Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Predictors of Higher-Intensity Care in Epinephrine-Treated Emergency Department Encounters for Anaphylaxis and Allergic Reactions: A Retrospective Cohort Study With Internal Validation

Ramazan Sami Aktaş
University of Health Sciences, Van Training and Research Hospital
Mehmet Şirin Büyükkaya
University of Health Sciences, Van Training and Research Hospital
Özge Atik
University of Health Sciences, Van Training and Research Hospital
Ömer Okucu
Ağrı İbrahim Çeçen University
Nagihan Adıgüzel
nagihan_adiguzel@hotmail.com

University of Health Sciences, Van Training and Research Hospital

Research Article

Keywords: anaphylaxis, epinephrine, emergency department, risk stratification, multiple imputation, internal validation, decision curve analysis

Posted Date: November 19th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11625521/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Background

Clinicians treating anaphylaxis in the emergency department (ED) have little objective guidance on which patients will need more than a single dose of epinephrine. We examined whether features recorded at the initial ED assessment are associated with higher-intensity care and tested the stability of these associations across approaches to missing data, repeat encounters, and outcome definition.

Methods

Adult ED encounters that received epinephrine for anaphylaxis or an allergic reaction were drawn from MIMIC-IV-ED. The primary outcome was a composite of recurrent epinephrine, intensive care unit (ICU) admission, mechanical ventilation, or vasopressor use. A prespecified seven-variable logistic model (age, sex, ambulance arrival, initial oxygen saturation, asthma, hypertension, food trigger) was fitted with multiple imputation. Sensitivity analyses used complete cases, a saturation-free model on all encounters, a first-encounter cohort, and patient-clustered standard errors. Internal validation nested the imputation step inside bootstrap resampling and repeated 10-fold cross-validation on all encounters. ICU admission, recurrent epinephrine, and life support (ventilation or vasopressor) were modeled separately.

Results

Among 424 encounters (399 patients), 185 (43.6%) met the primary outcome; none died in hospital. Asthma (adjusted odds ratio 3.65 (1.45–9.23)), hypertension (3.45 (1.64–7.29)), older age (1.16 (1.02–1.33) per decade), and a food trigger (2.50 (1.22–5.12)) were associated with the outcome. The food estimate weakened in complete cases (1.71 (0.77–3.80)) because encounters lacking an initial saturation were disproportionately food-triggered and severe; the saturation-free model on all 424 encounters reproduced the association (2.43 (1.19–4.95)). Age, asthma, and hypertension estimates were similar in the first-encounter cohort and with clustered errors. Discrimination was modest (nested cross-validated AUC 0.682; bootstrap-corrected calibration slope 0.90). Asthma, hypertension, and food trigger were more strongly related to ICU admission (AUC 0.883) than to recurrent epinephrine. The lowest tertile of cross-validated risk still had an event rate of 30.3%.

Conclusions

Asthma, hypertension, older age, and a food-related trigger identify epinephrine-treated encounters at higher risk of escalation, particularly ICU-level care. None of the examined risk strata had an event rate low enough to support routine immediate discharge. Timing of ICU-level events relative to the first dose could not be verified; external validation is required.

Background

Epinephrine is the only first-line drug for anaphylaxis, and its ability to reverse acute manifestations is not in doubt.[1–3] A separate question arises after that first dose: which patients will go on to need a second dose, intensive monitoring, airway support, or vasopressors, and which can be observed briefly and sent home? Practice parameters recommend a period of observation but acknowledge that the evidence on duration and patient selection is largely consensus-based.[4,5]

Studies of biphasic reactions have examined this problem indirectly. They report a wide range of incidences and inconsistent risk factors, and they usually treat the biphasic reaction as a phenomenon to describe rather than as a disposition decision to support.[6,7] Observation periods vary between institutions and between clinicians within the same department, and the choice of whom to observe rests heavily on impression.[8–11] Earlier reviews have called for outcome definitions anchored to the escalation of care rather than to symptom recurrence alone.[12–14]

We therefore asked whether characteristics available at the first ED assessment are associated with higher-intensity care in encounters treated with epinephrine. Because the clinical question is a disposition question, we restricted candidate predictors to information available before treatment decisions are made, and we evaluated clinical usefulness with decision-curve analysis rather than discrimination alone. In response to methodological review of an earlier draft, we also examined how the findings behave when missing data are handled differently, when repeat encounters by the same patient are accounted for, and when the composite outcome is separated into its components.

Methods

Study design and data source

This retrospective cohort study used the ED module of the Medical Information Mart for Intensive Care IV (MIMIC-IV), a de-identified database of ED visits to a single academic medical center in the United States between 2011 and 2019 (MIMIC-IV-ED, version 2.2).[15] Individual dates are shifted to protect privacy, so the accrual period of the study cohort within that window cannot be given more precisely. Reporting follows the TRIPOD + AI statement, which supersedes the 2015 TRIPOD checklist (Additional file 1).[16]

Participants and unit of analysis

The analytic extract comprised 2,689 adult (18 years or older) ED encounters for an allergic reaction, angioedema, or anaphylaxis. Encounters in which epinephrine was administered for the allergic presentation form the study cohort; encounters in which epinephrine was given for another indication, such as cardiac arrest, and encounters with insufficient data to determine the primary outcome had been

excluded during extraction (Additional file 1: Figure S1). The unit of analysis is the ED encounter. The 424 eligible encounters came from 399 distinct patients; 21 patients contributed more than one encounter (25 additional encounters, up to four per patient). The main analysis retains all encounters; two sensitivity analyses address within-patient correlation (see below).

