Predicting Ambulance Patient Wait Times: A Multicenter Derivation and Validation Study

Katie J. Walker, MBChB, FACEM*; Jirayus Jiarpakdee, BEng, PhD; Anne Loupis, BSc (Hons); Chakkrit Tantithamthavorn, BEng, PhD; Keith Joe, MBChB, FACEM; Michael Ben-Meir, MBBS, FACEM; Hamed Akhlaghi, FACEM, PhD; Jennie Hutton, FACEM, MPH; Wei Wang, PhD, MD; Michael Stephenson, BHealthSci, GradDipHealthSci; Gabriel Blecher, MBBS, FACEM; Paul Buntine, MBBS, FACEM; Amy Sweeny, RN, MPH (Epid); Burak Turhan, MSc, PhD; on behalf of the Australasian College for Emergency Medicine, Clinical Trials Network

*Corresponding Author. E-mail: katie_walker01@yahoo.com.au.

Study objective: To derive and internally and externally validate machine-learning models to predict emergency ambulance patient door–to–off-stretcher wait times that are applicable to a wide variety of emergency departments.

Methods: Nine emergency departments provided 3 years (2017 to 2019) of retrospective administrative data from Australia. 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.

Results: There were 421,894 episodes analyzed, and median site off-load times varied from 13 (interquartile range [IQR], 9 to 20) to 29 (IQR, 16 to 48) minutes. The global site prediction model median absolute errors were 11.7 minutes (95% confidence interval [CI], 11.7 to 11.8) using linear regression and 12.8 minutes (95% CI, 12.7 to 12.9) using elastic net. The individual site model prediction median absolute errors varied from the most accurate at 6.3 minutes (95% CI, 6.2 to 6.4) to the least accurate at 16.1 minutes (95% CI, 15.8 to 16.3). The model technique performance was the same for linear regression, random forests, elastic net, and rolling average. The important variables were the last k-patient average waits, triage category, and patient age. The global model performed at the lower end of the accuracy range compared with models for the individual sites but was within tolerable limits.

Conclusion: Electronic emergency demographic and flow information can be used to estimate emergency ambulance patient off-stretcher times. Models can be built with reasonable accuracy for multiple hospitals using a small number of point-of-care variables. [Ann Emerg Med. 2021;■:1-10.]

0196-0644/$-see front matter
Copyright © 2021 by the American College of Emergency Physicians.
https://doi.org/10.1016/j.annemergmed.2021.02.010

INTRODUCTION

Background

Deciding where to take ambulance patients for the care of emergency problems is complex and nuanced. Decisions are made after considering information about illness urgency and severity, geographic location, health service capabilities, emergency department queue lengths, patient and paramedic relationships with facilities and their preferences, and costs to patients. Decisions are made with little transparency from health services about their current capacity. Paramedics' off-load times can be prolonged and sick patients remain on stretchers or in ambulances while awaiting a suitable treatment space. This is an important problem for many communities,[1,2] and shorter off-loads are associated with better patient outcomes.[3] For each patient waiting on a stretcher in an emergency corridor, there is also a paramedic crew unable to attend the next emergency call in their region. While there is an increasing consumer interest in displaying emergency patient predicted wait times to improve patient journeys, there is limited predictive information for paramedics about off-load times.[4]

Many emergency departments collect a large volume of electronic point-of-care patient data relating to demographics, flow, and clinical care. This raises the possibility of data science techniques being used to augment patient journeys; however, how to do this remains uncertain. In a community where data from multiple emergency departments are available, knowledge of off-load times could inform optimal patient load-balancing across acute care facilities, potentially improving the performance and safety of both ambulance organizations and emergency departments.

Editor's Capsule Summary

What is already known on this topic

Various methods can be used to predict times to patient evaluation; few have been directed explicitly to patients arriving by ambulance.

What question this study addressed

Can machine learning models be used across a system of hospitals to predict door-to-stretcher time for ambulance patients?

What this study adds to our knowledge

A machine learning process can be built that predicts door-to-stretcher times across different hospitals to a median error of 12 minutes.

How this is relevant to clinical practice

Real time prediction of ambulance off-load times might inform better distribution of ambulance patient loads and so improve time to treatment for this disproportionately ill cohort.

Importance

There is limited knowledge regarding the ambulance off-load prediction problem. There is no literature regarding which predictor variables or techniques should be used, nor is there information about whether a single site-derived model would apply to multiple emergency departments for system-wide implementation. Cooney et al[5,6] considered whether the National Emergency Department Overcrowding Score (NEDOCS)[7,8] may be a useful predictor variable or model, with the association describing the relationship between access block and ambulance ramping. Others have explored centralized ambulance patient distribution optimization.[9,10] Some work reports models for emergency wait times (door to physician), but not for ambulance off-load.[11-14]

Goals of This Investigation

The primary objective of the study was to develop and internally validate predictive algorithms for ambulance patient wait times (door to off stretcher). The secondary objectives included determining the relative importance of each predictor variable and model method, whether one model was appropriate for multiple sites; and whether single-site models were transferable to different emergency departments (external validation).

MATERIALS AND METHODS

Study Design and Setting

This was an observational study using retrospective administrative data to develop, compare, and validate prediction models for ambulance patient off-stretcher wait times at emergency departments. Data from 2017 to 2019 from 9 emergency departments were used for the study.

