

Modelling Systemic Effects of Detecting and Reclassifying Under-Triaged Patients in a Pediatric Emergency Department

Helena Coggan

Nathan Georgette

Anne C. Bischops

Guilherme S. Imai Aldeia

William G. La Cava

Computational Health Informatics Program, Boston Children's Hospital, Boston, MA, USA

HELENA.COGGAN@CHILDRENS.HARVARD.EDU

NATHAN.GEORGETTE@CHILDRENS.HARVARD.EDU

ANNE.BISCHOPS@CHILDRENS.HARVARD.EDU

GUILHERME.IMAIALDEIA@CHILDRENS.HARVARD.EDU

WILLIAM.LACAVA@CHILDRENS.HARVARD.EDU

Abstract

Emergency department [ED] triage methods are imperfect and may mistakenly classify critically ill patients as ‘non-urgent’. ML-driven identification of under-triaged patients may prevent delays in life-saving care. Here we analyze a cohort of more than 200k visits to a large urban pediatric emergency department over four years. We use gradient-boosted tree methods to predict life-threatening deterioration in patients initially triaged as non-urgent (AUROC 0.80 [95% CI 0.78-0.81]). Such a model could be used in real-world EDs to reclassify flagged non-urgent patients as urgent. In deployment, however, this reclassification comes at a cost: if too many patients are up-triaged, this will increase competition for finite clinician attention and risk overall delays in care. To assess these tradeoffs, we present a novel discrete event modelling framework to simulate prospective deployment of our model at triage across two years of retrospective ED visits. Our simulations identify risk thresholds at which the model could be used without creating bottlenecks in the system: for instance, at a sensitivity of 49% and specificity of 89%, up-triaging of model-identified patients increased the average length of stay by 11 minutes [8.1-13], but decreased the time to admission for correctly-flagged patients by 23 minutes [10-34]. These results suggest that many cases of critical deterioration in the pediatric emergency department could be prevented through closer monitoring without large downstream effects on patient flow.

Keywords: emergency department, gradient boosting, crowding, simulation, deployment, pediatrics, discrete event modelling

Data and Code Availability This was a retrospective observational study using routinely collected data from the emergency department of a large urban pediatric medical center (Boston Children’s Hospital). Data is not publicly available due to data privacy laws. Code is publicly available at github.com/hcoggan/decompensation-prediction.

Institutional Review Board (IRB) This research was approved by the Institutional Review Board (IRB) at Boston Children’s Hospital (IRB #P00048571) under a waiver of informed consent, as this was a retrospective, observational study on de-identified data.

1. Introduction

Emergency departments [EDs] are fast-paced environments in which critically ill patients must be identified and treated quickly. Patients arriving at an emergency department [ED] are triaged at arrival to determine their need for care. Rapid and accurate triage is especially important in pediatric emergency care because of the speed and severity with which children can deteriorate (Frazier et al., 2011; Hazinski, 2012).

In the US, triage nurses generally use the Emergency Severity Index to assign acuity levels to patients (Wuerz et al., 2000; Wolf et al., 2023). This is a five-level system ranging from ESI 1 (patient requires immediate life-saving intervention) to ESI 5 (patient will likely require no hospital resources and can safely wait to be seen). In many EDs, patients with lower-acuity triage scores are treated in a separate ‘fast track room’ to improve patient flow (Wright et al., 1992). By design, these patients are monitored less closely and generally wait longer to be seen than

patients with high-acuity triage scores. This can lead to life-threatening delays in care when patients with severe conditions are misclassified as ‘non-urgent’ at triage (generally ESI 3-5) (Sax et al., 2025, 2026). It is therefore vital that high-acuity patients are correctly identified during the triage process, and receive triage scores that reflect their risk of deterioration.

Though widely used, ESI-based triage methods have a number of well-documented imperfections. The ESI system is deliberately coarse, designed to assign patients to a small number of categories to simplify resource allocation in a high-pressure environment, and impose large differences on patients (whose real clinical risk varies continuously) at the borders between triage acuity levels. In theory, ESI levels are assigned according to a strict protocol depending on symptoms, vital signs and anticipated resource needs (Wolf et al., 2023). However, the guidelines leave some room for clinical judgment and previous work has found that ESI scores can be influenced by a patient’s race, ethnicity and language (Dennis, 2021; Metzger et al., 2022; Joseph et al., 2022). More generally, ESI scores vary significantly between triage nurses (Alpert et al., 2013; Wolf et al., 2025) and hospital sites: in one comparison of US pediatric EDs, about a quarter of visits were assigned to ESI-5 at one site, but the category was almost never used at another (Milà et al., 2026). These differences can drive both under- and over triage of ED patients, which are common in pediatric ED populations (Geanakopoulou et al., 2025). Accurate machine learning (ML) methods deployed at triage may reduce unwarranted variations and improve identification of high-risk pediatric patients.

In this work, we focus on identifying under-triaged pediatric patients: those who were classified as ‘non-urgent’ at triage (ESI 3-5) but experienced critical deterioration later in their stay (severe hypotension, hypoxia, intubation, ICU transfer or death). A clinical decision support tool to predict deterioration risk at triage could, if deployed in a real ED, prompt the nurse to ‘up-triage’ a patient (i.e. reassign them to ESI 2, requiring clinician evaluation within 15 minutes). However, deterioration amongst non-urgent patients is a rare event and even a very accurate model would likely produce a large number of false alarms for every patient accurately detected. Up-triaging too many patients may have adverse system-wide effects: if *every* patient is classified as ‘urgent’, then in effect none of them are, because clinician attention is finite and the purpose of the triage system

is to preferentially allocate care to sicker patients. In addition to causing delays for urgent patients, high levels of up-triaging would push remaining non-urgent patients further down the queue and lengthen ED stays overall. The challenge then becomes not just how to develop a clinical decision support tool to identify high-risk ‘non-urgent’ patients, but how to deploy it in the real world to avoid overwhelming already-overcrowded EDs (Rabin et al., 2012; Morley et al., 2018).

Motivated by these challenges, we analyze a large dataset of visits to a US pediatric ED and identify critical deterioration within a subset of visits initially classified as non-urgent. We use XGBoost to predict clinical deterioration risk from information available at triage using a 50/50 chronological split of visits. We observe good discrimination (AUROC 0.80 [95% CI 0.78-0.81]) but a high number needed to screen (NNS): for instance, to catch half of all deteriorations requires 26 up-triages per critically ill patient. To assess whether and how this tool could be feasibly deployed, we design a simulation framework to assess the system-wide consequences of up-triaging patients at a given deterioration risk threshold. Observing that ED length of stay [LOS] is log-normally distributed, we fit a log-linear model to quantify the effect on ED LOS of up-triaging a given patient, as well as the effect *on other concurrent visits* of having to compete for hospital resources with an additional up-triaged patient. We then retrospectively simulate the effect of deploying the model in our test cohort at a given risk threshold r , using a discrete-event model to capture interaction effects between patients. Our simulations suggest that, across two years of visits, our CDS tool could have been deployed at a sensitivity of 49% without increasing the prolonged stay rate (LOS > 8h), a key reporting metric (Haimovich et al., 2025), by significantly more than 1%. At this threshold, correctly-flagged patients would be admitted 23 minutes faster on average. We present this pipeline—validating a clinical decision support tool with a novel discrete-event modelling framework to capture second-order effects of its deployment—as a general method to assess the feasibility of ML tools in health, and suggest that similar studies could inform the risk thresholds used in prospective clinical trials.

