs, such as temperature, pulse rate, respiratory rate, systolic blood pressure, and oxygen saturation, as features in the model to predict triage with high accuracy for patients with critical care outcomes, see, for example, Levin et al. (2018). These vital signs are typically available at the time of triage and can help the clinical staff make an informed triage decision. However, patient vital signs are not typically collected at the time of registration at EDs. Since our study aims to predict triage category at the time of registration, information regarding patient vital signs could not have been used as features during the modelling.

Previous research has investigated the effect of anticipated treatment on reducing ED crowdedness and patient wait times, whereby the triage nurses could be allowed to prescribe diagnostic tests before the patient is assessed by the ED physician (Visintin et al. 2019). In a similar vein, an interesting line of future work would be to use the TeleTriage and chief complaint to predict the likelihood of a patient needing a diagnostic test at the time of registration to try and streamline patient flow at the ED. Moreover, the advantage of combining distributional forecasts from different sources in a health setting has been demonstrated (Taylor and Taylor, 2023). It would thus be worth investigating combining the estimates of admission risk generated using different modelling approaches, such that the combined risk score is more accurate than the best performing model. An exciting line of future work would be to draw from recent developments in hybrid AI and combine the model predictions (for TeleTriage and the risk of admission) with corresponding estimates obtained from the clinical expert. Such a prioritization framework would leverage insights from the subjective domain expertise during the modelling.

Author contributions

SA: Conceptualization; Data curation; Formal analysis; Methodology; Validation; Visualization; Writing - original draft & editing; Revision.

JWT: Conceptualization; Validation; Writing - original draft & editing; Revision.

Declaration of interest

There is no funding to report for this work. The authors declare that they have no conflicts of interest.

Data availability

The authors do not have permission to share this data.

Declaration of generative AI in scientific writing

The authors declare that they have not used generative AI for writing this paper and take complete responsibility and accountability for the contents of this work.

References

Afnan, M. A. M., Netke, T., Singh, P., Worthington, H., Ali, F., Kajamuhan, C., and Nagra, A. 2021. "Ability of Triage Nurses to Predict, at the Time of Triage, the Eventual Disposition of Patients Attending the Emergency Department (ED): A Systematic Literature Review and Meta-Analysis," *Emergency Medicine Journal* (38:9), 694-700.

Ang, E., Kwasnick, S., Bayati, M., Plambeck, E. L., and Aratow, M. 2016. "Accurate Emergency Department Wait Time Prediction," *Manufacturing & Service Operations Man.* (18:1), 141-156.

Arora, S., Taylor, J. W., and Mak, H.-Y. 2023. "Probabilistic Forecasting of Patient Waiting Times in an Emergency Department," *Manufacturing & Service Operations Management* (25:4), 1209-1621.

Arora, S., and Taylor, J. W. 2016. "Forecasting Electricity Smart Meter Data Using Conditional Kernel Density Estimation," *Omega* 59, 47-59.

Batt, R. J., and Terwiesch, C. 2017. "Early Task Initiation and Other Load-Adaptive Mechanisms in the Emergency Department," *Management Science* (63:11), 3531-3997.

Batt, R. J., and Terwiesch, C. 2015. "Waiting Patiently: An Empirical Study of Queue Abandonment in an Emergency Department," *Management Science* (61:1), 39-59.

Bertsimas, D., Pauphilet, J., Stevens, J., and Tandon, M. 2022. "Predicting Inpatient Flow at a Major Hospital Using Interpretable Analytics", *Manufacturing & Service Operations Management* (24:6), 2809-2824.

Blunt, I., Bardsley, M., and Dixon, J. 2010. "Trends in Emergency Admissions in England 2004 - 2009," *Research report. Nuffield Trust*.

Brink, A., Alsma, J., van Attekum, L.A., Bramer, W.M., Zietse, R., Lingsma, H., and Schuit, S.C. 2022. "Predicting Inhospital Admission at the Emergency Department: A Systematic Review", *Emergency Medicine Journal* 39(3),191-198.

Breiman, L. 2001. "Random Forests," *Machine Learning* (45:1), 5-32.

Brier, G. W. 1950. "Verification of Forecasts Expressed in Terms of Probability," *Monthly Weather Review* (78:1), 1-3.