Mandatory point-of-care emergency patient demographic, flow, and clinical data were collected for every patient by clerks and clinicians in Australia. In Victoria, these defined data populated a state governmental hospital administrative database, the Victorian Emergency Minimum Dataset (VEMD).[15] Data available at the time of triage were used as predictor variables.

There are 24 million residents in Australia, and emergency medicine manages 8 million patient episodes annually. The majority (93%) of Australian residents attend government-funded, public emergency departments with no patient copayments. The remainder attend private emergency departments that require copayments. Ambulance services are funded mainly by government grants, patient transport fees, and membership fees. Ambulance services are independent of hospitals, and some undertake a diversion process when emergency departments are unable to off-load ambulance patients. There are no restrictions on individual choice of emergency departments for paramedics or patients. Regional to tertiary emergency departments were invited to participate if they were part of an academic health science center or were engaged by research networks. Nine Victorian emergency departments participated, comprising 1 pediatric, 3 major, 2 large metropolitan, and 3 medium metropolitan hospitals. No hospitals displayed wait times to the public or paramedics during this study period.

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

Data Sources and Measurements

Electronic medical record software applied time-date stamps to clinician activities (eg, triage). Clerical staff collected demographic data from patients at the initial registration. Paramedics provided emergency departments with their time of arrival for every patient (door) and the paramedic and receiving nurse agreed together on the off-stretcher time at completion of patient bedside handover. Both data were recorded by the receiving nurse in the hospital administrative dataset. Hospital clinical staff recorded data while attending to a patient. The VEMD datasets from each hospital were the primary source of data for this study. The VEMD data were 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, de-identified VEMD data were obtained from 12 hospitals in Australia (mainly Melbourne), and 9 of these hospitals provided ambulance data. Hospital names were randomly replaced by alphanumeric codes prior to analyses (Hospital #1 [H1] through Hospital #9 [H9]). 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 (2019) datasets, to maintain the temporal order based on patient arrival times, noting that predictions relied on the preceding attendances. The training dataset was used for exploratory analysis and training prediction models. The testing dataset was used to internally and externally validate the prediction models. These data were used to evaluate the stability of the model performance during unexpected circumstances.

Variables

The variables used in this study are presented in Appendix 1A (available at http://www.annemergmed.com). The variables collected after triage/registration were excluded from the models, except those required for calculation of the dependent variables. We used 18 predictor variables (12 VEMD and 6 derived) in total.

Outcomes

The primary outcome of this study was the ambulance door–to–off-stretcher wait times for ambulance patients, predicted at triage. The prediction models were built for the global dataset (all sites) and for each individual site. The secondary outcomes included the accuracy of each predictive model (internal validation); identification of the best technique to generate these models; the relative contribution of each variable to the models; and the assessment of how each model performed at different sites (cross-site, external validation). No researcher blinding was undertaken when assessing the outcomes.

The outcome choices and definitions were informed by 113 participants in a multicenter, qualitative study on the same population.[4] Community members, patients, paramedics, and hospital administrators were asked what times they would find most helpful when they were making a decision about which emergency department to attend or when they arrive at an emergency department. The acceptable accuracy of the wait-time estimate (either for door-to-provider and/or ambulance door–to–off-stretcher waits) was reported in 87 of the transcripts. Stakeholders wanted a median accuracy of ±30 minutes (interquartile range [IQR], 10 to 30 minutes).

Analysis

Study size. We opted for a time-based sampling similar to previous studies of wait-time prediction that have used time periods ranging from 1 month to 1 year.[12-14] We therefore requested 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 to use a minimum of 30 to 50 data points per variable for predictive modeling 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. This meant that we removed ambulance patients whose ambulance time at destination or ambulance off-load time were not recorded (n=1313 [0.3%]). Other missing values were replaced with the corresponding "unknown" or "other" category from the VEMD descriptors. The total number of unique patient episodes in which the value of at least 1 of the predictor variables was "unknown/other" was 70,900 (17%) and involved 8 predictor variables. The number of data rows with "unknown/other" values for each predictor variable is reported in Appendix 1B.

We removed patient data that were generated by administrative data entry errors or in which the wait time was negative (n=656 [0.2%]) or exceeded the maximum of 360 minutes plus the predefined statistical outlier threshold value (n=1237 [0.3%]), defined as 1.5 times the interquartile range (IQR=Q3−Q1).

Standardizing and encoding the data. One-hot encoding[16] was applied to all categorical variables to generate dummy variables prior to prediction model development as all categorical variables were nominal with the exception of the triage category. For example, applying one-hot encoding to Interpreter Required (where the possible values are Yes, No, and Not Stated) would generate 3 dummy variables (ie, IR_Yes, IR_No, and IR_NotStated). Patients who required interpreters would have an IR_Yes value of 1 while having IR_No and IR_NotStated values of 0. One-hot encoding is useful in machine learning when there is no ordinal relationship between categorical variables. We assessed that it was preferable to lose order information for triage by applying one-hot encoding than to treat the triage category as a continuous variable as the distances within levels of triage category were nonlinear.

Model building and recalibration. Guided by door-to-provider wait-time prediction literature,[11-13,17] we used 3 statistical and machine-learning techniques (ie, linear regression, random forests, and elastic net regression) and a rolling average approach (ie, the average calculation of the outcome of previous "k" observations). We included all predictor variables in the model construction and undertook a post hoc variable importance analysis. We rebuilt the models with the most important variables only and compared the performance of the simplified models with that of the initial models.