2. Related Work

Over the last two decades, a variety of ML methods have been developed to predict hospital outcomes,

but have faced significant barriers to clinical deployment (Brink et al., 2022). Common prediction targets include hospital admission from the emergency department (Barak-Corren et al., 2017; Hong et al., 2018; Barak-Corren et al., 2021; La Cava et al., 2023), sepsis (Raita et al., 2019; Liu et al., 2025) and in-hospital deterioration or mortality (Levin et al., 2018; Perng et al., 2019; Fernandes et al., 2020; Rust et al., 2023). Such models, which are generally based on gradient-boosting or deep learning techniques, may fail to accurately capture critical illness by using definitions that exclude interventions performed in the ED itself. These patients may never be admitted to the ICU but are likely to have been under-triaged if they were initially classified as non-urgent and monitored accordingly. Our work focuses on identifying and intervening on such patients.

In this study, our intended tool would prompt a nurse to up-triage a patient as soon as possible after the triage process takes place, and so we limit our inputs to information collected within 30 minutes of the patient’s arrival. As this requires a one-time binary prediction, we use gradient-boosted classification algorithms on structured data, which have previously been shown to perform well in predicting ED outcomes (Barak-Corren et al., 2017; Raita et al., 2019; Barak-Corren et al., 2021; Poursoltan et al., 2025; Strutz et al., 2025).

Perhaps most relevantly to our current work, Ha et al. (2026) used a number of ML models, including gradient-boosted trees and neural networks, to predict deterioration using information available at triage at a large urban pediatric ED (not ours). These authors considered critical care interventions amongst all ED patients (except those triaged at ESI 1, many of whom will have required such interventions at arrival), not simply those classified as non-urgent. Practically speaking, this model, if implemented, would replace the existing ESI system, or at least split existing ESI levels into more fine-grained categories (‘high-risk ESI 2’, ‘low-risk ESI 2’, ‘high-risk ESI 3’, and so on). This approach would face significant practical barriers within EDs which are currently designed around the five-level ESI system. By focusing on ‘non-urgent’ patients, we can create models whose predictions are actionable within the existing ESI system (by up-triaging ‘non-urgent’ patients to ESI 2). Further, these authors assessed their model counterfactually by assuming that timestamps would be swapped between patients, and evaluated their model’s impact *only on patients who received*

critical care interventions. Such evaluations ignore the system-wide effects of real-world deployment, including effects on other patients who may well have been extremely ill and who would face delays in care under the counterfactual systems. Here, we use discrete-event simulations of patient flow (Doudareva and Carter, 2022) to evaluate our model’s effect on all patients.

3. Data

3.1. Preprocessing

This was a retrospective cohort study conducted on visits to the ED of a large urban pediatric medical center between January 2019 and May 2024. Available information included sex, gender (if different from sex), self-reported race, triage acuity score, age, weight, mode of arrival at ED, chief complaint (categorised based on the schema developed by Gorelick et al. (2005)), comorbidities diagnosed in the year preceding the visit, comorbidities diagnosed from the visit (categorised using the Pediatric Comorbidity Index developed by Sun et al. (2021)), vitals measured throughout the stay (heart rate, respiratory rate, blood pressure, blood oxygen), pain scores, Children’s Hospital Early Warning Scores [CHEWS, used to predict deterioration in hospitalised patients and occasionally used in our ED when admission is likely], ED visits and/or admissions in the 30 days preceding the stay, insurance source, primary language, time of arrival (year, season, time of day, whether the visit falls on a weekend), crowding (i.e. the number of other patients present in the ED at the time of arrival), miles travelled to the hospital, state of origin, index of socio-economic deprivation, and whether or not the patient had been referred to the ED by a physician. Socio-economic deprivation was approximated using the Robert Graham Center’s Social Deprivation Index associated with the patient’s zip code (Robert Graham Center - Policy Studies in Family Medicine & Primary Care, 2018). Distance travelled to the hospital was calculated on the assumption that a patient travelled from their home address. When predicting deterioration, all vitals measured after the first thirty minutes of the patient’s stay were excluded, along with visit disposition and all resulting diagnoses, in order to ensure predictions were made with information available around the time of triage (and at latest within 30 minutes of arrival). Visits were excluded if they met any of the following cri-

teria: missing unique visit identifier; missing patient identifier; no recorded sex; disposition recorded as ‘left without being seen’ or similar, or missing disposition with no associated vitals (suggesting departure before triage); or irreconcilable timestamps (e.g. departure time earlier than arrival time).

3.2. Outcome definition

After preprocessing, our dataset contained records of 305,963 visits (the ‘full cohort’). We were interested in predicting critical illness amongst patients who were under 18 at the time of arrival, were initially triaged as ‘non-urgent’ (i.e. ESI score 3, 4, or 5), and had no unstable vital signs within the first 30 minutes of their visit (see below). A total of 211,243 visits met these criteria (the ‘screening cohort’).

We sought to predict deterioration at least 60 minutes into the stay (i.e. at least 30 minutes after prediction time). Deteriorations occurring within 30 minutes of prediction time were not counted, to ensure a meaningful gap between prediction and outcome (i.e. to avoid label leakage, where a model’s predictions unwittingly reflect real-time clinical suspicion instead of actually predicting future events). For the same reason, visits shorter than 60 minutes (3% of the screening cohort) were not excluded in our screening cohort, to avoid conditioning on information (departure time) not known at the time of the prediction. Deterioration was identified if any of the following outcomes were observed in the patient’s record: new unstable vital signs, intubation, or administration of vasoactives during the ED stay; ICU admission; or in-hospital death. Unstable vital signs included severe hypotension, defined as a systolic blood pressure at least 5 mmHg below the age-dependent PALS criterion for pediatric hypotension (Haque and Zaritsky, 2007), or severe hypoxia (SpO2 below 85%). Intubation was inferred from the presence of intravenous rocuronium or etomidate in the patient’s medication record. Vasoactives indicative of critical deterioration included intravenous epinephrine (a proxy for CPR), norepinephrine, dopamine, vasopressin, and milrinone.

3.3. Cohort description

After filtering, the screening cohort comprised 211,243 visits from 126,184 unique patients. Critical deterioration occurred during 2,012 of these visits (1.0%). 52.8% of visits were from male patients. Visits were 30.7% from Non-Hispanic White patients,

28.5% Hispanic, 15.1% Non-Hispanic Black, 11.0% from patients of unknown race, 10.6% Other race, 4.0% Asian and 0.2% Hispanic White. The median patient age was 5.3 years (IQR 2.0-11.0). 61.1% of visits were assigned ESI 3 at triage, with 35.3% allocated to ESI 4 and the remainder (3.6%) ESI 5. The most common chief complaints (non-exclusive) were fever (15.3%), extremity-related complaints (7.6%), abdominal pain (7.1%), vomiting (5.8%) and cough (5.7%). Over half of visits (54.1%) had public insurance and most (77.8%) spoke English as their preferred language; common alternatives included Spanish (12.9%), Portuguese (2.2%) and Arabic (1.3%).

4. Predicting critical deterioration

Visits were split 50/50 chronologically into training and testing cohorts (roughly January 2019-September 2021 and Oct 2021-May 2024). We tested three different gradient-boosting classifiers to predict deterioration: XGBoost (Chen and Guestrin, 2016), CatBoost (Dorogush et al., 2018) and LightGBM (Ke et al., 2017). We also included a small neural network (a multilayer perceptron), and a simple baseline logistic regression model trained on ESI score alone. All classifiers were tuned using 5-fold cross-validation, using 100 randomly chosen hyperparameter sets. Mean imputation was performed to handle missing values for the neural network (age-group-specific for heart rate, respiratory rate and weight), with all means fit on the training data to avoid leakage. The best-performing XGBoost model (AUROC 0.80) was found to outperform the ESI-only baseline (0.65), CatBoost (0.69), LightGBM (0.74) and the neural network (0.79), and was chosen as the basis of our later simulations. This model showed good discrimination (see Table 1), with an AUROC of 0.80 (95% CI 0.78-0.81) and an AUPRC of 0.06 (95% CI 0.05-0.07). Given the baseline prevalence of 1.0%, this suggests our model is roughly six times better at balancing precision and recall than random chance (AP Lift: 6.4 [5.3-7.5]). The model was well-calibrated, with an expected calibration error (ECE) of 0.22% when visits were grouped by predicted risk decile and a calibration slope of 0.92 (see Figure 5). Model performance did not differ by sex, race/ethnicity, preferred language, or year of arrival, though AUROC was slightly higher for patients with public insurance (see Figure 6).

