Emergency medicine patient wait time multivariable prediction models: a multicentre derivation and validation study

Katie Walker ,[1,2,3] Jirayus Jiarpakdee,[4] Anne Loupis,[3] Chakkrit Tantithamthavorn,[4] Keith Joe,[3,5] Michael Ben-Meir,[3,6] Hamed Akhlaghi,[7,8] Jennie Hutton,[7] Wei Wang,[9,10] Michael Stephenson,[11,12] Gabriel Blecher ,[13,14] Buntine Paul,[15,16] Amy Sweeny ,[17,18] Burak Turhan,[4,19] Australasian College for Emergency Medicine, Clinical Trials Network

Handling editor Shammi L Ramlakhan

► Additional supplemental material is published online only. To view, please visit the journal online (http://dx.doi.org/10.1136/emermed-2020-211000).

For numbered affiliations see end of article.

Correspondence to
Professor Katie Walker, Emergency Department, Casey Hospital, Berwick, VIC 3806, Australia; katie_walker01@yahoo.com.au

Received 1 December 2020
Accepted 7 August 2021
Published Online First 25 August 2021

ABSTRACT

Objective Patients, families and community members would like emergency department wait time visibility. This would improve patient journeys through emergency medicine. The study objective was to derive, internally and externally validate machine learning models to predict emergency patient wait times that are applicable to a wide variety of emergency departments.

Methods Twelve emergency departments provided 3 years of retrospective administrative data from Australia (2017–2019). Descriptive and exploratory analyses were undertaken on the datasets. Statistical and machine learning models were developed to predict wait times at each site and were internally and externally validated. Model performance was tested on COVID-19 period data (January to June 2020).

Results There were 1 930 609 patient episodes analysed and median site wait times varied from 24 to 54 min. Individual site model prediction median absolute errors varied from±22.6 min (95% CI 22.4 to 22.9) to ±44.0 min (95% CI 43.4 to 44.4). Global model prediction median absolute errors varied from ±33.9 min (95% CI 33.4 to 34.0) to ±43.8 min (95% CI 43.7 to 43.9). Random forest and linear regression models performed the best, rolling average models underestimated wait times. Important variables were triage category, last-k patient average wait time and arrival time. Wait time prediction models are not transferable across hospitals. Models performed well during the COVID-19 lockdown period.

Conclusions Electronic emergency demographic and flow information can be used to approximate emergency patient wait times. A general model is less accurate if applied without site-specific factors.

Key messages

What is already known on this subject

- Patients and families want to know approximate emergency wait times, which will improve their ability to manage their logistical, physical and emotional needs while waiting.
- There are a few small studies from a limited number of jurisdictions, reporting model methods, important predictor variables and accuracy of derived models for predicting wait times.

What this study adds

- Our study demonstrates that predicting wait times from simple, readily available data are complex and provides estimates that are not as accurate as patients would like; however, rough estimates may still be better than no information.
- We present the most influential variables regarding wait times and advise against using rolling average models, preferring random forest or linear regression techniques.
- Emergency medicine machine learning models may be less generalisable to other sites than we hope for when we read manuscripts or buy commercial off-the-shelf models or algorithms. Models developed for one site lose accuracy at another site and global models built for whole systems may need customisation to each individual site. This may apply to data science clinical decision instruments as well as operational machine learning models.

INTRODUCTION

Deciding where to seek care for acute medical problems is complex and nuanced. Many decisions are made with limited information and at times little transparency from health services. Most people hope to be seen by a definitive provider immediately on arrival but usually have to wait for treatment. Emergency department (ED) proximity and wait times are the major influencers on patient choice of facility.[1–4] Wait time visibility assists with meeting physical, logistic and psychological needs of patients.[5] There is increasing consumer advocacy for transparency of information about health service resources. There is also health service interest in displaying wait times. Many examples exist in the USA, Canada and emerging interest has been seen in Australia and other jurisdictions.[6]

Information technology capabilities and applied data science techniques are becoming increasingly available to acute care services. Many EDs collect a large volume of electronic point-of-care patient data, relating to demographics, flow and clinical

To cite: Walker K, Jiarpakdee J, Loupis A, *et al*. *Emerg Med J* 2022;**39**:386–393.

care. In a community where data from multiple EDs are available, knowledge of queue lengths could facilitate optimal patient load-balancing across acute care facilities. This has the potential to reduce the harms of long waits.

Previously published information is available regarding how to predict wait times in emergency medicine. Manuscripts report a variety of predictor variables, model techniques and accuracy, from either a single centre or small number of sites. Sun *et al*[7] used quantile regression techniques in a single ED and found that by using triage categories, the number of unseen patients and the number of new patients treated by physicians in the last hour, they could predict wait times to an accuracy of ±12 min. Ang *et al*[8] compared statistical to machine learning models using data from four EDs in the USA in response to inaccurate commercial wait time predictors, finding a regularised regression model (Q Lasso) to be the best model with less underestimation of wait times using only time-of-day and day-of-week data. Arha[9] used a tree-based regression model, with simple predictor variables available at triage. Senderovich *et al*[10] found that by adding congestion variables, patient flow predictors had increased accuracy.

There is limited knowledge regarding wait time predictors performance across a variety of jurisdictions, patient catchments and healthcare resources. There is no knowledge about whether one model might be able to be applied across multiple EDs for system-wide implementation or how predictive models perform during unexpected events with variations in demand (eg, COVID-19).

Objectives

The primary objective of the study was to develop and internally validate predictive algorithms for patient wait times (triage-to-provider). Secondary objectives include determining the relative importance of each predictor variable and model method, whether models are transferable across different EDs (external validation) and the performance of the models during special events (COVID-19).

METHODS

Study design and setting

This is an observational study using retrospective administrative data to develop, compare and validate prediction models for patient wait times at EDs. Data from 2017 to 2019 from 12 EDs were used for the main study, followed by data from January to June 2020 from three EDs to test performance during COVID-19 conditions.

Mandatory point-of-care emergency patient demographic, flow and clinical data are collected for every patient by clerks and clinicians in Australia. In Victoria, these defined data populate the governmental Victorian Emergency Minimum Dataset (VEMD).[11] Data available at time of triage were used as predictor variables.

There are 24 million residents in Australia and emergency medicine manages eight million patient episodes annually. The majority (93%) of Australian residents attend government-funded, public EDs with no patient copayments. There are no restrictions on individual choice of ED. Regional to tertiary departments were invited to participate if they were part of an academic health science centre or were engaged via research networks. Ten Melbourne and two Queensland EDs participated, comprising one private, one paediatric, four major, two large metropolitan and four medium metropolitan hospitals. Hospital #7 (H7) displayed predicted patient wait times (prior in-house models) online, in their waiting room and to Ambulance Victoria during the study.

Data sources and measurements

Electronic medical record software applies time-date stamps to clinician activities (eg, triage). Clerical staff collect demographic data from patients at initial registration. Clinical staff record data while attending to a patient. The VEMD datasets from each hospital were the primary source of data for this study. VEMD data are routinely checked for completeness, accuracy and administrative errors by an emergency physician at each site prior to submission to the Victorian Government.

Three years of retrospective, deidentified VEMD data were obtained from 12 hospitals in Australia (mainly Melbourne). Hospital names were replaced with alphanumeric codes prior to analyses. All episodes of care were eligible for inclusion in the study. Data were collected in early 2020 and arranged into training (2017 and 2018) and testing datasets (2019), maintaining the temporal order based on patient arrival times. The training dataset was used for exploratory analysis and learning prediction models. The testing dataset was used to internally and externally validate the prediction models. Further retrospective data were collected from three EDs, for the period of 1 January 2020 to 30 June 2020 coinciding with major variations in emergency attendances secondary to COVID-19 concerns (April, May 2020). These data were used to evaluate the stability of model performance during unexpected circumstances.