Cameron, A., Rodgers, K., Ireland, A., Jamdar, R., and Mckay, G. A. 2015. "A Simple Tool to Predict Admission at the Time of Triage," *Emergency Medicine Journal* (32:3), 174-179.

Cardoso, L.T., Grion, C.M., Matsuo, T., Anami, E.H., Kauss, I.A., Seko, L., and Bonametti, A.M. 2011. "Impact of Delayed Admission to Intensive Care Units on Mortality of Critically ill Patients: A Cohort Study", *Critical Care* (15:1), R28.

Chen, W., Argon, N. T., Bohrmann, T., Linthicum, B., Lopiano, K., Mehrotra, A., Travers, D., and Ziya, S. 2022. "Using Hospital Admission Predictions at Triage for Improving Patient Length of Stay in Emergency Departments," *Operations Research* (71:5), 1733-1755.

Chen, W. Y., Linthicum, B., Argon, N. T., Bohrmann, T., Lopiano, K., Mehrotra, A., Travers, D., and Ziya, S. 2020. "The Effects of Emergency Department Crowding on Triage and Hospital Admission Decisions," *American Journal of Emergency Medicine* (38:4), 774-779.

Chen, Z., Russo, N. W., Miller, M. M., Murphy, R. X., and Burmeister, D. B. 2021. "An Observational Study to Develop a Scoring System and Model to Detect Risk of Hospital Admission Due to Covid-19," *J Am Coll Emerg Physicians Open* (2:2), e12406.

Collins, G. S., Reitsma, J. B., Altman, D. G., and Moons, K. G. 2015. "Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD): The Tripod Statement," *BMC Med* (13), 1.

Cummins, N. M., Dixon, M., Garavan, C., Landymore, E., Mulligan, N., and O'Donnell, C. 2013. "Can Advanced Paramedics in the Field Diagnose Patients and Predict Hospital Admission?," *Emergency Medicine Journal* (30:12), 1043-1047.

Dash, S., Shakyawar, S. K., Sharma, M., and Kaushik, S. 2019. "Big Data in Healthcare: Management, Analysis and Future Prospects," *Journal of Big Data* (6:1), 1-25.

Ding, Y. C., Park, E., Nagarajan, M., and Grafstein, E. 2019. "Patient Prioritization in Emergency Department Triage Systems: An Empirical Study of the Canadian Triage and Acuity Scale (CTAS)," *Manufacturing & Service Operations Management* (21:4), 723-741.

Duma, D., and Aringhieri, R. 2023. "Real-time Resource Allocation in the Emergency Department: A Case Study," *Omega* 117, 102844.

Fabbri, C., Lombardi, M., Malaguti, E., and Monaci, M. 2024. "On-line Strategy Selection for Reducing Overcrowding in an Emergency Department," *Omega* 127, 103098.

Feizi, A., Carson, A., Jaeker, J. B., and W.E., Baker. 2022. "To Batch or Not to Batch? Impact of Admission Batching on Emergency Department Boarding Time and Physician Productivity," *Operations Research* (71:3), 939–957.

Galarraga, J. E., and Pines, J. M. 2016. "Costs of ED Episodes of Care in the United States," *American Journal of Emergency Medicine* (34:3), 357-365.

Gerlee, P., Karlsson, J., Fritzell, I., Brezicka, T., Spreco, A., Timpka, T., Joud, A., and Lundh, T. 2021. "Predicting Regional Covid-19 Hospital Admissions in Sweden Using Mobility Data," *Sci Rep* (11:1).

Hale, T., Angrist, N., Goldszmidt, R., Kira, B., Petherick, A., Phillips, T., Webster, S., Cameron-Blake, E., Hallas, L., Majumdar, S., and Tatlow, H. 2021. "A Global Panel Database of Pandemic Policies (Oxford COVID-19 Government Response Tracker)," *Nat Hum Behav* (5:4), 529-538.

He, H. B., Bai, Y., Garcia, E. A., and Li, S. T. 2008. "ADASYN: Adaptive Synthetic Sampling Approach for Imbalanced Learning," *IEEE International Joint Conference on Neural Networks, 1:8*, 1322-1328.