SHapley Additive exPlanations (SHAP values, Lundberg and Lee (2017)) suggested that the model

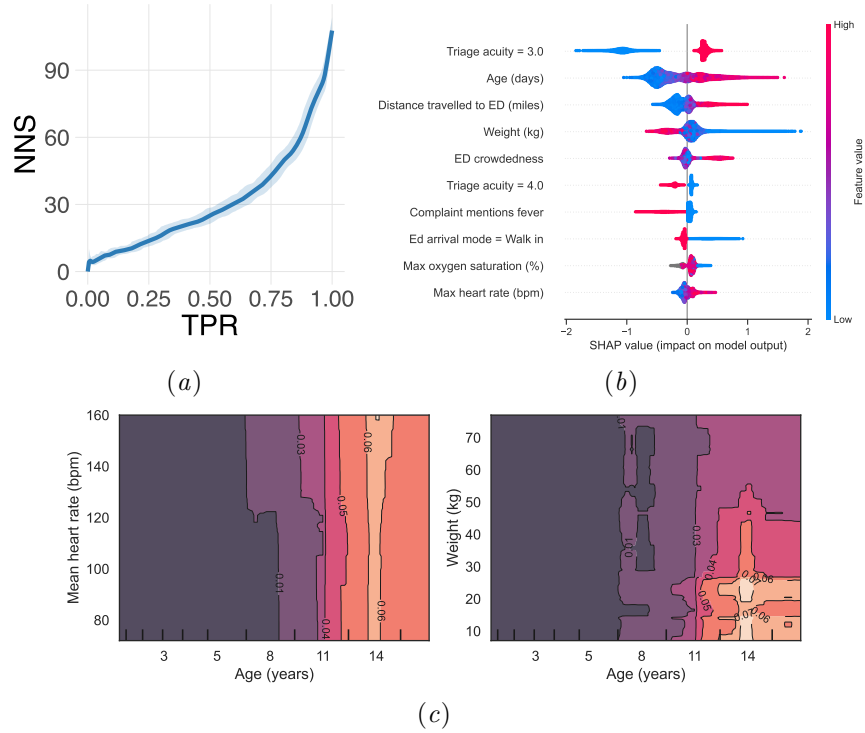


Figure 1: Post-hoc explanation of the deterioration prediction model. (a) Number needed to screen (NNS, y-axis) as a function of sensitivity (x-axis), with bootstrapped 95% confidence intervals. (b) SHAP values summarising the direction and relative magnitude of each feature’s contribution to the predicted risk. (c) Partial dependence of predicted deterioration risk on mean heart rate/ weight and age. Contour lines describe the mean predicted risk, computed over 1000 randomly-selected visits with complete heart rate and weight information.

Table 1: Performance on the held-out (chronologically later) test set. AP Lift: AUPRC divided by outcome prevalence. 95% CIs (1000 bootstrap resamples) in parentheses; best model in bold.

Model	AUROC	AUPRC	AP Lift
ESI LR	.645 (.634–.654)	.013 (.013–.014)	1.43 (1.42–1.49)
CatBoost	.692 (.668–.703)	.024 (.021–.028)	2.63 (2.23–2.93)
LightGBM	.736 (.716–.747)	.035 (.028–.042)	3.74 (3.08–4.46)
MLP	.788 (.772–.801)	.044 (.036–.050)	4.68 (4.00–5.31)
XGBoost	.800 (.783–.813)	.060 (.050–.071)	6.42 (5.33–7.51)

was likelier to predict deterioration for patients with ESI level 3, for older children, and for those who lived far away from the ED (Figure 1(b)). SHAP analy-

sis suggested that complaints including ‘allergic reaction’, ‘chronic disease’, ‘other respiratory’, ‘seizure’ and ‘cough’ were predictive of deterioration, along with previous diagnoses of congenital malformations, epilepsy, cardiovascular conditions or developmental delays.

We did not normalise vital signs by age group (the approach taken by Barak-Corren et al. (2017), amongst others), as this tended to decrease model performance in early tests. However, partial dependence plots of the interactions between vital signs and age suggested that the model correctly inferred known clinical relationships between age and measured vital signs, worrying more about a low weight and elevated heart rate in older children (see Figure 1(c)). This suggests that predictions are improved when XGBoost is allowed to learn age-vital relation-

ships on its own, and meaningful information is lost when vitals are normalised around arbitrary age cut-offs.

Given the low baseline prevalence of critical deterioration, real-world deployment of the model would result in a large number of false alarms (see Figure 1(a)). At a sensitivity of 25% we would need to up-triage 13.6 people to catch one critical deterioration; at 52%, we would need 26.5; and at 75% we would need 44.7. To better understand whether the model would be workable in practice, and at what sensitivity, we developed a simulation framework to assess the effects of these false alarms on system-level crowding.

5. Quantifying the effect of crowding on length of stay

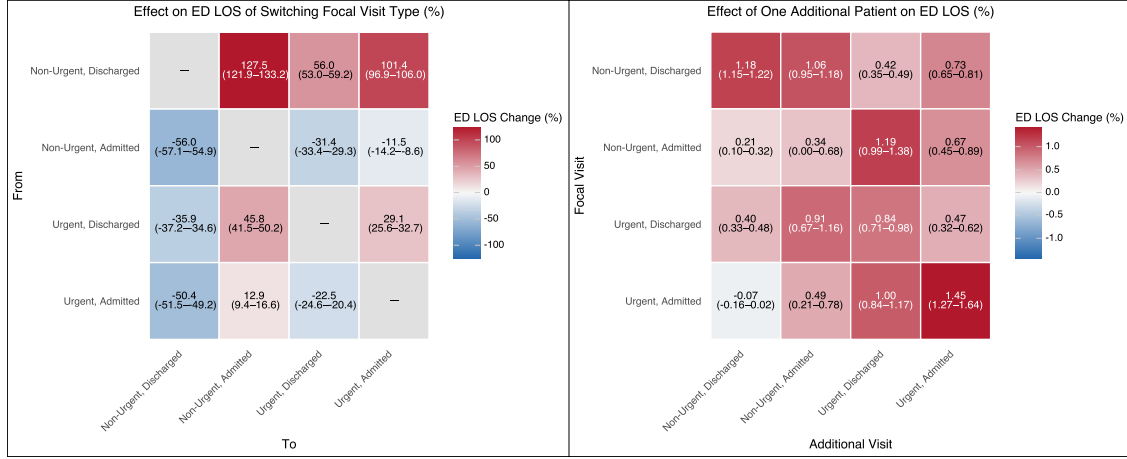
Predicting individual ED length of stay (LOS) is a notoriously difficult problem (Farimani et al., 2024; Vallée, 2026). The duration of individual stays are affected by dynamic hospital-level factors including clinician availability, bed occupancy in specific units, the length of time required for particular lab tests, and so on. As we are interested in simulating the aggregate system-wide effects of deploying our deterioration risk model, we based our simulation on general, cohort-wide estimates of the effect of crowding levels and triage urgency. Unlike a simulation approach based on counterfactual ML-based predictions of ED LOS, this high-level approach allows us to quantify and report our simulation parameters.