Variables

Variables used in this study are presented in online supplemental appendix 1A. Variables collected after triage/registration were excluded from the models, except those required for calculation of the dependent variables. We used 19 predictor variables (13 VEMD and 6 derived) in total.

Outcomes

The primary outcomes of this study were triage-to-provider wait times for all patients, predicted at triage. Secondary outcomes included the accuracy of each predictive model (internal validation); determining if a global model or individual models performed better; identifying the best technique to generate these models; the relative contribution of each variable to the models; assessment of how each model performs at different sites (cross-site, external validation) and evaluation of how the models perform during COVID-19 conditions or unusual circumstances. Researchers weren't blinded to outcomes. The outcome choices and definitions were informed by a large, multisite, qualitative study of community members, consumers, paramedics and health administrators.[5] These participants recommended a prediction accuracy of ±30 min (unpublished data).

Analysis

Study size

We used time-based sampling similar to previous studies of wait time prediction that have used time periods ranging from 1 month to 1 year. Time-based sampling uses sampling over a set period, so that there is temporal separation of data used to train the model and those used for validation.[7–9] We obtained 3 years of data from each hospital to account for seasonal variations in patient visits. Multiple hospitals were enrolled to allow cross-site validation evaluations. The accepted convention of using a minimum of 30–50 data points per variable was applied.

Data cleaning, outliers and missing data

Patient data rows were checked for missing values related to the primary outcomes and episodes were removed from analysis if the primary outcome variables were missing. We therefore removed patients who left without being seen by a provider (n=133 204 (6.85%)). Other missing values were replaced with 'unknown' or 'other' categories using VEMD descriptors. Three hospitals did not collect ambulance data. The total number of unique patient episodes where the value of at least one of the predictor variables is 'unknown/other' is (n=1 733 247), covering a total of eight predictor variables (online supplemental appendix 1B).

Negative values for triage-to-provider time (n=236 (0.01%)) were removed from the analysis. We also removed patient data where the wait time exceeded the maximum of 360 min and the predefined statistical outlier threshold value (defined as 1.5 times the IQR (Q3−Q1) over Q3) which were mainly generated by administrative data entry errors for triage-to-provider time (n=13 612 (0.7%)).

Standardising and encoding data

Hospitals providing non-VEMD formatted data (Queensland) had their data converted to VEMD format. One-hot encoding[12] was applied to all categorical variables prior to prediction model development as all categorical variables were nominal with the exception of triage category. We assessed that it was preferable to lose order information for triage by applying one-hot encoding than to treat triage category as a continuous variable as the distances within levels of triage category were non-linear.

Model building and recalibration

A Python-based machine learning library (scikit-learn) and analysis tool (statsmodels) were used for model development. Guided by wait-time prediction literature,[8–10 13] we used three statistical and machine learning techniques (linear regression, random forests and elastic net regression) and a rolling average approach (the mean wait time of the previous k=4 observations). We included all predictor variables in model construction and undertook a posthoc variable importance analysis. We rebuilt the models with the most important variables only and compared the performance of the simplified models to the initial models. To foster future replications, we provide code snippets for model construction in an online repository: https://doiorg/105281/zenodo459978.

The 'last-k' variable is the mean triage-to-doctor wait time for the last 'k' patients seen by a provider. To determine the appropriate value of 'k' for this study, we performed a sensitivity analysis by observing the performance of prediction models constructed using different k values (ie, 3–10). We found that the performance differences were statistically indistinguishable across different values of k and thus selected k=4 as this produced models with the best performance.

Validation

For site-specific accuracy testing (internal validation), we used a time-wise holdout validation approach, where data were split according to defined sampling periods, and a time-defined dataset put aside for use as a test set.[14] Patient records were sorted for each hospital by their arrival time. Data from 2017 to 2018 were used to construct site-specific prediction models for all individual hospitals, while 2019 data were used to evaluate prediction models within each hospital. For cross-site evaluation of site-specific models, we used a time-wise cross-site external validation approach.[15] Site-specific models were validated against 2019 data from other hospitals (eg, train with Hospital A data from 2017 to 2018, then test with Hospital B data from 2019), resulting in 132 pairwise combinations.

We also undertook geographical and temporal cross-state external validation. Global prediction models were constructed using 2017 and 2018 data from hospitals in Victoria and evaluated using 2019 data from Queensland. We also tested the global model performance against combined 2019 Victorian all-site data.

We used two boosting ensemble algorithms (light Gradient Boosting Machines (GBM) and eXtreme gradient boosting) and one hyperparameter-optimised random forest algorithm (Random Forests) in our model validation. We found non-statistically significant improvements using these models and as they came at considerable computational cost, we excluded them from the main analysis.

For validation during unexpected events, we compared model accuracies between the first 6 months of 2019 (surrogate for normal conditions) and the first 6 months of 2020 (surrogate for unexpected events, for example, COVID-19), using data and models from three EDs.

To assess model performance, we calculated the absolute errors (AE) between the actual time and the predicted time for all models and hospitals. We then calculated the median of these distributions of absolute errors (MAE) to identify the best model for each hospital and across all 12 hospitals.

Statistical methods

The Scott-Knott effect size difference test was used to rank performance, based on MAE, of the prediction models for internal validation.[16] This is a multiple comparison approach that produces statistically distinct and non-negligible (effect size) groups of distributions. We used the implementation provided by the sk_esd function of the ScottKnottESD R package V.2.0.3. Mann Whitney U tests were used to identify whether the performance (MAE) difference between two models was statistically significant; then Cliff's delta tests,[17] ldl, were used to measure the effect size. The interpretation of Cliff's delta values is as follows: ldl <0.147 negligible, ldl <0.33 small, ldl <0.474, otherwise large.[18] We used cliff.delta function of the effsize R package V.0.7.8 for calculating Cliff's delta. For all statistical tests, we used a statistical significance level of α=0.05 and sought non-negligible effect sizes.

Patient and public involvement

The primary outcome of this study was determined by a qualitative study involving patients, the public and other stakeholders.[5] Consumers and community stakeholders contributed to the design and write up of the study.

RESULTS

Characteristics of study subjects

Twelve EDs contributed data. Two sites were unable to obtain ethics approval (regional referral, medium metropolitan). Flow through the study is presented in figure 1. Department and patient demographics are presented in table 1. The total number of patient episodes included in the study was 1 930 609 with 1 388 509 in the training and 542 100 in the testing datasets. Overall admission rates were 29% and 23% of patients arrived by ambulance.