The "last-k" variable is the average ambulance door–to–off-stretcher wait time for the last "k" patients off-loaded. It does not include those still awaiting off-load who are accounted for by the number of patients in the ambulance off-load queue. To determine the appropriate value of "k" for this study, we performed a sensitivity analysis by observing the performance of the prediction models constructed using different k values (ie, 4, 8, and 10). We found that the performance differences were statistically indistinguishable across different k values. We selected the k value of 4 for this study, which produced models with the best performance.

Multicollinearity often exists in continuous predictor variables. Thus, we used a variance inflation factor (VIF) analysis[18] with its threshold value of 5 to detect multicollinearity in the studied datasets. We found that none of the studied continuous predictor variables had VIF scores of more than 5, suggesting that multicollinearity did not exist in our studied datasets.

Validation. We used a time-wise hold-out validation approach for internal validation.[19] Patient records were sorted for each hospital using their ambulance at destination (door) time. A fixed time-based setting was used to generate training and testing samples. Patient data from 2017 and 2018 were used to construct prediction models, whereas 2019 patient data were used to evaluate the prediction models. We then constructed site-specific prediction models in which each model was derived and evaluated using its own hospital data.

For external validation, we tested the models in 2 ways. The first was a pairwise, cross-site approach.[20] The site-specific models were tested using 2019 data from other hospitals (eg, train with Hospital A data from 2017 and 2018, then test with Hospital B data from 2019), resulting in 72 pairwise combinations. The second was a global prediction model in which the model was constructed and evaluated with patient data from all hospitals.

To assess the model performance, we calculated the absolute errors between the actual time and the predicted time for all models and hospitals. We then calculated the median of these distributions of absolute errors to identify the best model for and across the 9 hospitals.

Simplified models. To construct simplified models, we first constructed predictive models with all predictor variables. Then, we conducted a post hoc variable importance analysis. We selected the top-ranked, most important variables that accounted for 95% relative variable importance as a simplified subset of predictor variables for constructing simplified models.

Statistical methods. The Scott-Knott effect size difference test[21] was used to identify the performance ranking, based on median absolute error, of the prediction models for internal validation. The Scott-Knott effect size difference test 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 version 2.0.3. Mann-Whitney *U* tests were used to identify whether the performance (median absolute error) difference between 2 models was statistically significant, then Cliff's delta tests,[22] $|d|$, were used to measure the effect size. The interpretation of Cliff's delta values is as follows: $|d|<0.147$ negligible, $|d|<0.33$ small, $|d|<0.474$ medium, otherwise large.[23] We used the cliff.delta function of the effsize R package version 0.7.8 for calculating Cliff's delta. For all statistical tests, we used a statistical significance level of $\alpha=0.05$ and sought non-negligible effect sizes.

RESULTS

Characteristics of Study Subjects

Nine emergency departments contributed data. Two sites were unable to obtain ethics approval (regional referral, medium metropolitan), and 3 sites were unable to obtain ambulance patient data (private, major, medium metropolitan). Participant flow through the study is presented in Figure 1. The demographics of the patients are presented in the Table. The total number of ambulance patient episodes after data cleaning and outlier removal was 421,894, with 276,390 in the training and 145,504 in the test datasets. Twenty-eight percent of the emergency patients arrived at the hospital by ambulance and 56% of the ambulance patients were admitted.

Off-load wait time proportional distributions were similar among sites, although specific wait times at each site varied, with a median site range of 13 to 29 minutes (Figure 2). Distributions of outcomes were right skewed, but we did not apply any transformation to require positive predictions, because predictions of negative values were expected to be rare events and could be replaced with zero in deployment.[14]