When a patient is up-triaged from ‘non-urgent’ to ‘urgent’, this will directly affect their own length of stay and, later, indirectly affect the crowding levels experienced by other patients. We first aimed to quantify the effect of visit type (split by disposition into ‘urgent, admitted’, ‘urgent, discharged’, ‘non-urgent, admitted’, and ‘non-urgent, discharged’) and crowding levels on length of stay. Observing that ED LOS is approximately log-normally distributed, we fit a log-linear model to estimate the effects of visit type and crowding. Here we defined LOS as the duration between a patient’s arrival time and their ED departure. The crowding experienced by a focal visit was defined as the number of patients in the ED at the time of arrival and was further split by visit type. We used a log-linear model to estimate the effect of focal visit-type on LOS, as well as its interaction with visit-type-specific crowding (e.g. the effect on urgent-admitted visits of crowding from

non-urgent-discharged visits, and so on). These effects were further adjusted for temporal factors, demographic factors (race, sex, age group, insurance source, language, miles travelled to the hospital), and medical factors (presence of physician referral, ED arrival mode, number of previous recent admissions/ED visits, chief complaint, previous and visit diagnoses). We emphasise that this LOS model is descriptive, not predictive; it was fit to all visits in the test and training cohort, controlled for variables not known at triage (including disposition and final visit diagnosis), and its individual visit-level ‘predictions’ were never used. The function of the LOS model was to identify the *general* effects of visit-type and crowding on a patient’s LOS (see Figures 2(a) and 2(b)). These effects were then used as inputs to the counterfactual simulations (see later).

This approach allowed us to estimate the theoretical ‘switching effect’ on ED LOS of switching patients from visit type s to s' ($a'_s - a_s$, see Figure 2(a)) and the ‘crowding effect’ on a visit of type s of an additional visit of type s' ($b_{s,s'}$, see Figure 2(b)). On average, we would expect that up-triaging a patient from non-urgent to urgent would reduce their stay by 11.5% (95% CI 8.6-14.2%) if they were eventually admitted, and increase it by 56.0% (53.0-59.2%) if they were discharged (as urgent patients receive more thorough investigation before discharge). Crowding effects, unsurprisingly, were much smaller, but still significant between almost all visit types. The largest interaction effects were observed between urgent, admitted patients, each of whom would be expected to increase the other’s length of stay by 1.45% (1.27-1.64%). Each additional urgent-discharged patient would be expected to increase an urgent-admitted patient’s stay by 1.00% (0.84-1.17%), but non-urgent discharged patients had no significant effect on urgent-admitted patients. These results validated our intuition that up-triaging a large number of visits would increase competition amongst urgent patients for finite hospital resources, and delay admission for the sickest patients.

Because the LOS model controlled for the patient’s true clinical severity as far as possible, this allowed us to capture the effect on LOS of *urgent classification*, over and above how sick the patient actually was. We could therefore apply these effects counterfactually to non-urgent patients— even ‘false positives’ who did not deteriorate— to estimate what their LOS would have been, had they been uptriaged. Motivated by this, we designed a novel discrete-event



(a) Effect of switching visit type.

(b) Effect of one additional concurrent visit ('crowding effects').

Figure 2: Effect of visit type and crowding on emergency department (ED) length of stay (LOS). (a) Estimated relative change in LOS if a patient were reclassified from one visit type to another. In simulations only 'non-urgent, admitted' to 'urgent, admitted' and 'non-urgent, discharged' to 'urgent, discharged' would actually take place; other hypothetical 'switches' are included for comparison. (b) Estimated effect on ED LOS of a focal visit (rows) from each additional concurrent patient of a given visit type present at arrival (columns). Values in parentheses show 95% confidence intervals.

model to quantify the effect of our deterioration-risk model on patient flow.

6. Simulating deployment of the deterioration-risk model

The workings of our discrete-event model are described in Figure 3. Beginning at the start of our test-set period, we iterated over all visits sequentially and simulated the effect of deploying our deterioration-risk model with an operational action threshold r_0 . Each non-urgent visit was reclassified as urgent if its predicted deterioration risk $r_i \geq r_0$ and left as non-urgent otherwise. We assumed that disposition (admitted/discharged) was unchanged, but as deterioration risk was modelled as a function of crowding, we recalculated risk predictions using simulated crowding levels. (Before feeding it to the model, we capped simulated crowding levels at 250, about twice the maximum observed in the real world, to avoid pulling the model too far away from its training distribution and generating unreliable outputs.) We then generated a simulated ED LOS by multiplying the

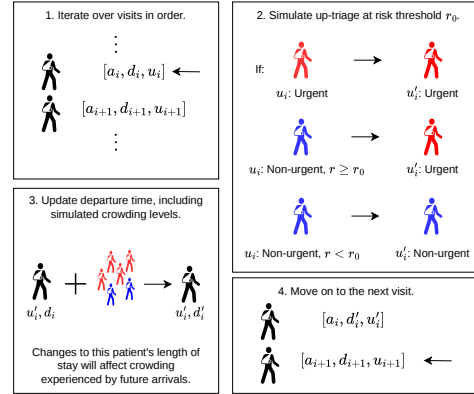


Figure 3: Schematic of the discrete-event modelling framework used to simulate model deployment at risk threshold r_0 . The i th visit has real-world arrival time a_i , departure time d_i , and urgency u_i , and predicted deterioration risk r_i .

real-world ED LOS by a stochastic factor drawn from the coefficients calculated in Figure 2.

Within the counterfactual simulations, we assumed that all demographic (sex, race, age), clinical (chief complaint, mode of arrival) and temporal (season, year) factors were unchanged, along with visit-level random variation in LOS. We described each visit using a 19-element vector s . The first three elements were binary and described the focal visit-type a_s ($[0, 0, 0]$ described a non-urgent, admitted visit). The other 16 elements described the interaction between possible focal visit type and visit-type specific crowding (4×4 terms). Any change in a patient’s visit type or in the number or type of visits ongoing when they arrived would result in a change in s (i.e. s_{sim} , the counterfactual state, different from s_{real} , the real-world state).

Regression of LOS onto this vector s gave a vector v with known mean and covariance, describing the effect of focal visit type and crowding on a patient’s log-length of stay. For each of 200 random seeds, we drew a vector v' from the distribution $N(\mathbb{E}[v], \text{Cov } v)$ to account for simulation-level noise. For each visit within that simulation, then, the counterfactual length of stay LOS_{sim} was given by

$$LOS_{sim} = LOS_{real} \times \exp[v' \cdot (s_{sim} - s_{real})]$$

where LOS_{real} was the real-world length of stay. The counterfactual length of stay depended on whether the visit had been ‘switched’ from its real-world visit type (only applicable to up-triaged patients) and on whether crowding had shifted from its real-world distribution (applicable to all visits, including originally-urgent visits, because of the downstream effects of previous up-triages). The simulation iterated over each visit, up-triaged according to dynamic prediction risk at a given risk threshold, and adjusted LOS where necessary. By design, at a sensitivity of 0 (i.e. with no uptriages), the simulation replicated the real-world LOS and crowding patterns exactly. We measured the change in ED LOS over the test set period using 95% bootstrapped confidence intervals. Visits were clustered into 24-hour blocks before bootstrapping to account for interdependence in LOS induced by crowding. Each random seed therefore determined both the simulation-level noise in switching and crowding effects and the bootstrap draw.