Wait time proportional distributions were similar throughout sites, although specific wait times at each site varied, with a

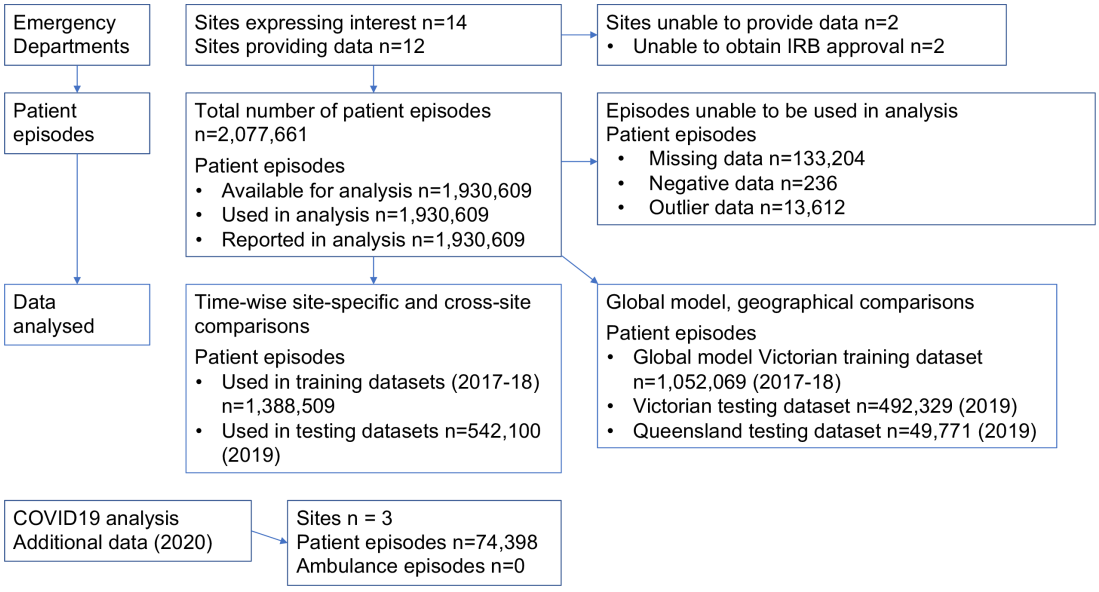


Figure 1 Participant flow through the study.

median site range of 24–54 min for triage-to-provider time (figure 2). Distribution of outcomes is right skewed but we did not apply any transformation to require positive predictions, since predictions of negative values are expected to be rare events and can be replaced with zero in deployment.[7]

Main results

Internal validation of site-specific models (same hospital, using later time period)

The performance rankings produced by the Scott-Knott effect size difference test showed that Random Forests and Linear Regression performed the best (first rank) for all studied hospitals (n=12) followed by Elastic Net (n=11) and Rolling Average (n=5). Random Forests, Linear Regression, and Elastic Net outperformed Rolling Average at seven hospitals. The MAE of Random Forests varied from 22.6 min (95% CI 22.4 to 22.9) for H7 to 44.0 min (95% CI 43.4 to 44.4) for H2. The distributions of the AE of internal validation for ED wait-time (triage-to-provider) prediction are shown (figure 3). The prediction models predicted a wait time to within ±30 min of the actual wait time between 40% and 63% of the time, depending on the site.

The performance differences between Random Forests and Linear Regression were negligible for all hospitals according to Cliff's delta effect size. Rolling average models consistently underestimated patient wait times. More details regarding the actual errors are shown in online supplemental appendix 1C.

Variable importance analysis

Triage Category (median importance score=65%, IQR (54%–74%)), Arrival Time (Hour) (median importance score=15%, IQR (12%–25%)), and the average wait time of the last k-patients (median importance score=15%, IQR (7%–21%)) were the three most important Random Forest predictive variables across all sites. The distributions of importance scores are shown in figure 4.

Table 1 Emergency department and patient demographics

ED	Type of ED (AIHW)*	Patients in training dataset (n)	Patients in testing dataset (n)	Females (n,%)	Admissions (n,%)	Median age of patients (median, IQR)	Ambulance patients (n, %)
Angliss	Medium metro	76 709	39 895	60 332 (52%)	41 933 (36%)	35 (IQR 16–58)	20 038 (17%)
Box Hill	Large metro	130 499	67 103	101 365 (51%)	115 008 (58%)	47 (IQR 23–72)	65 575 (33%)
Cabrini	Private (not-for-profit)	45 314	28 138	40 857 (56%)	32 900 (45%)	Not available	Not available
Casey	Medium metro	151 227	61 062	112 887 (53%)	42 233 (20%)	35 (IQR 18–56)	43 850 (21%)
Clayton (adult)	Major	132 602	52 868	95 032 (51%)	60 963 (33%)	53 (IQR 35–73)	68 678 (37%)
Clayton (children)	Specialist children's	75 271	35 298	48 115 (44%)	15 973 (14%)	5 (IQR 3–10)	12 117 (11%)
Dandenong	Large metro	165 795	61 470	112 526 (49%)	61 066 (27%)	42 (IQR 25–64)	73 891 (33%)
Gold Coast University Hospital	Major	215 034	29 426	121 449 (50%)	Not available	36 (IQR 19–59)	Not available
Maroondah	Major	109 431	55 355	82 013 (49.77%)	81 652 (50%)	43 (IQR 21–65)	56 002 (34%)
Robina	Medium metro	121 406	20 345	72 561 (51%)	Not available	43 (IQR 23–66)	Not available
St Vincent's	Major	97 575	52 671	68 413 (46%)	65 693 (44%)	52 (IQR 37–76)	58 831 (39%)
Werribee	Medium metro	67 646	38 469	56 937 (54%)	40 953 (39%)	39 (IQR 25–60)	24 147 (23%)
Total		1 388 509	542 100	971 487 (50%)	558 374 (29%)	43 (IQR 23–65)	423 129 (22%)

AIHW, Australian Institute of Health and Welfare; ED, emergency department.

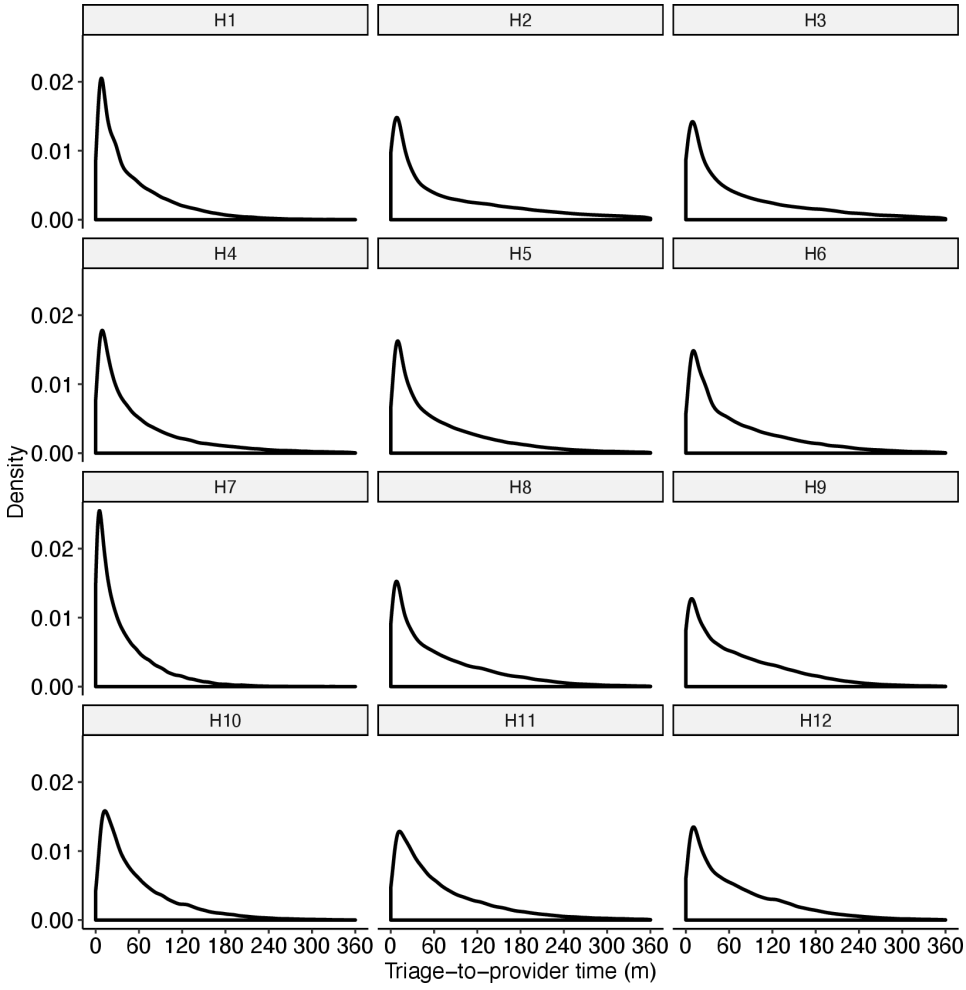
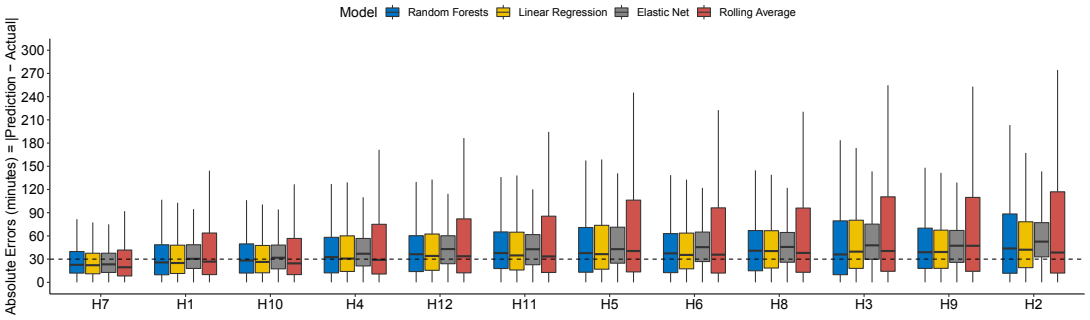