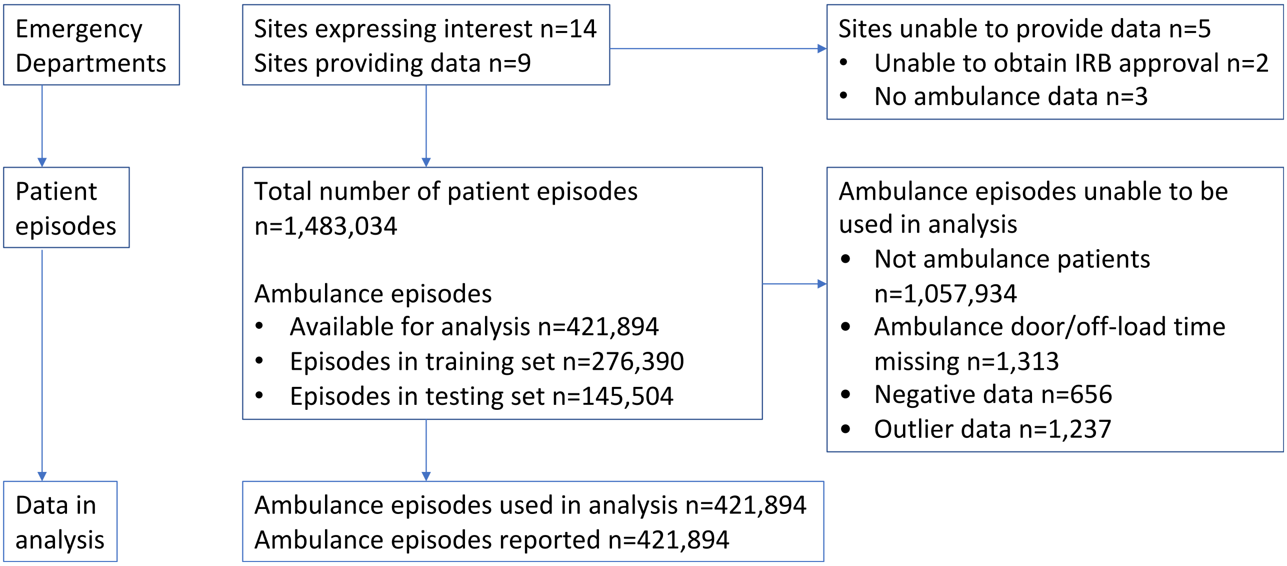

Figure 1. Participant flow through the study. *IRB*, Institutional review board.

Main Results

Internal validation (global model). Global models generated by all techniques performed with similar accuracy and were statistically indistinguishable according to the Scott-Knott effect size difference test. The median absolute errors of these global models varied from 11.7 minutes (95% confidence interval [CI], 11.7 to 11.8) for linear regression to 12.8 minutes (95% CI, 12.7 to 12.9) for elastic net. The distributions of the absolute errors of the global external validation for ambulance door–to–off-stretcher time prediction are shown in Appendix 1C.

Internal validation (site-specific models). All individual models performed with similar accuracy and were statistically indistinguishable according to the Scott-Knott effect size difference test. Random forest models yielded median absolute errors that varied from 6.3 minutes (95% CI, 6.2 to 6.4) for H1 to 16.1 minutes (95% CI, 15.9 to 16.3) for H2. Despite the lack of statistical difference between the site-specific model performances, these differences may be clinically significant for patients or paramedics. Site-specific models overestimated wait times for 54% (median, 14.3; IQR, 7.4

Table. Emergency department characteristics and ambulance patient demographics.

Emergency Department			Ambulance Patients (n=421,894)				
Emergency Department	**Type of ED (AIHW)**	**Total Patients in ED 2017 to 2019**	**Patients in Training Dataset, No.**	**Patients in Testing Dataset, No.**	**Female Patients, No. (%)**	**Admissions, No. (%)**	**Median Age of Patients, y (IQR)**
Angliss	Medium metro	116,719	12,554	7,474	11,263 (56)	12470 (62)	63 (39, 81)
Box Hill	Large metro	198,824	42,073	23,430	34,764 (53)	54195 (83)	71 (46, 84)
Casey	Medium metro	213,428	29,585	14,154	24,385 (56)	14,286 (33)	55 (32, 78)
Clayton (Adult)	Major	189,699	46,807	21,447	33,533 (50)	32,719 (49)	69 (49, 83)
Clayton (Children)	Specialist children's	110,636	7,745	4,299	5,265 (43)	3,768 (31)	6 (3, 13)
Dandenong	Large metro	229,335	50,815	22,835	37,217 (51)	29,662 (41)	60 (38, 79)
Maroondah	Major	165,261	35,860	20,108	29,876 (54)	40,008 (72)	61 (38, 80)
St Vincents	Major	151,322	36,994	21,773	26,503 (45)	33,094 (57)	61 (41, 86)
Werribee	Medium metro	107,810	13,957	9,984	13,359 (56)	13,105 (55)	58 (37, 79)
Total	**9 sites**	**1,483,034**	**276,390**	**145,504**	**216,165 (52)**	**233,307 (56)**	**62 (38, 81)**

AIHW, Australian Institute of Health and Welfare.