We simulated deployment across a range of decision thresholds (see Figure 4). We report metrics at two thresholds, chosen for their impact on the ‘prolonged stay’ reporting metric (i.e. the fraction of

ED visits exceeding 8 hours). In the conservative-use regime, we might limit up-triages to ensure *no significant difference* in prolonged stay rate. This would allow a maximum mean operating sensitivity of 1.5% (0.9-2.3%) and mean specificity of 99.9% (99.9-100%). This would have no effect on ED LOS overall or for urgent visits (see Table 2), but would reduce time to admission for true positives (i.e. correctly up-triaged patients) by an average of 55 minutes [95% CI 34-85 mins]. Alternatively, in a ‘liberal-use’ regime, we might require that up-triages increase prolonged stay rate by an amount *not significantly greater than 1%*. This would allow a mean sensitivity of 49% (46-51%) and mean specificity of 89% (89-89%). The average visit would be approximately 11 minutes longer [8.1-13 mins] (see Table 4), but correctly-flagged patients would be admitted 23 minutes faster [10-34 mins]. Induced delays would be similar between male and female patients and between the publicly- and privately-insured, but longer for Non-Hispanic White and English-speaking patients than for Hispanic and Spanish-speaking patients respectively (see Table 5).

7. Conclusions

In this study we present a model designed to detect under-triaged patients in the pediatric ED. If deployed, this model could be used to reclassify flagged patients from ‘non-urgent’ to ‘urgent’, allowing closer monitoring and possibly preventing the critical deterioration events the model predicts. To avoid overwhelming the ED with too many urgent patients and allow efficient prioritization of hospital resources, we quantified the direct and indirect effects on ED LOS of up-triaging a large number of patients, and show that the model could be used to detect half of all critical deterioration events and decrease admission times for flagged patients without impacting the prolonged-stay rate ($LOS > 8h$, a key reporting metric) by more than 1%.

Our results suggested that the model would slightly speed up processing of discharged patients who were not up-triaged, as ‘fast track’ rooms emptied out and more patients were classified to urgent pathways. However, *admitted* patients who were still classified as non-urgent could expect increased delays at most risk thresholds, as they would compete for attention with a larger pool of urgent patients.

Our study is single-site and has a number of limitations. For simplicity, we excluded patients who left without being seen or against medical advice, though

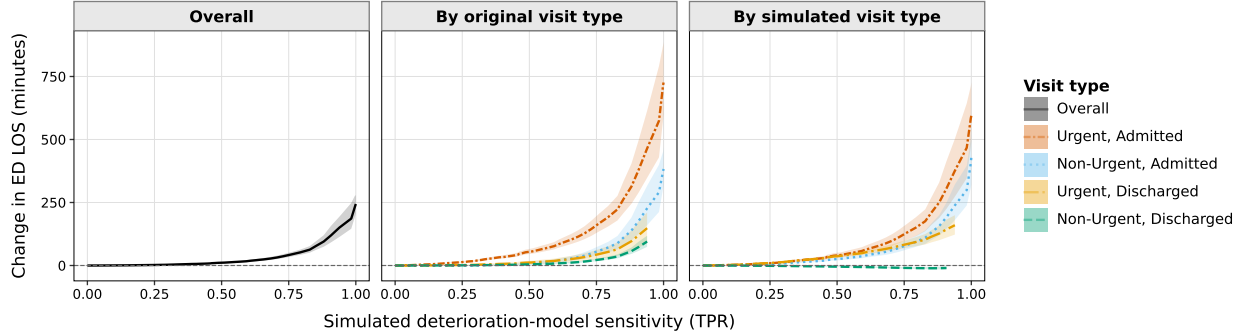


Figure 4: Simulated mean change in ED LOS (y-axis) when deploying the model at a given sensitivity (x-axis). Means are computed across 200 simulations with different random seeds. Left: mean change across all visits; middle: mean change split by real-world visit type; right: mean change split by simulated visit type (i.e. incorporating counterfactual uptrriages). Shaded areas indicate 95% confidence intervals.

this phenomenon may also be affected by crowding. We note that this exclusion likely leads us to overestimate the effect of crowding on visit length of stay (by ‘under-counting’ patients during periods of high crowding, when patients are more likely to leave without being seen), which in turn increases our estimates of the adverse impact of our model. This will drive more conservative deployment, which may be clinically desirable.

Our discrete-event model assumed that the general effects of visit type and crowding extracted from a log-linear model could be applied to individual visits, which may not be appropriate to all contexts; hospitalists may be particularly interested in the model’s downstream effects on, for example, cancer patients or those experiencing seizures. We also assume that crowding effects extracted from observational data are causal when we use them as inputs to our simulation. While we cannot prove causation, we have strong *a priori* reasons to believe that crowding does increase ED length of stay: unless ED crowding correlates with unobserved sources of variation in hospital resources (i.e. staffing levels), which is unlikely, a large influx of patients in the ED will force clinicians to divide their time between more patients and delay necessary care. By controlling for chief complaint, visit type, seasonal and temporal factors in our LOS model, we aimed to isolate this causal component, though we admit that we may be biased by unobserved confounders.

In future work, we hope to further control for hospital-wide occupancy rates, NEDOCS scores, or staffing levels, none of which were available to us during this study. Other models may also include more complex nonlinear descriptions of the effect of crowding on LOS (e.g. saturating functions); but we assumed a linear relationship here for simplicity. Nonetheless our work presents a novel framework for evaluating the feasibility of clinical decision support tools. In addition to the model itself, our simulated retrospective clinical trial approach could be useful in a variety of contexts beyond the emergency department.

Acknowledgments

Research reported in this publication was supported by the National Library of Medicine of the National Institutes of Health (NIH) Award R01LM014300. The content is the responsibility of the authors and does not necessarily represent the official views of NIH. The authors acknowledge the use of AI (Claude Code) to assist with code generation.

References

Evan A. Alpert, Ari M. Lipsky, Dvora Hertz, Jonathon Rieck, and Jacob Or. Simulated evaluation of two triage scales in an emergency depart-