He, H. B., and Garcia, E. A. 2009. "Learning from Imbalanced Data," *IEEE Transactions on Knowledge and Data Engineering* (21:9), 1263-1284.

Hogan, A. J., and Bouknight, R. R. 2002. "The Emergency Department as a Gateway for Hospitalization: Evidence from the National Medical Expenditure Survey," *Journal of Health Care for the Poor and Underserved* (13:3), 288-297.

Jeffery, M. M., D'Onofrio, G., Paek, H., Platts-Mills, T. F., Soares, W. E., 3rd, Hoppe, J. A., Genes, N., Nath, B., and Melnick, E. R. 2020. "Trends in Emergency Department Visits and Hospital Admissions in Health Care Systems in 5 States in the First Months of the Covid-19 Pandemic in the US," *JAMA Intern Med* (180:10), 1328-1333.

Johnson, K., Lindsell, C., Froehle, C., and Gillespie, G. 2022. "Interruption of Initial Patient Assessment in the Emergency Department and Its Effect on Patient Perception of Care Quality," *International Journal for Quality in Health Care* (34:3), mzab146.

Kamali, M. F., Tezcan, T., and Yildiz, O. 2019. "When to Use Provider Triage in Emergency Departments," *Management Science* (65:3), 1003-1019.

Kim, S-H., Overman, W., Pauphilet, J., and Cha, W.C. 2024. "Occupancy Prediction with Patient Data: Evaluating Time-Series, Patient-Level Aggregation, and Deep Set Models", Available at SSRN: https://ssrn.com/abstract=4523116.

Kelly, A. M., Bryant, M., Cox, L., and Jolley, D. 2007. "Improving Emergency Department Efficiency by Patient Streaming to Outcomes-Based Teams," *Aust Health Rev* (31:1), 16-21.

King's Fund. 2023. "Accident and Emergency (A&E) Waiting Times." https://www.kingsfund.org.uk/projects/nhs-in-a-nutshell/waiting-times#demand.

Leo, G., Lodi, A., Tubertini, P., and Di Martino, M. 2016. "Emergency Department Management in Lazio, Italy," *Omega* 58, 128-138.

Levin, S., Toerper, M., Hamrock, E., Hinson, J.S., Barnes, S., Gardner, H., Dugas, A., Linton, B., Kirsch, T., Kelen, G. 2018. "Machine-Learning-Based Electronic Triage More Accurately Differentiates Patients With Respect to Clinical Outcomes Compared With the Emergency Severity Index," *Annals of Emergency Medicine* (71:5), 565-574.

Levine, S. D., Colwell, C. B., Pons, P. T., Gravitz, C., Haukoos, J. S., and McVaney, K. E. 2006. "How Well Do Paramedics Predict Admission to the Hospital? A Prospective Study," *Journal of Emergency Medicine* (31:1), 1-5.

Li, W., Sun, Z., and Hong, L.J. 2023. "Who is Next: Patient Prioritization Under Emergency Department Blocking," *Operations Research* (71:3), 821-842.

Li, M., Carter, A., Goldstein, J., Hawco, T., Jensen, J., and Vanberkel, P. 2021. "Determining Ambulance Destinations When Facing Offload Delays Using a Markov Decision Process," *Omega* 101, 102251.

Long, E. F., and Mathews, K. S. 2018. "The Boarding Patient: Effects of ICU and Hospital Occupancy Surges on Patient Flow," *Production and Operations Management* (27:12), 2122-2143.

Lundberg, S. M., and Lee, S. I. 2017. "A Unified Approach to Interpreting Model Predictions," *Advances in Neural Information Processing Systems 30 (NIPS 2017)* (30).

Luo, D., Bayati, M., Plambeck, E. L., and Aratow, M. 2021. "Low-Acuity Patients Delay High-Acuity Patients in the Emergency Department," *Stanford Working Paper*.

Malaki, S., Izady, N., and de Menezes, L.M. 2023. "A Framework for Optimal Recruitment of Temporary and Permanent Healthcare Workers in Highly Uncertain Environments," *European Journal of Operational Research* (308:2), 768-781.