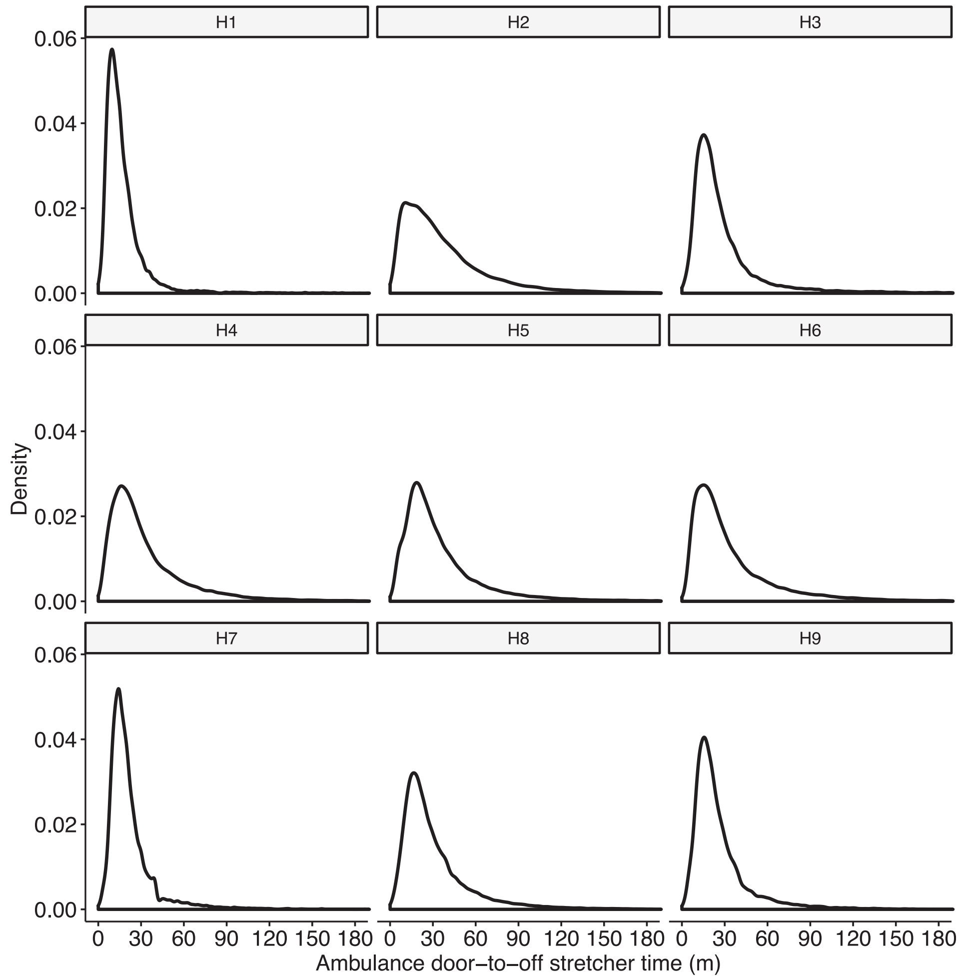

Figure 2. Site distributions of ambulance door–to–off-stretcher time.

to 22.0) and underestimated wait times for 46% (median, 20.2; IQR, 8.0 to 42.6), using H2 as an example. The distributions of the absolute errors and actual errors of site-specific internal validation for ambulance door–to–off-stretcher time prediction are shown in Figure 3 and Appendix 1D.

Variable importance analysis. The average wait time of the last k-patients (median importance score = 45.8%, IQR [32.9%, 51.2%]), triage category (median importance score=16.9%, IQR [6.0%, 28.8%]), and age (median importance score=7.5%, IQR [6.3%, 9.1%]) were the 3 most important predictor variables in terms of percentage of relative importance. The distributions of importance scores at all sites are shown in Figure 4. Variables collected after triage/registration were not included in this variable importance analysis and do not appear in Figure 4 (eg, departure time of the patient of interest). The simplified models demonstrated similar accuracy to that of the full models. More details are shown in Appendix 1E.

External validation (pairwise cross-site comparison). The models performed better for the specific hospital that they were developed for and were less accurate when applied to other hospitals. Out of 72 pairwise combinations, when we applied random forests to other hospitals, 68 combinations had statistically significant differences with negligible to medium effect sizes. Within these 68 combinations, 41 (∼60.3%) yielded higher errors by 0.01 to 9.2 minutes and 27 (∼39.7%) yielded lower errors by 0.01 to 4.9 minutes in comparison with internal (within site) validation. This suggested that the ambulance wait-time models lost accuracy when they were transferred to different hospitals. More details are shown in Appendix 1F.

LIMITATIONS

Limitations of the study included using only administrative demographic and emergency department flow data, and using only Australian data. The accuracy of the data was not verified by the researchers. There were no