Figure 2 Site distribution of triage-to-provider wait times.

Simplified models

Simplified models were built with the top-ranked variables, which accounted for 95% of the relative variable importance. They demonstrated similar accuracy to full models with all variables. Performance differences between Random Forest simplified models and full models were not statistically significant and had negligible effect sizes for all hospitals. Distributions of the AE of these simplified models are shown in online supplemental appendix 1D.

Figure 3. Distributions of the Absolute Errors (AE) for triage−to−provider time (Internal Validation, site−specific models).



Box and whisker plots of median absolute errors of the wait−time estimates. The horizontal bold black line within each boxplot indicates the median value of the absolute errors. The box describes the interquartile range (IQR, Q1 and Q3). The lower whisker extends from Q1 to the lowest value at most Q1 − 1.5xIQR. The upper whisker extends from Q3 to the highest value at most Q3 + 1.5xIQR. The horizontal dashed line indicates the acceptable threshold of +/− 30 minute wait time estimate errors. Each box colour represents a different model generation technique.

Figure 3 Distributions of absolute errors for triage-to-provider wait times (internal validation, site-specific models). Box and whisker plots of median absolute errors of the wait-time estimates. The horizontal bold black line within each boxplot indicates the median value of the absolute errors. The box describes the IQR (Q1 and Q3). The lower whisker extends from Q1 to the lowest value at most Q1 – 1.5×IQR. The upper whisker extends from Q3 to the highest value at most Q3+1 5×IQR. The horizontal dashed line indicates the acceptable threshold of ±30 minute wait time estimate errors. Each box colour represents a different model generation technique.

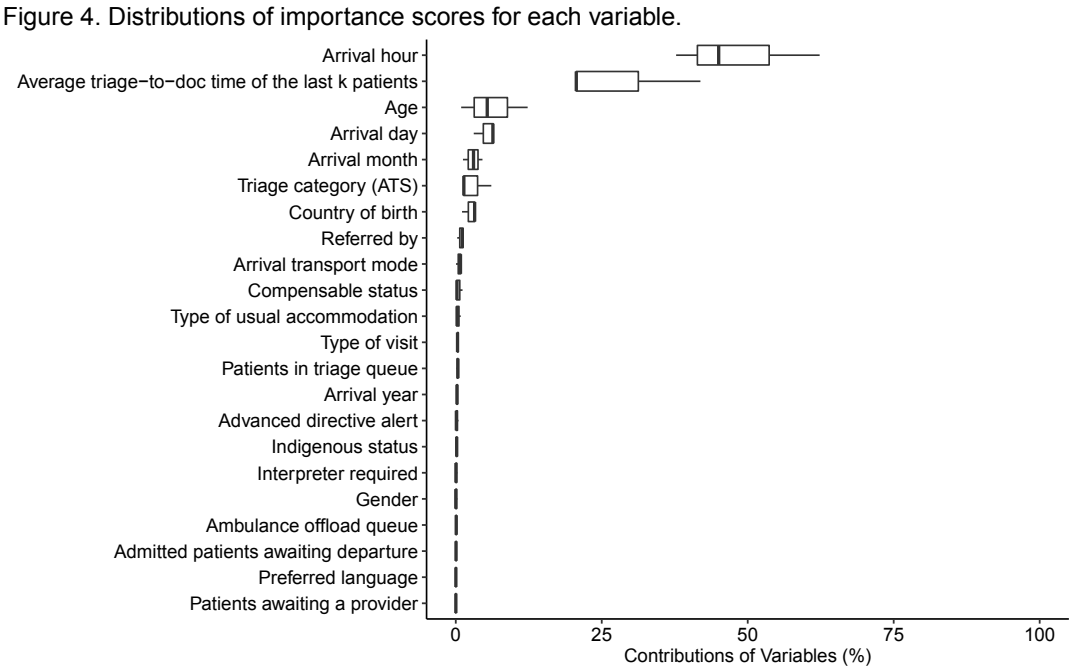


Figure 4 Distribution of importance scores for each variable.

Site-specific model performances at other single sites

Single site models perform better for the specific hospital they were developed for, compared with other sites. Out of 132 pairwise Random Forest combinations, 119 combinations had statistically significant performance differences with negligible to medium effect size. Ninety-seven (~82%) yielded higher errors by 0.02–39.6 min and 22 (~20%) yielded lower errors by 0.01–12.4 min compared with their site of origin. This suggests that site-specific patient wait time prediction models are not transferable to different hospitals. The distributions of AE for triage-to-provider time prediction are shown in online supplemental appendix 1E.

Global model performance

Multihospital global models that were constructed from Victorian 2017–2018 data performed similarly when tested with 2019 data from hospitals in Victoria and Queensland. The global models didn't perform as well as site-specific models. The MAE of these global models varied from 32.2 min (95% CI 32.1 to 32.3) for Linear Regression to 42.6 min (95% CI 42.5 to 42.7) for Elastic Net when tested with Victorian hospital data and varying from 36.1 min (95% CI 35.6 to 36.6) for Linear Regression to 42.4 min (95% CI 42.1 to 42.7) for Elastic Net when tested with patients from hospitals in Queensland. The distributions of the AE of these global models are shown in figure 5.

Impact of COVID-19 on model accuracy

Three hospitals provided 2020 data (n=74 398) covering some of the reduced attendances COVID-19 period in Victoria. We observed that patient wait time models that were built using past data from 2017 and 2018 still performed at reasonable accuracy with MAE differences ranging from 0.03 to 6.0 min. Though these performance differences are statistically significant, except for Random Forests at H10 and Elastic Net at H12, the effect sizes were all negligible for these hospitals. Three models yielded higher errors by 1.7–5.00 min and four yielded lower errors by 0.5–6.0 min. Distributions of the AE of models during COVID-19 are shown in online supplemental appendix 1F.