Natsui, S., Silvestri, D. M., Salway, R. J., Iavicoli, L., Birnbaum, A., Flaherty, C., Gonzalez, O., Bouton, M., Mendez, K., Dibari, D., Albertson, P., and Wei, E. K. 2020. "Envisioning the Post-Covid-19, Pre-Vaccine Emergency Department," *Health Secur*.

NHS. 2019. "Handbook to the NHS Constitution for England."

ONS. 2017. "Overview of the UK Population: March 2017," *Office for National Statistics Report*).

Parker, C.A., Liu, N., Wu, S.X., Shen, Y., Lam, S.S.W., and Ong, M.E.H. 2019. "Predicting Hospital Admission at the Emergency Department Triage: A Novel Prediction Model", *The American Journal of Emergency Medicine* (37:8), 1498-1504.

Paton, A., Mitra, B. and Considine, J. 2019. "Longer Time to Transfer from the Emergency Department after Bed Request is Associated with Worse Outcomes", *Emergency Medicine Australasia* 31, 211-215.

Pines, J. M., Mutter, R. L., and Zocchi, M. S. 2013. "Variation in Emergency Department Admission Rates across the United States," *Med Care Res Rev* (70:2), 218-231.

RCEM. 2017. "Initial Assessment of Emergency Department Patients," *The Royal College of Emergency Medicine Report (Feb 2017)*).

Reschen, M. E., Bowen, J., Novak, A., Giles, M., Singh, S., Lasserson, D., and O'Callaghan, C. A. 2021. "Impact of the Covid-19 Pandemic on Emergency Department Attendances and Acute Medical Admissions," *BMC Emergency Medicine* (21:1), 143.

Robert, J. 2022. "Homerton Emergency Department Is Set to Roll out Etriage to Help Reduce Waiting Times." https://econsult.net/blog/homerton-emergency-department-rolls-out-etriage-to-help-reduce-waiting-times.

Rostami-Tabar, B. and Ziel, F. 2022. "Anticipating Special Events in Emergency Department Forecasting," *International Journal of Forecasting* (38:3), 1197-1213.

Seiffert, C., Khoshgoftaar, T. M., Van Hulse, J., and Napolitano, A. 2010. "Rusboost: A Hybrid Approach to Alleviating Class Imbalance," *IEEE Transactions on Systems Man and Cybernetics Part A-Systems and Humans* (40:1), 185-197.

Somanchi, S., Adjerid, I., and Gross, R. 2022. "To Predict or Not to Predict: The Case of the Emergency Department," *Production and Operations Management* (31), 799-818.

Steventon, A., Deeny, S., Friebel, R., Gardner, T., and Thorlby, R. 2018. "Briefing: Emergency Hospital Admissions in England: Which May Be Avoidable and How?," *Health Foundation Briefing Report*.

Steyerberg, E. W., Vickers, A. J., Cook, N. R., Gerds, T., Gonen, M., Obuchowski, N., Pencina, M. J., and Kattan, M. W. 2010. "Assessing the Performance of Prediction Models: A Framework for Traditional and Novel Measures," *Epidemiology* (21:1), 128-138.

Sun, Y., Heng, B.H., Tay, S.Y., and Seow, E. 2011. "Predicting Hospital Admissions at Emergency Department Triage using Routine Administrative Data", *Academic Emergency Medicine* 18(8), 844-850.

Taylor, J. W., and Jeon, J. 2018. "Probabilistic Forecasting of Wave Height for Offshore Wind Turbine Maintenance," *European Journal of Operational Research* (267:3), 877-890.

Taylor, J. W., and Taylor, K.S. 2023. "Combining Probabilistic Forecasts of COVID-19 Mortality in the United States," *European Journal of Operational Research* (304:1), 25-41.

Visintin, F., Caprara, C., and Puggelli, F. 2019. "Experimental Design and Simulation Applied to a Paediatric Emergency Department: A Case Study," *Computers & Industrial Engineering* 128, 755-781.

Zlotnik, A., Alfaro, M. C., Pérez, M. C., Gallardo-Antolín, A., and Martínez, J. M. 2016. "Building a Decision Support System for Inpatient Admission Prediction with the Manchester Triage System and Administrative Check-in Variables", *Computers, Informatics, Nursing* (34:5), 224-230.