ARTICLE IN PRESS

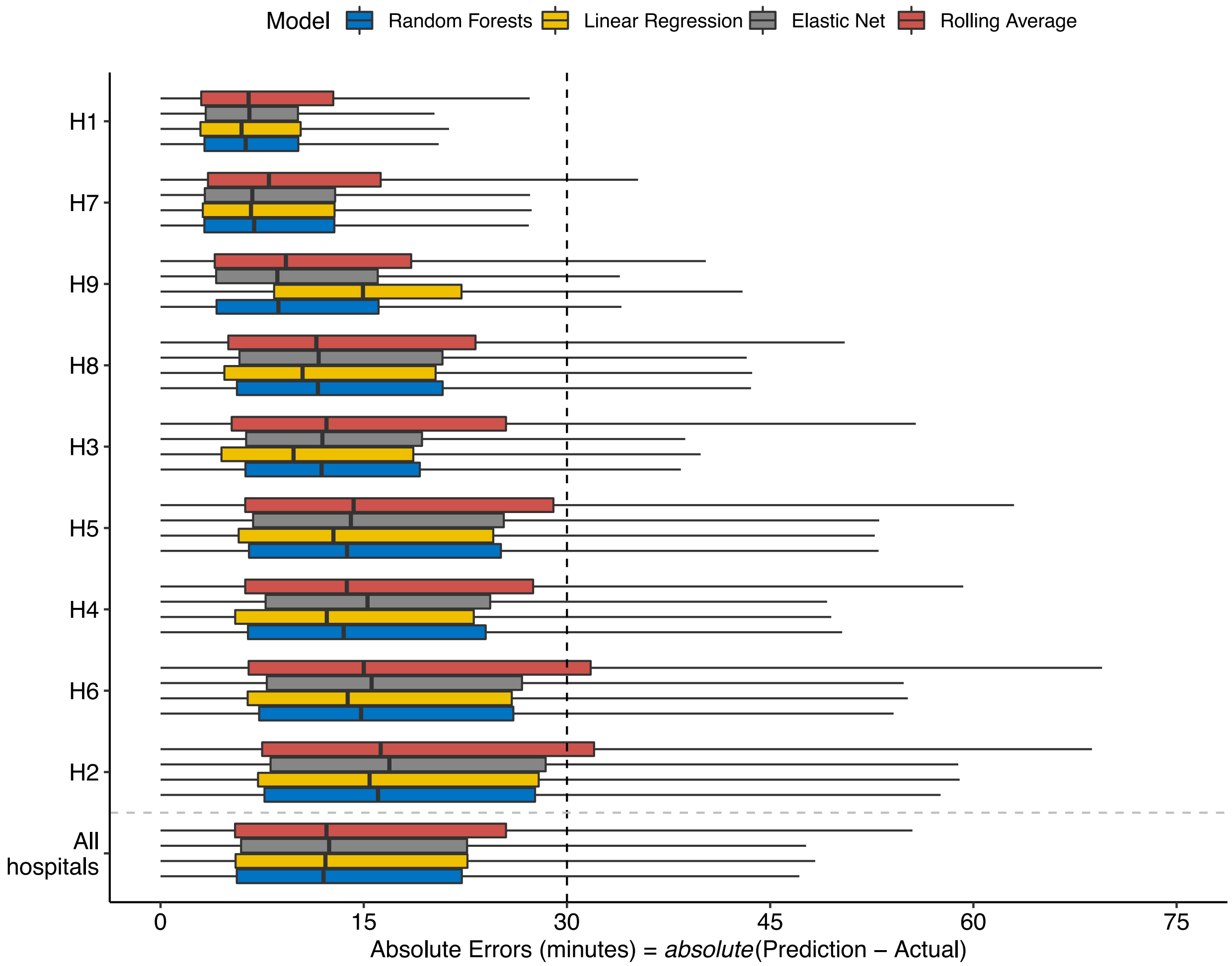

Figure 3. Box and whisker plots of median absolute errors of the ambulance door–to–off stretcher time estimates. The horizontal bold black line within each boxplot indicates the median value of the absolute errors. The box describes the interquartile range (IQR, Q1, and Q3). The lower whisker extends from Q1 to the lowest value at most Q1 − 1.5 × 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 color represents a different model generation technique.

direct measures of resource availability within the emergency department (eg, nursed cubicles), hospital capacity (eg, available beds), or community resources (eg, ambulances to transfer patients back to nursing homes). There were no measurements of patient comorbidity or diagnoses used in the models. Inclusion of this information may improve predictive accuracy in future models. Additionally, we do not have information about how this model would perform during a disaster with a rapid surge in attendances.

The model provided a generalized prediction about the ambulance door–to–off-stretcher time for all patients at a point in time, rather than an individualized prediction for a particular patient checking in. The prediction does not incorporate drive times for ambulances nor predict how many patients are about to arrive at an emergency department; therefore, it will be less accurate when decisions are made at a scene rather than at the door of an emergency department.

Models can generate nonsensical outputs; for example, linear regression can generate negative predictions (negative values of triage-to-provider and ambulance door–to–off-stretcher time) that do not make practical sense. In practice, negative predictions can be replaced with 0.

DISCUSSION

By using emergency ambulance patient demographic and flow data from the 9 studied hospitals, it was possible

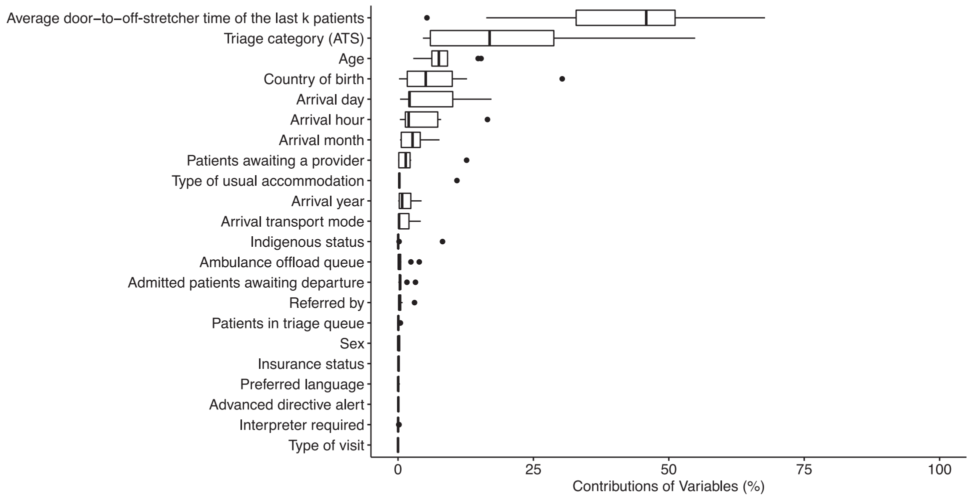

Figure 4. Distributions of importance scores for ambulance door—to—off stretcher time prediction at all sites.

to build ambulance door–to–off-stretcher predictive waiting time models to an accuracy of ±6.3 to 16.1 minutes. More accurate models were built when hospitals used their own data, but a global model can still provide clinically useful information. All model techniques performed about the same. Average wait time of the last k-patients (k=4), triage category, and age were the 3 most important variables for the ambulance door–to–off-stretcher prediction models.