Limitations

Limitations of the study include using only administrative demographic and ED flow data and using only Australian data. There are no direct measures of resource availability or processes within the ED (eg, nursed cubicles, streaming within the ED), hospital capacity (eg, available beds) or community resourcing (eg, ambulances to transport patients to nursing homes). There were also no measurements of patient comorbidity or diagnosis used in the models. Inclusion of this information may improve prediction accuracy in future models. Excluding did not wait patients or including triage category 1 patients from the study may have over or underestimated true wait times. Additionally, we do not have information about how this model would perform during

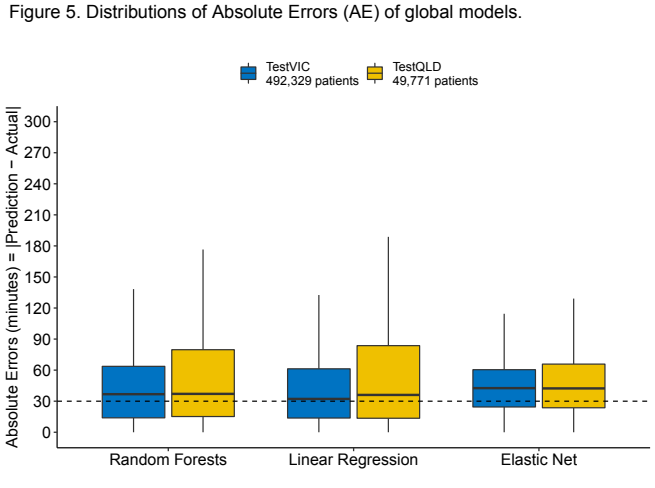


Figure 5 Distribution of absolute errors of global models.

a disaster with a rapid surge in attendances. Models present general estimates of patient wait times for those arriving at the ED, and they do not generate individualised wait times for each patient.

Models can generate nonsensical outputs, for example, linear regression can generate negative predictions which do not make practical sense. In practice, negative predictions can be replaced with 0. We observe the best prediction results in H7 which is the only hospital where wait-time predictions, from an in-house developed model, are shared on their website and on site. This may have affected the behaviour of lower acuity patients in choosing to visit H7 with more flexibility, at times when less waiting is expected, which could have resulted in a more homogenous—and easier to model—distribution than the other hospitals in our sample. Machine learning models may or may not introduce or amplify bias in healthcare. There are currently no reliable ways of testing for bias in machine learning when applied to healthcare datasets (personal communication, Dr Aldeida Aleti, Monash University) and so we are unable to determine if these outputs are biased for or against any particular group of patients.

DISCUSSION

In summary, using emergency patient demographic and flow data from 12 studied hospitals, it is possible to build triage-to-provider predictive waiting time models to an accuracy of ±22.6 to 44.0 min. Predictions were within ±30 min of the actual wait time between 40% and 63% of the time, varying by site. The best performing models used Random Forests and linear regression methods for triage-to-provider prediction. Average wait time of the last k-patients, triage category and patient arrival time were the most important predictor variables. Accuracy is reduced when a model developed for one site is used at another site or a global model (developed by multiple sites) is used. When special events occur such as COVID-19 reduction in attendances, prediction accuracy is maintained.

The accuracies obtained for triage-to-provider times (23–44 min) are less than those reported from a Singapore study.[7] Sun *et al* built different models for each acuity level and this approach to prediction at a finer granularity level may explain the performance difference. They omitted acuity category 1 and reported accuracies of 11.9 min for acuity category 2 and 15.7 min for acuity category 3. We chose a single prediction based on consumer feedback that triage categories are not understood by patients and families, but this may not suit all populations, especially as data and health literacy increases.[5] Ang *et al* modelled low-acuity patients only and reported performance in terms of mean squared error, arguing that median based measures tend to underestimate due to right skewed distribution of wait-times.[8] We observed this with Rolling Average models, but not others. Ang *et al* reported 9.4 min for non-absolute median error, which we outperform at a range of 1.1–8.9 min depending on the site, even after inclusion of all triage categories in the analyses.[8]

To our knowledge, this is the one of few emergency medicine prediction model studies to date, to undertake broad internal (between sites, global model) and external (temporally, geographically, COVID-19) validations. Importantly, we found that using a model derived from data from one site can be used at new sites, but will come at a cost in accuracy, prompting caution prior to promoting a 'one size fits all' model. Similarly, global models trained using aggregate data from all sites may reduce time and cost spent developing individual models but are less accurate at individual sites. We should be wary of seemingly accurate models which have only been tested on data similar to that on which they were trained. External validation of data science models should be undertaken prior to implementation at new jurisdictions, particularly where clinical care decisions might be assisted by machine learning algorithms or models.

This is the first literature describing how models perform during unusual events. We demonstrated that COVID-19 lockdowns did not have a negative impact on model accuracy. In Australia, this period of time was one of both significantly reduced emergency attendances and reduced productivity for physicians due to the increased complexity of managing patients and departments.[19] These data do not cover periods of surge.

Qualitative work has shown that patients want access to wait times and would use times to address a large variety of needs.[5] This study has shown that it is possible to predict approximate wait times; however, it has also demonstrated that the range of predictions is less accurate than desired by stakeholders. Some information about wait times may still be useful to patients, even if the prediction range is broad. Overestimated and underestimated predictions may be perceived differently by patients and families. Overestimated predictions may be perceived positively if patients wait a shorter time than predicted, but could deter patients from seeking care. Random Forests tend to overestimate more than linear regression. Rolling average models underestimated wait times the most (figure 5) and could either make emergency flow seem better than it is or generate dissatisfaction from patients when waits exceed predictions.[20–24]

In summary, using limited data available at point-of-care, wait times can be predicted to ±23–44 min. Models should be individually built for each hospital and are likely to perform the same during COVID-19 like conditions.

Author affiliations
[1]Emergency Department, Casey Hospital, Berwick, Victoria, Australia
[2]Health Services, Faculty of Medicine Nursing and Health Sciences, Monash University, Melbourne, Victoria, Australia
[3]Emergency Department, Cabrini Institute, Melbourne, Victoria, Australia
[4]Department of Software Systems and Cybersecurity, Monash University, Melbourne, Victoria, Australia
[5]MADA, Monash University, Clayton, Victoria, Australia
[6]Emergency Department, Austin Health, Heidelberg, Victoria, Australia
[7]Department of Emergency Medicine, St Vincent's Hospital Melbourne Pty Ltd, Fitzroy, Victoria, Australia
[8]Medicine, Dentistry and Health Sciences, The University of Melbourne, Melbourne, Victoria, Australia
[9]Biostatistics, Cabrini Health, Malvern, Victoria, Australia
[10]Faculty of Medicine Nursing and Health Sciences, Monash University, Clayton, Victoria, Australia
[11]Ambulance Victoria, Doncaster, Victoria, Australia
[12]Community Emergency Health and Paramedic Practice, Monash University, Melbourne, Victoria, Australia
[13]Emergency Program, Monash Health, Clayton, Victoria, Australia
[14]School of Clinical Sciences, Monash University, Melbourne, Victoria, Australia
[15]Emergency Medicine, Eastern Health, Melbourne, Victoria, Australia
[16]Eastern Health Clinical School, Monash University, Melbourne, Victoria, Australia
[17]Emergency, Gold Coast Hospital and Health Service, Southport, Queensland, Australia
[18]Griffith University School of Medicine, Gold Coast, Queensland, Australia
[19]Faculty of Information Technology and Electrical Engineering, University of Oulu, Oulu, Pohjois-Pohjanmaa, Finland

Twitter Gabriel Blecher @gabyblech and Amy Sweeny @EpidemicAmy

Acknowledgements Lisa Kuhn, Anne Spence, Cathie Piggot: governance assistance. John Papatheohari, David Rankin: project sponsors and advisors. Mrs Katarina Tomka: project facilitator, Monash University.