- ment in Israel. *European Journal of Emergency Medicine*, 20(6):431, December 2013. ISSN 0969-9546. doi: 10.1097/MEJ.0b013e32835e2b42.
- Yuval Barak-Corren, Andrew M. Fine, and Ben Y. Reis. Early prediction model of patient hospitalization from the pediatric emergency department. *Pediatrics*, 139, 5 2017. ISSN 10984275. doi: 10.1542/peds.2016-2785.
- Yuval Barak-Corren, Pradip Chaudhari, Jessica Perniciaro, Mark Waltzman, Andrew M. Fine, and Ben Y. Reis. Prediction across healthcare settings: a case study in predicting emergency department disposition. *npj Digital Medicine*, 4, 12 2021. ISSN 23986352. doi: 10.1038/s41746-021-00537-x.
- Anniek Brink, Jelmer Alsma, Lodewijk Aam van Attekum, Wichor M. Bramer, Robert Zietse, Hester Lingsma, and Stephanie Ce Schuit. Predicting in-hospital admission at the emergency department: a systematic review. *Emergency medicine journal: EMJ*, 39(3):191–198, March 2022. ISSN 1472-0213. doi: 10.1136/emered-2020-210902.
- Tianqi Chen and Carlos Guestrin. Xgboost. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 785–794. ACM, 8 2016. ISBN 9781450342322. doi: 10.1145/2939672.2939785.
- Jeff A Dennis. Racial/ethnic disparities in triage scores among pediatric emergency department fever patients. *Pediatric Emergency Care* •, 37, 2021. ISSN 0749-5161. URL www.pec-online.com.
- Anna Veronika Dorogush, Vasily Ershov, and Andrey Gulin. CatBoost: gradient boosting with categorical features support, 2018. URL <https://arxiv.org/abs/1810.11363>. Version Number: 1.
- Evgueniia Doudareva and Michael Carter. Discrete event simulation for emergency department modelling: A systematic review of validation methods. *Operations Research for Health Care*, 33: 100340, June 2022. ISSN 22116923. doi: 10.1016/j.orhc.2022.100340. URL <https://linkinghub.elsevier.com/retrieve/pii/S2211692322000029>.
- Raheleh Mahboub Farimani, Hesam Karim, Alireza Atashi, Fariba Tohidinezhad, Kambiz Bahaadini, Ameen Abu-Hanna, and Saeid Eslami. Models to predict length of stay in the emergency department: a systematic literature review and appraisal. *BMC Emergency Medicine*, 24(1):54, April 2024. ISSN 1471-227X. doi: 10.1186/s12873-024-00965-4. URL <https://bmccemergmed.biomedcentral.com/articles/10.1186/s12873-024-00965-4>.
- Marta Fernandes, Rúben Mendes, Susana M. Vieira, Francisca Leite, Carlos Palos, Alistair Johnson, Stan Finkelstein, Steven Horng, and Leo Anthony Celi. Risk of mortality and cardiopulmonary arrest in critical patients presenting to the emergency department using machine learning and natural language processing. *PLOS ONE*, 15(4):e0230876, April 2020. ISSN 1932-6203. doi: 10.1371/journal.pone.0230876. URL <https://dx.plos.org/10.1371/journal.pone.0230876>.
- Aisha Frazier, Elizabeth A. Hunt, and Kathryn Holmes. Pediatric cardiac emergencies: Children are not small adults. *Journal of Emergencies, Trauma, and Shock*, 4(1):89–96, January 2011. ISSN 0974-519X. doi: 10.4103/0974-2700.76842.
- Alexandra T. Geanacopoulos, Alon Peltz, Katherine Melton, Mark I. Neuman, Colleen K. Gutman, Kathleen E. Walsh, Margaret E. Samuels-Kalow, and Kenneth A. Michelson. Pediatric Triage Accuracy in Pediatric and General Emergency Departments. *Hospital Pediatrics*, 15(1):37–45, January 2025. ISSN 2154-1663, 2154-1671. doi: 10.1542/hpeds.2024-008063. URL <https://publications.aap.org/hospitalpediatrics/article/15/1/37/200006/Pediatric-Triage-Accuracy-in-Pediatric-and-General>.
- Marc H. Gorelick, Elizabeth R. Alpern, and Evaline A. Alessandrini. A system for grouping presenting complaints: The pediatric emergency reason for visit clusters. *Academic Emergency Medicine*, 12:723–731, 8 2005. ISSN 1069-6563. doi: 10.1197/j.aem.2005.03.530.
- Trang Ha, Brandon Kappy, James M. Chamberlain, and Kenneth W. McKinley. Early Prediction of Critical Care Interventions From Pediatric Emergency Department Triage. *Hospital Pediatrics*, 16(8):e617–e626, August 2026. ISSN 2154-1663, 2154-1671. doi: 10.1542/hpeds.2025-009127. URL <https://publications.aap.org/hospitalpediatrics/article/16/8/e617/208272/Early-Prediction-of-Critical-Care-Interventions>.
- Adrian D. Haimovich, Sarah D. Berry, and Bruce E. Landon. Prolonged Emergency Department

- Stays for Older US Adults. *JAMA Internal Medicine*, 185(8):1037, August 2025. ISSN 2168-6106. doi: 10.1001/jamainternmed.2025.2006. URL <https://jamanetwork.com/journals/jamainternalmedicine/fullarticle/2835685>.
- Ikram U. Haque and Arno L. Zaritsky. Analysis of the evidence for the lower limit of systolic and mean arterial pressure in children. *Pediatric Critical Care Medicine*, 8(2):138–144, March 2007. ISSN 1529-7535. doi: 10.1097/01.PCC.0000257039.32593.DC. URL <https://www.ovid.com/00130478-200703000-00007>.
- M. S. N. Hazinski. *Nursing Care of the Critically Ill Child*. Mosby, Chantilly, 3rd ed edition, 2012. ISBN 978-0-323-02040-4 978-0-323-08603-5.
- Woo Suk Hong, Adrian Daniel Haimovich, and R. Andrew Taylor. Predicting hospital admission at emergency department triage using machine learning. *PLOS ONE*, 13(7):e0201016, July 2018. ISSN 1932-6203. doi: 10.1371/journal.pone.0201016. URL <https://dx.plos.org/10.1371/journal.pone.0201016>.
- Joshua W. Joseph, Alden M. Landry, Maura Kennedy, Da’Marcus Eugene Baymon, Alice K. Bukhman, Noémie Elhadad, and León D. Sanchez. Association of race and ethnicity with triage emergency severity index scores and total visit work relative value units for emergency department patients. *JAMA Network Open*, 5(9):e2231769–e2231769, 09 2022. ISSN 2574-3805. doi: 10.1001/jamanetworkopen.2022.31769. URL <https://doi.org/10.1001/jamanetworkopen.2022.31769>.
- Guolin Ke, Qi Meng, Thomas Finley, Taifeng Wang, Wei Chen, Weidong Ma, Qiwei Ye, and Tie-Yan Liu. LightGBM: A Highly Efficient Gradient Boosting Decision Tree. In I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017. URL https://proceedings.neurips.cc/paper_files/paper/2017/file/6449f44a102fde848669bdd9eb6b76fa-Paper.pdf.
- William G La Cava, Elle Lett, and Guangya Wan. Fair admission risk prediction with proportional multicalibration. Technical report, 2023. URL <https://proceedings.mlr.press/>.
- Scott Levin, Matthew Toerper, Eric Hamrock, Jeremiah S. Hinson, Sean Barnes, Heather Gardner, Andrea Dugas, Bob Linton, Tom Kirsch, and Gabor Kelen. 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.e2, May 2018. ISSN 01960644. doi: 10.1016/j.annemergmed.2017.08.005. URL <https://linkinghub.elsevier.com/retrieve/pii/S0196064417314427>.
- Zheng Liu, Wenqi Shu, Teng Li, Xuan Zhang, and Wei Chong. Interpretable machine learning for predicting sepsis risk in emergency triage patients. *Scientific Reports*, 15(1):887, January 2025. ISSN 2045-2322. doi: 10.1038/s41598-025-85121-z. URL <https://www.nature.com/articles/s41598-025-85121-z>.
- Scott Lundberg and Su-In Lee. A Unified Approach to Interpreting Model Predictions, November 2017. URL <http://arxiv.org/abs/1705.07874>. arXiv:1705.07874 [cs.AI].
- Peter Metzger, Leyla Allum, Erin Sullivan, Frankline Onchiri, and Maya Jones. Racial and language disparities in pediatric emergency department triage. *Pediatric Emergency Care* •, 38, 2022. ISSN 0749-5161. URL www.pec-online.com.
- Blanca Romero Milà, Helena Coggan, Andrew M. Fine, Yuval Barak-Corren, Ben Y. Reis, Jaya Aysola, Pradip P. Chaudhari, and William G. La Cava. The Benefit of the Doubt Phenomenon in Emergency Triage Assignment Disparities, February 2026.
- Claire Morley, Maria Unwin, Gregory M. Peterson, Jim Stankovich, and Leigh Kinsman. Emergency department crowding: A systematic review of causes, consequences and solutions. *PLOS ONE*, 13(8):e0203316, August 2018. ISSN 1932-6203. doi: 10.1371/journal.pone.0203316. URL <https://dx.plos.org/10.1371/journal.pone.0203316>.
- Jau-Woei Perng, I.-Hsi Kao, Chia-Te Kung, Shih-Chiang Hung, Yi-Horng Lai, and Chih-Min Su. Mortality Prediction of Septic Patients in the Emergency Department Based on Machine Learning. *Journal of Clinical Medicine*, 8(11):1906, November 2019. ISSN 2077-0383. doi: 10.3390/jcm8111906.