The accuracies obtained by the models may be useful to ambulance organizations and hospitals. It was a simpler problem, generating more accurate algorithms, in comparison with predicting patient wait times. If a community-wide, centralized distribution approach was taken in a jurisdiction with multiple emergency department choices (eg, a large city), total ambulance patient load could be optimized, delivering patients evenly across facilities according to various hospital capacities, capabilities, and locations.[24] Centralized distribution would need to take into account more variables than just ambulance door–to–off-stretcher waits, but these models provide a good starting point for distribution algorithm development. Whether a provider of ambulance services decides to pursue the use of wait-time predictions will depend on a variety of political, legislative, stakeholder, and economic factors. It is important to note that optimization is different from diversion. Optimization spreads loads evenly and continuously, aiming to prevent overloading, whereas ambulance diversion is a response to overcapacity emergency departments and hospitals. Optimization may allow for smoother processing of all patients, with more availability of resources at any given time and therefore fewer periods of crowding and fewer prolonged waits.

Our generalized model accuracy was toward the less accurate end of the performance ranges of individual site models but still within the tolerance of the paramedic stakeholders interviewed prior to the model creation. Decisions about whether one model or multiple models should be deployed are likely to be pragmatic decisions about accuracy versus deployment resources.

Qualitative work has shown that patients want access to wait times and paramedics want access to off-load times. Both paramedics and patients would use predicted times to address a variety of needs once in emergency department corridors.[4] Paramedics would use the information to manage patient needs; logistical, emotional, and physical. They would also use the predictions to manage paramedic needs, both physical and operational. Examples of these include informing patients and families about queue lengths, sourcing new oxygen bottles, toileting patients, and arranging for reinforcement paramedics to attend during waits that cross shift change times. There is little to no out-of-hospital research available on whether, how, and when paramedics would use this information while in the community rather than in the emergency department. We also do not understand how best to visualize or display the off-load times or whether to combine them in an algorithm with information about drive times, hospital capabilities, costs, or patient-specific variables.

In summary, electronic emergency demographic and flow information can be used to approximate emergency ambulance patient off-stretcher times. Models can be built with reasonable accuracy for multiple hospitals using a small number of point-of-care variables.

The authors acknowledge Lisa Kuhn, PhD, Anne Spence, and Cathie Piggot, MPHC, for governance assistance; John Papatheohari, MAICD, and David Rankin (project sponsors and advisors), MHA; Mrs Katarina Tomka (project facilitator), MBM; and the following collaborators: Rachel Rosler (network sponsor), FACEM, Melanie Stephenson (literature review), Kim Hansen (risk advisor), FACEM, Ella Martini (consumer), MPH, Hamish Rodda (emergency informatics advisor, project sponsor), MD, FACEM, and Judy Lowthian (district nursing researcher), PhD.

Supervising editor: Stephen Schenkel, MD, MPP. Specific detailed information about possible conflict of interest for individual editors is available at https://www.annemergmed.com/editors.

Author affiliations: From Cabrini Emergency Department, Malvern, Melbourne, Victoria, Australia (Walker, Joe, Ben-Meir, Blecher); Cabrini Institute, Malvern, Melbourne, Victoria, Australia (Walker, Loupis, Wang); Casey Emergency Department, Berwick, Melbourne, Victoria, Australia (Walker); School of Clinical Sciences at Monash Health, Monash University, Clayton, Melbourne, Victoria, Australia (Walker, Loupis, Blecher); Department of Software Systems and Cybersecurity, Monash University, Clayton, Melbourne, Victoria, Australia (Jiarpakdee, Tantithamthavorn); Monash Art, Design and Architecture, Monash University, Caulfield, Melbourne, Victoria, Australia (Joe); School of Public Health and Preventive Medicine, Monash University, Melbourne, Victoria, Australia (Ben-Meir); Emergency Department, St Vincent's Hospital, Fitzroy, Melbourne, Victoria, Australia (Akhlaghi, Hutton); Medicine, Dentistry and Health Sciences, University of Melbourne, Parkville, Melbourne, Victoria, Australia (Hutton); Faculty of Medicine, Nursing and Health Sciences, Monash University, Melbourne, Victoria, Australia (Wang); Ambulance Victoria, Doncaster, Melbourne, Victoria, Australia (Stephenson); Department of Epidemiology and Preventive Medicine, Monash University, Melbourne, Victoria, Australia (Stephenson); Department of Community Emergency Health and Paramedic Practice, Frankston, Melbourne, Victoria, Australia (Stephenson); Monash Medical Centre, Emergency Department, Clayton, Melbourne, Victoria, Australia (Blecher); Emergency Department, Box Hill Hospital, Eastern Health, Box Hill, Melbourne, Victoria, Australia (Buntine); Eastern Health Clinical School, Monash University, Box Hill, Melbourne, Victoria, Australia (Buntine); Emergency Department, Gold Coast University Hospital, Southport, Queensland, Australia (Sweeny); Faculty of Health Sciences and Medicine, Bond University, Robina, Queensland, Australia (Sweeny); and the Faculty of Information Technology and Electrical Engineering, University of Oulu, Oulu, Finland (Turhan).

Author contributions: KJW was the principal investigator. KJ, KJW, and MBM were responsible for funding. KJW, BT, CT, JJ, and WW developed the study design and protocol. All authors revised the study protocol. KJW and AL provided ethics/governance. HA, GB, PB, KJW, and AS were the site chief investigators. AL, HA, PB, KJW, and AS collected the data. JJ, CT, and BT analyzed the data. KJW, JJ, CT, and BT wrote the manuscript and all authors revised the manuscript. KJW and BT take responsibility for the paper as a whole.