Collaborators Rachel Rosler: network sponsor, Melanie Stephenson: literature review, Kim Hansen: risk advisor, Ms Ella Martini: consumer, Dr Hamish Rodda: emergency informatics advisor, project sponsor, Dr Judy Lowthian: district nursing researcher.

Contributors Principal investigator: KW. Funding: KJ, KW, MB-M. Study design and protocol: KW, BT, CT, JJ, WW. Study protocol revisions: all authors. Ethics/governance: KW, AL. Site chief investigators: HA, GB, BP, KW, AS. Data collection: AL, HA, BP, KW, AS. Data analysis: JJ, CT, BT. Manuscript: KW, JJ, CT, BT. Manuscript revisions: all authors. Manuscript guaranteed by KW and BT.

Funding The Australian government, Medical Research Future Fund, via Monash Partners, funded this study. Researchers contributed in-kind donations of time. The Cabrini Institute and Monash University provided research infrastructure support.

Competing interests Some authors and collaborators are emergency physicians or directors, and others work in community health (prehospital and district nursing). One collaborator is a consumer. The Australian government, Medical Research Future Fund, via Monash Partners, funded this study. Researchers contributed in-kind donations of time. The Cabrini Institute and Monash University provided research infrastructure support.

Patient consent for publication Not required.

Ethics approval The study received Monash Health ethics committee approval (RES-19-0000-763A).

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement No data are available.

ORCID iDs
Katie Walker http://orcid.org/0000-0002-5313-5852
Gabriel Blecher http://orcid.org/0000-0001-8537-2011
Amy Sweeny http://orcid.org/0000-0001-8392-5612

REFERENCES

1 Grafstein E, Wilson D, Stenstrom R, *et al*. A regional survey to determine factors influencing patient choices in selecting a particular emergency department for care. *Acad Emerg Med* 2013;20:63–70.
2 He J, Toloo G-S, Hou X-Y, *et al*. Qualitative study of patients' choice between public and private hospital emergency departments. *Emerg Med Australas* 2016;28:159–63.
3 Xie B, Youash S. The effects of publishing emergency department wait time on patient utilization patterns in a community with two emergency department sites: a retrospective, quasi-experiment design. *Int J Emerg Med* 2011;4:29.
4 Shearer FM, Bailey PM, Hicks BL, *et al*. Why do patients choose to attend a private emergency department? *Emerg Med Australas* 2015;27:62–5.
5 Walker K, Stephenson M, Loupis A, *et al*. Displaying emergency patient estimated wait times: a multi-centre, qualitative study of patient, community, paramedic and health administrator perspectives. *Emerg Med Australas* 202010.1111/1742-6723.13640. [Epub ahead of print: 28 Sep 2020].
6 Strobel S, Ren KY, Dragoman A, *et al*. Do patients respond to posted emergency department wait times: time-series evidence from the implementation of a wait time publication system in Hamilton, Canada. *Ann Emerg Med* 2021. doi:10.1016/j.annemergmed.2021.04.009. [Epub ahead of print: 17 Jun 2021].
7 Sun Y, Teow KL, Heng BH, *et al*. Real-Time prediction of waiting time in the emergency department, using quantile regression. *Ann Emerg Med* 2012;60:299–308.
8 Ang E, Kwasnick S, Bayati M. Accurate emergency department wait time prediction. *M&SOM* 2016;18:141–56.
9 Arha G. *Reducing wait time prediction in hospital emergency room: lean analysis using a random forest model*. University of Tennessee, 2017.
10 Senderovich A, Beck JC, Gal A, *et al*. Congestion graphs for automated time predictions. *Proc Conf AAAI Artif Intell* 2019;33:4854–61.
11 State government of Victoria. VEMD user manual 2018-19, section 3, data definitions. in: department of health and human services. Melbourne, Australia: department of health and human services, 2018. Available: https://www2.health.vic.gov.au/about/publications/policiesandguidelines/vemd-manual-2018-19-sec-3-data-definitions [Accessed 28 Nov 2020].
12 Gori M. *Machine Learning : A Constraint-Based Approach*. 1 edn. Amsterdam: Elsevier, 2018.
13 Dong J, Yom-Tov E, Yom-Tov GB. The impact of delay announcements on hospital network coordination and waiting times. *Management Science* 2019;65:1969–94.
14 Hyndman RJ, Hyndman RJ. *Forecasting : principles and practice*. 2 edn, 2018.
15 Cho I, Boo E-H, Chung E, *et al*. Novel approach to inpatient fall risk prediction and its Cross-Site validation using Time-Variant data. *J Med Internet Res* 2019;21:e11505.
16 Tantithamthavorn C, McIntosh S, Hassan AE, *et al*. The impact of automated parameter optimization on defect prediction models. *IEEE Transactions on Software Engineering* 2019;45:683–711.
17 Cliff N. Dominance statistics: ordinal analyses to answer ordinal questions. *Psychol Bull* 1993;114:494–509.
18 Romano J, Kromrey J, Coraggio J. Appropriate statistics for ordinal level data : Should we really be using t-test and Cohen's d for evaluating group differences on the NSSE and other surveys? Florida Association of Institutional Research Annual Meeting, 2006:1–33.
19 Lim A, Gupta N, Lim A, *et al*. Description of the effect of patient flow, junior doctor supervision and pandemic preparation on the ability of emergency physicians to provide direct patient care. *Aust Health Rev* 2020;44:741–7.
20 Thompson DA, Yarnold PR, Williams DR, *et al*. Effects of actual waiting time, perceived waiting time, information delivery, and expressive quality on patient satisfaction in the emergency department. *Ann Emerg Med* 1996;28:657–65.
21 Hedges JR, Trout A, Magnusson AR. Satisfied patients exiting the emergency department (speed) study. *Acad Emerg Med* 2002;9:15–21.
22 Sonis JD, Aaronson EL, Lee RY, *et al*. Emergency department patient experience: a systematic review of the literature. *J Patient Exp* 2018;5:101–6.
23 Reimann M, Lünemann UF, Chase RB. Uncertainty avoidance as a Moderator of the relationship between perceived service quality and customer satisfaction. *J Serv Res* 2008;11:63–73.
24 Soremekun OA, Takayesu JK, Bohan SJ. Framework for analyzing wait times and other factors that impact patient satisfaction in the emergency department. *J Emerg Med* 2011;41:686–92.

Appendix for "Emergency Medicine Patient Wait Time Multivariable Prediction Models: A Multicentre Derivation and Validation Study"

Table of Contents

Appendix 1a. Collected and derived predictor variables.