- Lily Poursoltan, Jie Cao, Brian Clay, Beth Trimble, Leah Adrid, Jeffrey Pan, Andrew Chua, John Bell, Christopher A. Longhurst, Kevin Zhu, and Karandeep Singh. Prospective comparison of econometric, machine learning, and foundation models for forecasting emergency department boarding patients. *npj Health Systems*, 2(1):49, December 2025. ISSN 3005-1959. doi: 10.1038/s44401-025-00054-z. URL <https://www.nature.com/articles/s44401-025-00054-z>.
- Elaine Rabin, Keith Kocher, Mark McClelland, Jesse Pines, Ula Hwang, Niels Rathlev, Brent Asplin, N. Seth Trueger, and Ellen Weber. Solutions To Emergency Department ‘Boarding’ And Crowding Are Underused And May Need To Be Legislated. *Health Affairs*, 31(8):1757–1766, August 2012. ISSN 0278-2715. doi: 10.1377/hlthaff.2011.0786.
- Yoshihiko Raita, Tadahiro Goto, Mohammad Kamal Faridi, David F. M. Brown, Carlos A. Camargo, and Kohei Hasegawa. Emergency department triage prediction of clinical outcomes using machine learning models. *Critical Care*, 23(1):64, February 2019. ISSN 1466-609X. doi: 10.1186/s13054-019-2351-7.
- Robert Graham Center - Policy Studies in Family Medicine & Primary Care. Social deprivation index (SDI), November 2018. URL <https://www.graham-center.org/maps-data-tools/social-deprivation-index.html>.
- Laura O. H. Rust, Tyler J. Gorham, Sven Bambach, Ryan S. Bode, Tensing Maa, Jeffrey M. Hoffman, and Steven W. Rust. The Deterioration Risk Index: Developing and Piloting a Machine Learning Algorithm to Reduce Pediatric Inpatient Deterioration*. *Pediatric Critical Care Medicine*, 24(4):322–333, April 2023. ISSN 1529-7535. doi: 10.1097/PCC.0000000000003186. URL <https://journals.lww.com/10.1097/PCC.0000000000003186>.
- Dana R. Sax, E. Margaret Warton, Dustin G. Mark, and Mary E. Reed. Emergency Department Triage Accuracy and Delays in Care for High-Risk Conditions. *JAMA Network Open*, 8(5):e258498, May 2025. ISSN 2574-3805. doi: 10.1001/jamanetworkopen.2025.8498. URL <https://jamanetwork.com/journals/jamanetworkopen/fullarticle/2833434>.
- Dana R. Sax, E. Margaret Warton, Dustin G. Mark, Tina J. Vitale-McDowell, Daniel D. DiLena, Adina S. Rauchwerger, and Mary E. Reed. Association Between Emergency Department Undertriage or Overtriage With Timeliness of Care and Patient Outcomes. *Annals of Emergency Medicine*, 87(6):723–734, June 2026. ISSN 01960644. doi: 10.1016/j.annemergmed.2025.11.018. URL <https://linkinghub.elsevier.com/retrieve/pii/S0196064425013861>.
- Sierra Strutz, Huan Liang, Kyle Carey, Fereshteh Bashiri, Priti Jani, Emily Gilbert, Julie L. Fitzgerald, Nicholas Kuehnel, Maya Dewan, L. Nelson Sanchez-Pinto, Dana Edelson, Majid Afshar, Matthew Churpek, and Anoop Mayampurath. Machine Learning for Predicting Critical Events Among Hospitalized Children. *JAMA network open*, 8(5):e2513149, May 2025. ISSN 2574-3805. doi: 10.1001/jamanetworkopen.2025.13149.
- Jenny W Sun, Florence T Bourgeois, Sebastien Haneuse, Sonia Hernández-Díaz, Joan E Landon, Brian T Bateman, and Krista F Huybrechts. Development and validation of a pediatric comorbidity index. *American Journal of Epidemiology*, 190:918–927, 5 2021. ISSN 0002-9262. doi: 10.1093/aje/kwaa244.
- Alexandre Vallée. Patient-level, Temporal, and Dynamic Operational Indicators of Emergency Department Prolonged Length of Stay: A Retrospective Cohort Study. *Archives of Academic Emergency Medicine*, 14(1):e25, 2026. ISSN 2645-4904. doi: 10.22037/aaem.v14i1.3048.
- Lisa Wolf, Katrina Ceci, Danielle McCallum, and Deena Brecher. *Emergency Severity Index Handbook Fifth Edition*. Emergency Nurses Association, 2023.
- Lisa Wolf, Altair Delao, Francine M. Jodelka, and Claire Simon. Determining emergency severity index acuity: Key triage elements identified by emergency nurses. *Journal of Emergency Nursing*, 51(3):472–479, 2025. ISSN 0099-1767. doi: <https://doi.org/10.1016/j.jen.2024.11.003>. URL <https://www.sciencedirect.com/science/article/pii/S0099176724003337>.
- Seth W. Wright, Terri L. Erwin, Donald M. Blanton, and Charlotte M. Covington. Fast track in the emergency department: A one-year experience

with nurse practitioners. *The Journal of Emergency Medicine*, 10(3):367–373, May 1992. ISSN 07364679. doi: 10.1016/0736-4679(92)90345-T. URL <https://linkinghub.elsevier.com/retrieve/pii/073646799290345T>.

Richard C. Wuerz, Leslie W. Milne, David R. Eitel, Debbie Travers, and Nicki Gilboy. Reliability and validity of a new five-level triage instrument. *Academic Emergency Medicine*, 7(3):236–242, 2000. doi: <https://doi.org/10.1111/j.1553-2712.2000.tb01066.x>. URL <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1553-2712.2000.tb01066.x>.

Appendix A. Appendix

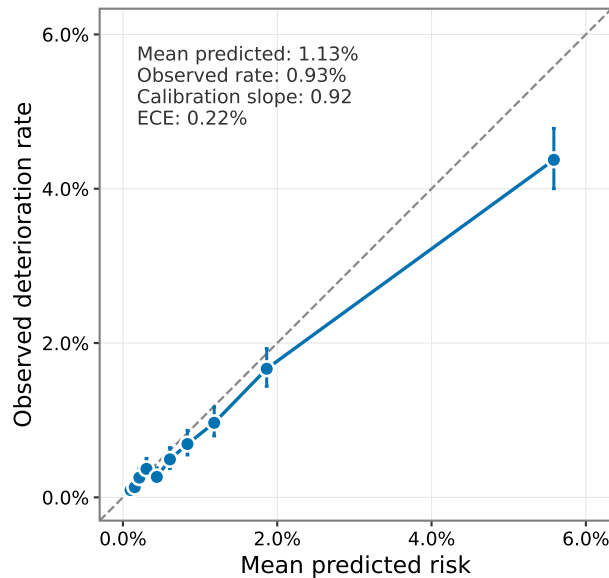


Figure 5: Calibration of the deterioration model. Predictions are grouped into deciles of predicted risk (equal numbers of visits per bin); error bars show Wilson 95% confidence intervals and the dashed line indicates perfect calibration. Expected calibration error (ECE) is computed over the same deciles. The calibration slope is from a logistic regression of the observed outcome on the logit of predicted risk.