All authors attest to meeting the four ICMJE.org authorship criteria: (1) Substantial contributions to the conception or design of the work; or the acquisition, analysis, or interpretation of data for the work; AND (2) Drafting the work or revising it critically for important intellectual content; AND (3) Final approval of the version to be published; AND (4) Agreement to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Funding and support: By *Annals* policy, all authors are required to disclose any and all commercial, financial, and other relationships in any way related to the subject of this article as per ICMJE conflict of interest guidelines (see www.icmje.org). The Medical Research Future Fund, by Monash Partners, funded this study. Researchers contributed in kind donations of time. The Cabrini Institute and Monash University provided research infrastructure support.

Publication dates: Received for publication November 30, 2020. Revision received January 31, 2021. Accepted for publication February 5, 2021.

REFERENCES

1. Cooney DR, Millin MG, Carter A, et al. Ambulance diversion and emergency department offload delay: resource document for the National Association of EMS Physicians position statement. *Prehosp Emerg Care*. 2011;15:555-561. https://doi.org/10.3109/10903127.2011.608871.
2. Kingswell C, Shaban RZ, Crilly J. Concepts, antecedents and consequences of ambulance ramping in the emergency department: a scoping review. *Australas Emerg Nurs J*. 2017;20:153-160. https://doi.org/10.1016/j.aenj.2017.07.002.
3. Crilly J, Keijzers G, Tippett V, et al. Improved outcomes for emergency department patients whose ambulance off-stretcher time is not delayed. *Emerg Med Australas*. 2015;27:216-224. https://doi.org/10.1111/1742-6723.12399.
4. 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*. https://doi.org/10.1111/1742-6723.13640
5. Cooney DR, Wojcik S, Seth N. Can NEDOCS score be used to predict ambulance offload delay? *Ann Emerg Med*. 2011;58(Suppl):S217. https://doi.org/10.1016/j.annemergmed.2011.06.144.
6. Cooney DR, Wojcik S, Seth N, et al. Evaluation of ambulance offload delay at a university hospital emergency department. *Int J Emerg Med*. 2013;6:15. https://doi.org/10.1186/1865-1380-6-15.
7. Hargreaves D, Snel S, Dewar C, et al. Validation of the National Emergency Department Overcrowding Score (NEDOCS) in a UK non-specialist emergency department. *Emerg Med J*. 2020;37:801-806. https://doi.org/10.1136/emermed-2019-208836.
8. Weiss SJ, Derlet R, Arndahl J, et al. Estimating the degree of emergency department overcrowding in academic medical centers: results of the National ED Overcrowding Study (NEDOCS). *Acad Emerg Med*. 2004;11:38-50. https://doi.org/10.1197/j.aem.2003.07.017.
9. Mengyu L. Designing emergency medical services processes to minimize the impact of ambulance offload delay. Dalhousie University Web site. 2019. Accessed March 26, 2021. http://hdl.handle.net/10222/76289

ARTICLE IN PRESS

10. Yang W, Qiang S, Huang S, et al. Simulation modeling and optimization for ambulance allocation considering spatiotemporal stochastic demand. *Journal of Management Science and Engineering*. 2019;4:252-265. https://doi.org/10.1016/j.jmse.2020.01.004.
11. Senderovich A, Beck JC, Gal A, et al. Congestion graphs for automated time Predictions. *Proc Conf AAAI Artif Intell*. 2019;33:4854-4861.
12. Arha G. *Reducing* Wait Time Prediction In Hospital Emergency Room: Lean Analysis Using a Random Forest Model. Thesis. University of Tennessee; 2017.
13. Ang E, Kwasnick S, Bayati M, et al. Accurate emergency department wait time prediction. *Manuf Serv Oper Manag*. 2016;15:170-175.
14. 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. https://doi.org/10.1016/j.annemergmed.2012.03.011.
15. State government of Victoria. VEMD user manual 2018-19, section 3, data definitions. Accessed November 28, 2020. https://www2.health.vic.gov.au/about/publications/policiesandguidelines/vemd-manual-2018-19-sec-3-data-definitions
16. Gori M. *Machine Learning: A Constraint-Based Approach*. Elsevier; 2018; 5, 6, 86.
17. Dong J, Yom-Tov E, Yom-Tov GB. The impact of delay announcements on hospital network coordination and waiting times. *Manage Sci*. 2019;65:1969-1994.
18. Fox J. *Applied Regression Analysis and Generalized Linear Models*. Sage Publications; 2015.
19. Hyndman RJ, Athanasopoulos G. *Forecasting: Principles and Practice*. 2nd ed. OTexts; 2018.
20. 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. https://doi.org/10.2196/11505.
21. Tantithamthavorn C, McIntosh S, Hassan A, et al. Impact of automated parameter optimization on defect prediction models. *IEEE Transactions on Software Engineering*. 2019;45:683-711. https://doi.org/10.1109/TSE.2018.2794977.
22. Cliff N. Dominance statistics: ordinal analyses to answer ordinal questions. *Psychol Bull*. 1993;114:494-509.
23. Romano J, Kromrey J, Coraggio J, et al. *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. Accessed April 26, 2021. https://www.bibsonomy.org/bibtex/1040942465797dca02d5527fd876088dc
24. 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. https://doi.org/10.1186/1865-1380-4-29.