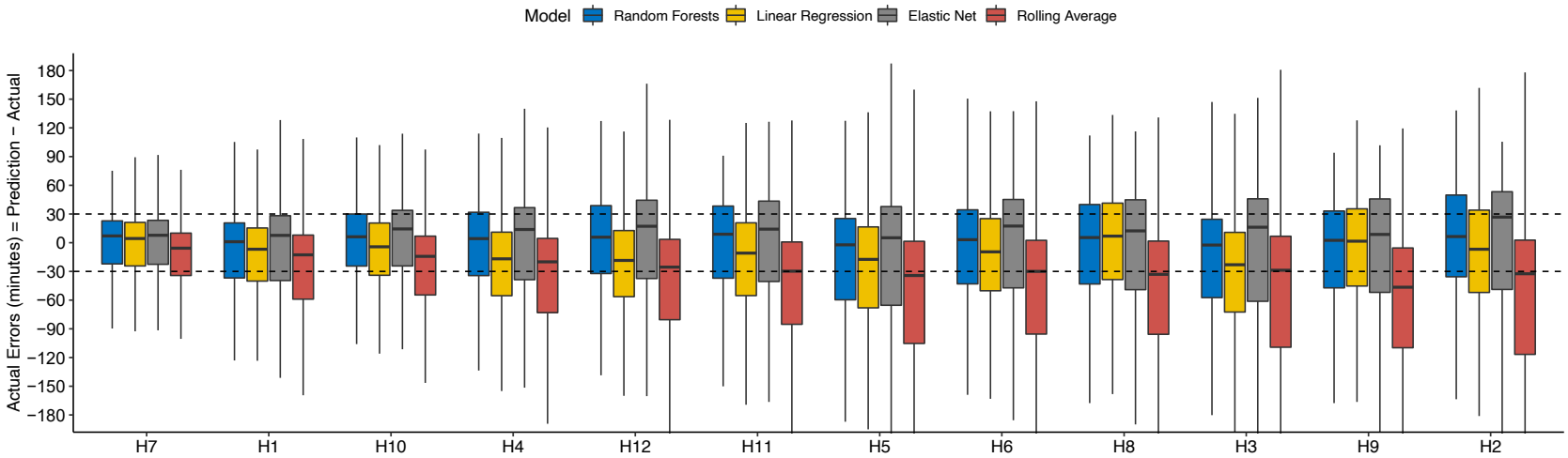
Collected predictor variables	**Type of Variable**
Required to calculate triage-to-provider time	
Triage time	Date/Time
First seen by provider time	Date/Time
Variables required to calculate proposed predictor variables	
Ambulance at destination/door (Front door time only available for ambulance patients)	Date/Time
Ambulance handover complete (Off stretcher time, only for ambulance patients)	Date/Time
Clinical decision to admit time (previous patients)	Date/Time
Date of Birth	Date/Time
Departure time (previous patients)	Date/Time
Other predictor variables available for model development	
Advanced care directive alert	Categorical
Arrival transport mode	Categorical
Campus code	Categorical
Compensable status	Categorical
Country of birth	Categorical
Indigenous status	Categorical
Interpreter required	Categorical
Preferred language	Categorical
Referred by	Categorical
Gender	Categorical
Triage category (Australasian Triage Scale)	Categorical
Type of usual accommodation	Categorical
Type of visit	Categorical
Calculated predictor variables	
Age (We used age rather than date of birth to preserve privacy)	Continuous
Patients in triage queue: The number of patients who arrived before the patient of interest but have not been triaged; requires Arrival and Triage date/time	Continuous
Patients awaiting a provider: The number of patients who have completed triage before the patient of interest, but have not yet seen a provider; requires Triage and First seen by provider Date/Time	Continuous
Admitted patients awaiting departure: The number of patients who have had an admission decision made and have not yet departed from the Emergency Department; requires Clinical decision to admit and departure date/time	Continuous

Ambulance offload queue: The number of ambulance patients arrived, but not yet off their stretcher; requires Ambulance at destination and Ambulance off-stretcher date/time	Continuous
Average wait-time of the last k-patients: The average calculation of the triage-to-provider time of the last k-patients that are seen by the provider prior to the patient of interest arriving; requires Triage-to-provider time of previous patients and First seen by provider of previous patients date/time plus the Triage date/time of the patient of interest	Continuous

Appendix 1b. Incomplete data columns and number of episodes impacted

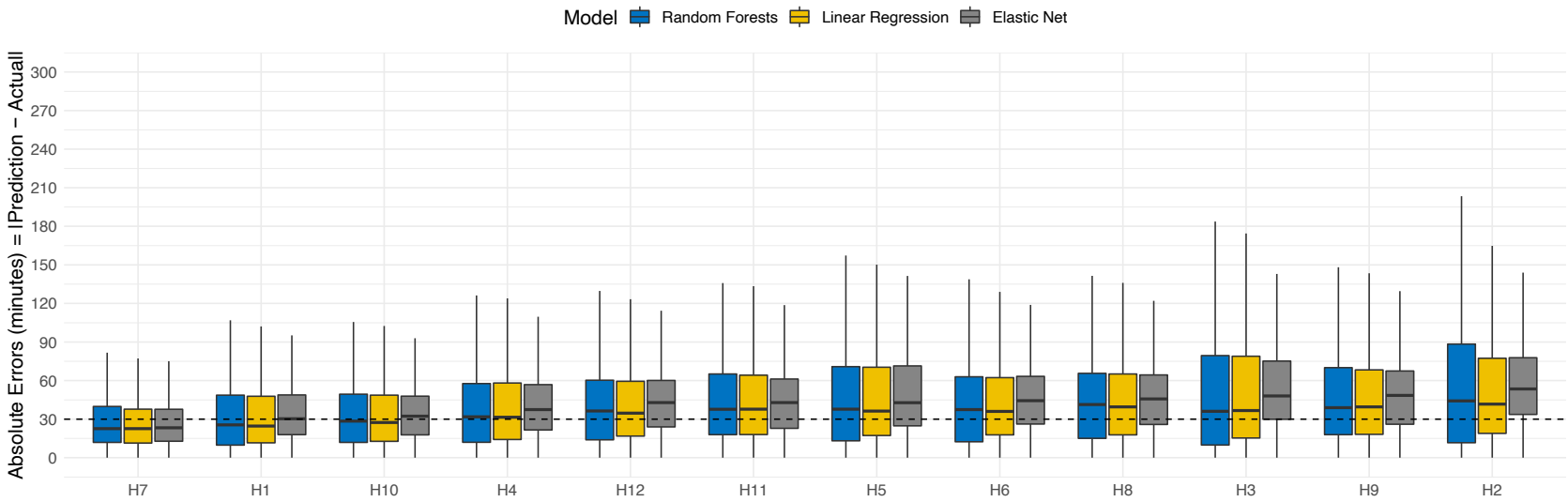
Predictor variable	**Descriptor**	**Number of episodes**
Type of usual accommodation	Unknown/unable to determine	613,474
Arrival transport mode	Other	1,349,317
Compensable status	Compensable status unknown	478,415
Country of birth	Not Stated	772,685
Interpreter required	Not Stated / Inadequately Described	2,942
Referred by	Other	232,848
Preferred language	Not Stated	1,040,001
Indigenous status	Question unable to asked	15,926

Appendix 1c. Full models: Internal validation of each site-specific, full model using its own hospital 2019 testing data; distributions of absolute errors for wait time predictions