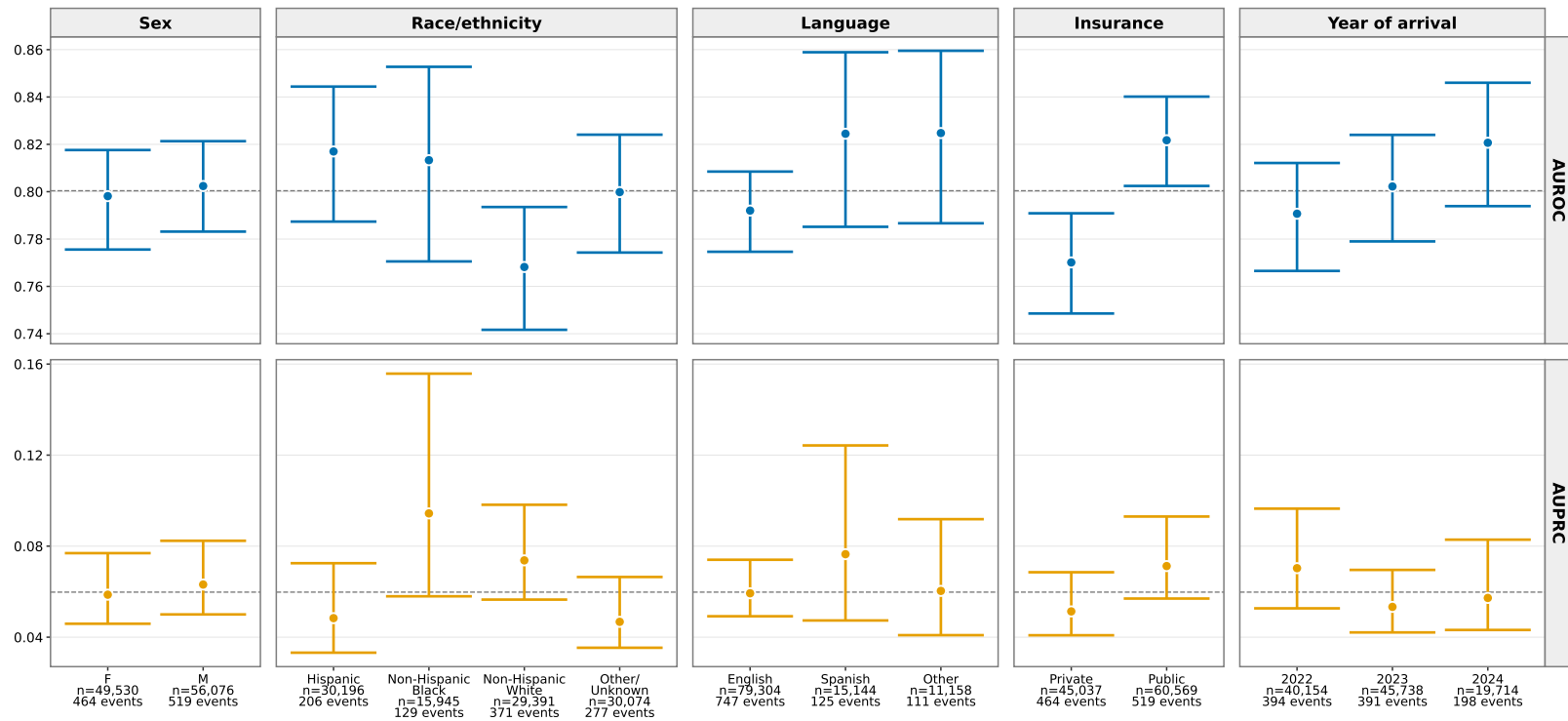


Figure 6: Subgroup-specific AUROC (top) and AUPRC (bottom) of the deterioration model by sex, race/ethnicity, language, insurance, and year of arrival. Error bars indicate 95% confidence intervals; horizontal dotted lines indicate performance across the full test cohort.

Table 2: Change in ED length of stay (LOS) by visit type at 1.5% model sensitivity, the lowest threshold with no significant increase in the prolonged stay rate (LOS > 8 h). Values are means with 95% bootstrap confidence intervals in brackets; negative values indicate shorter stays.

	Change in ED LOS (min)
Overall	-0.07 [-0.11, -0.03]
<i>Original visit type</i>	
Urgent, admitted	0.16 [-0.06, 0.35]
Urgent, discharged	-0.12 [-0.19, -0.06]
Non-urgent, admitted	-0.20 [-0.30, -0.11]
Non-urgent, discharged	-0.08 [-0.10, -0.05]
<i>Counterfactual visit type</i>	
Urgent, admitted	0.02 [-0.19, 0.23]
Urgent, discharged	0.05 [-0.07, 0.17]
Non-urgent, admitted	0.02 [-0.05, 0.08]
Non-urgent, discharged	-0.11 [-0.14, -0.10]

Table 3: Change in ED length of stay (LOS) by patient characteristics at 1.5% model sensitivity, the lowest threshold with no significant increase in the prolonged stay rate. Values are means with 95% bootstrap confidence intervals in brackets; negative values indicate shorter stays.

	Change in ED LOS (min)
Overall	-0.07 [-0.11, -0.03]
<i>Sex</i>	
Female	-0.05 [-0.11, 0.01]
Male	-0.08 [-0.11, -0.04]
<i>Race</i>	
Hispanic	-0.12 [-0.15, -0.09]
Non-Hispanic Black	-0.07 [-0.16, 0.00]
Non-Hispanic White	-0.04 [-0.10, 0.02]
Other/Unknown	-0.05 [-0.09, 0.00]
<i>Language</i>	
English	-0.05 [-0.10, -0.01]
Spanish	-0.14 [-0.18, -0.10]
Other	-0.08 [-0.14, -0.02]
<i>Insurance</i>	
Public	-0.11 [-0.15, -0.07]
Private	-0.01 [-0.07, 0.03]

Table 4: Change in ED length of stay (LOS) by visit type at 48.6% model sensitivity, the lowest threshold with a change in the prolonged stay rate not significantly greater than 1%. Values are means with 95% bootstrap confidence intervals in brackets; negative values indicate shorter stays.

	Change in ED LOS (min)
Overall	10.59 [8.09, 13.15]
<i>Original visit type</i>	
Urgent, admitted	52.51 [43.90, 63.15]
Urgent, discharged	11.23 [5.34, 17.93]
Non-urgent, admitted	8.50 [4.12, 13.36]
Non-urgent, discharged	4.27 [2.86, 5.95]
<i>Counterfactual visit type</i>	
Urgent, admitted	38.58 [30.10, 48.67]
Urgent, discharged	35.85 [30.14, 42.58]
Non-urgent, admitted	24.64 [19.48, 30.40]
Non-urgent, discharged	-4.52 [-5.93, -2.98]

Table 5: Change in ED length of stay (LOS) by patient characteristics at 48.6% model sensitivity, the lowest threshold with a change in the prolonged stay rate not significantly greater than 1%. Values are means with 95% bootstrap confidence intervals in brackets; negative values indicate shorter stays.

	Change in ED LOS (min)		
Overall	10.59	[8.09,	13.15]
<i>Sex</i>			
Female	10.89	[8.29,	13.58]
Male	10.31	[7.89,	12.79]
<i>Race</i>			
Hispanic	7.05	[4.82,	9.37]
Non-Hispanic Black	9.86	[7.03,	12.91]
Non-Hispanic White	14.33	[11.39,	17.46]
Other/Unknown	9.99	[7.77,	12.32]
<i>Language</i>			
English	11.64	[9.01,	14.36]
Spanish	4.97	[3.15,	7.35]
Other	9.34	[6.63,	11.55]
<i>Insurance</i>			
Public	8.50	[6.07,	10.84]
Private	13.00	[10.27,	15.91]