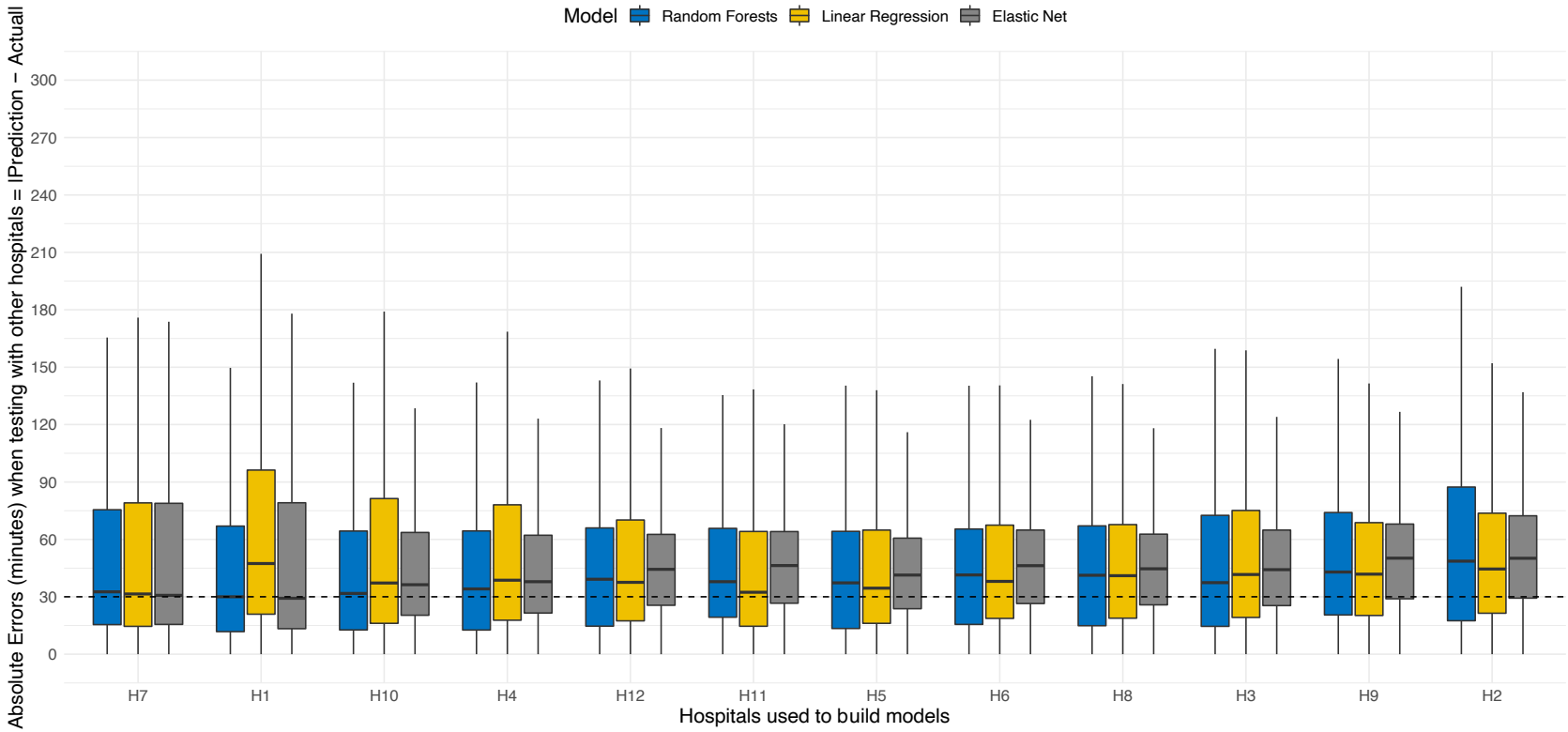
Box and whisker plots of median absolute errors of the wait−time estimates. The horizontal bold black line within each boxplot indicates the median value of the absolute errors. The box describes the interquartile range (IQR, Q1 and Q3). The lower whisker extends from Q1 to the lowest value at most Q1 − 1.5xIQR. The upper whisker extends from Q3 to the highest value at most Q3 + 1.5xIQR. The horizontal dashed line indicates the acceptable threshold of +/− 30 minute wait time estimate errors. Each box colour represents a different model generation technique.

Appendix 1d. Simplified models: Internal validation of each site-specific, simplified model using its own hospital 2019 testing data; distributions of absolute errors for wait time predictions

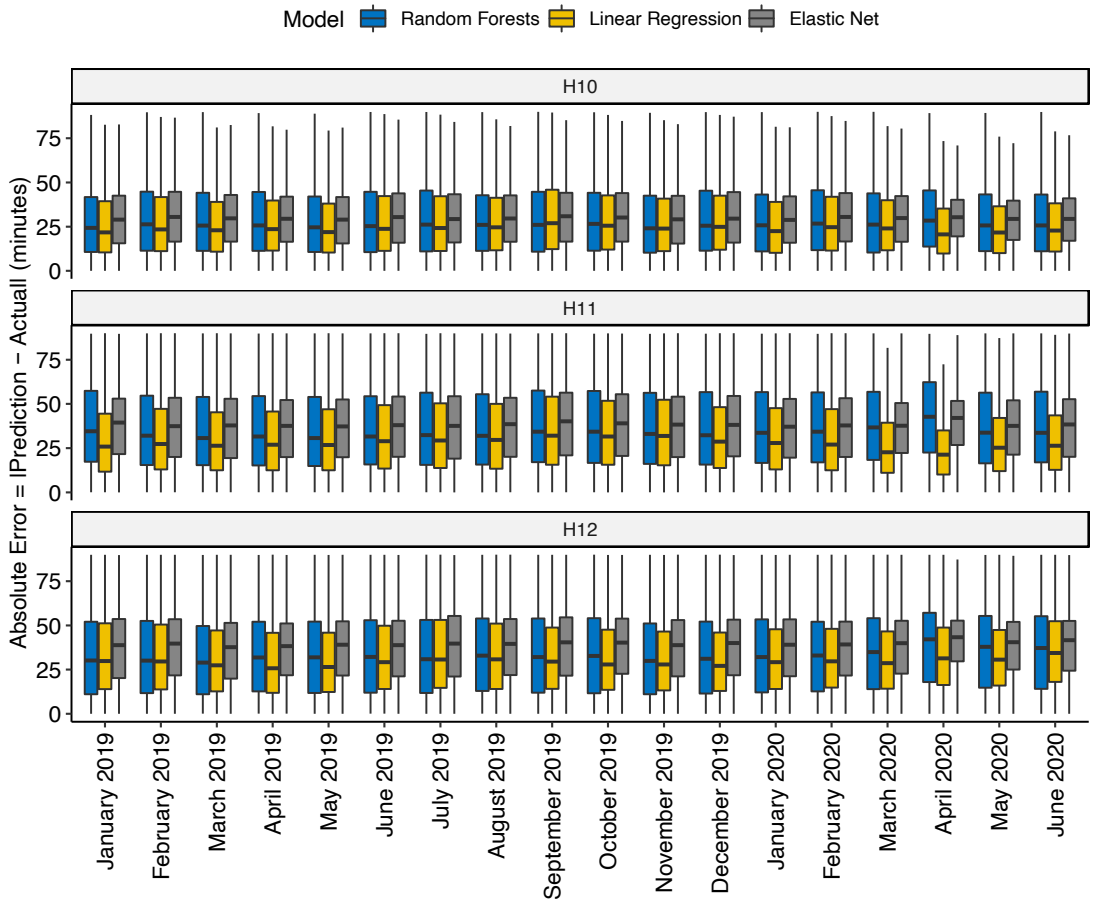


Simplified models built with top−ranked variables that account for 95% of the relative importance.

Appendix 1e. Cross-site, site-specific comparisons: distributions of absolute errors



Appendix 1f. Distributions of absolute errors for wait time predictions before and during COVID19 reduced attendances in 2020



ED presentations over time	H10	H11	H12
January 2019	3164	5598	4478
February 2019	3104	5109	4173
March 2019	3346	5903	4771
April 2019	3325	5440	4613
May 2019	3449	5865	4836
June 2019	3310	5592	4658
July 2019	3323	5643	4722
August 2019	3455	5832	4675
September 2019	3332	5509	4440
October 2019	3367	5685	4736
November 2019	3312	5405	4621
December 2019	3437	5546	4663
January 2020	3259	5665	4565
February 2020	3252	5261	4607
March 2020	3063	7369	4107
April 2020	2200	5538	3177
May 2020	2460	4425	3874
June 2020	2714	4846	3932

*The first wave of COVID−19 in Victoria occured from March 2020