The extract records whether epinephrine was given and the door-to-epinephrine interval (median 8 min), but it does not contain time stamps for ICU admission, intubation, or vasopressor initiation. Recurrent epinephrine is, by definition, subsequent to the first dose. For the other outcome components the sequence relative to the first dose could not be established, and we therefore describe the cohort as epinephrine-treated encounters rather than as patients who had clinically improved after treatment.

Outcomes

The primary outcome was higher-intensity care, a composite of any of the following during the index encounter: a second or further dose of epinephrine, ICU admission, mechanical ventilation, or vasopressor use. Each component represents a departure from uncomplicated recovery. A database flag for 24-hour deterioration was found to coincide exactly with this composite and was not used as a separate component.

Three secondary outcomes were analyzed individually: ICU admission; recurrent epinephrine; and a life-support outcome defined as mechanical ventilation or vasopressor use. In this cohort every ventilated or vasopressor-treated encounter was also admitted to the ICU, so a composite of ICU admission, ventilation, and vasopressor use is identical to ICU admission alone. An earlier draft had described that composite as a “hard” outcome; we now report it as ICU admission and present life support as a separate, smaller outcome.

Candidate predictors

Candidate predictors were restricted to information available at the initial ED assessment: age, sex, arrival mode, initial vital signs (systolic blood pressure, heart rate, respiratory rate, oxygen saturation), shock index, documented hypotension and hypoxemia, comorbidities (asthma, chronic obstructive pulmonary disease, coronary artery disease, hypertension), recorded trigger category (food, drug, insect), and triage acuity. Variables that reflect downstream management (admission, observation time, treatments after the initial assessment) were excluded to avoid conditioning on consequences of the outcome. A composite severity flag that was collinear with the outcome by construction was also excluded.

The multivariable model was specified in advance on clinical grounds and contained seven terms: age (per 10 years), female sex, ambulance arrival, initial oxygen saturation (per 1% lower), asthma, hypertension, and a food-related trigger. Oxygenation was entered as a continuous saturation rather than

as the binary hypoxemia flag, which was present in only 11 encounters and produced quasi-complete separation.

Missing data

Demographic, comorbidity, trigger, and outcome variables were complete. Missing values were confined to initial vital signs (Table 4). Initial oxygen saturation, the only model variable with missing values, was absent in 37 encounters (8.7%). We compared encounters with and without a recorded saturation on all model variables and outcomes to characterize the missingness mechanism, recognising that observed data can support but cannot prove a missing-at-random assumption. The prespecified primary analysis used multiple imputation by chained equations with predictive mean matching (20 imputations, 10 burn-in iterations), with the outcome and the other model variables, together with initial blood pressure, heart rate, and respiratory rate, in the imputation model; estimates were combined with Rubin's rules.[17] Two further analyses show how the handling of missing saturation affects the estimates: a complete-case analysis, and a model without the saturation term fitted to all 424 encounters, which involves no imputation because the remaining six variables are complete.

Table 4
Encounters with and without a recorded initial oxygen saturation (the only model variable with missing values).

Characteristic	SpO_2 missing (n = 37)	SpO_2 recorded (n = 387)	p
Age, years, median (IQR)	37 (22–56)	34 (23–50)	0.96
Female sex	23 (62.2%)	260 (67.2%)	0.58
Ambulance arrival	11 (29.7%)	99 (25.6%)	0.56
Asthma	5 (13.5%)	25 (6.5%)	0.17
Hypertension	6 (16.2%)	44 (11.4%)	0.42
Food trigger	11 (29.7%)	33 (8.5%)	< 0.001
Higher-intensity care (primary outcome)	24 (64.9%)	161 (41.6%)	0.009
Recurrent epinephrine	18 (48.6%)	140 (36.2%)	0.16
ICU admission	17 (45.9%)	54 (14.0%)	< 0.001
Vasopressor use	6 (16.2%)	12 (3.1%)	0.003
Life support (ventilation or vasopressor)	12 (32.4%)	21 (5.4%)	< 0.001

Values are n (%) unless stated. Missing initial vital signs in the full cohort: Initial SpO2 37 (8.7%); Initial SBP 43 (10.1%); Initial heart rate 27 (6.4%); Initial respiratory rate 33 (7.8%); Shock index 45 (10.6%); Triage acuity (ESI) 7 (1.7%). Demographic, comorbidity, trigger, and outcome variables were complete.

Statistical analysis

Continuous variables are summarized as medians with interquartile ranges and compared with the Mann–Whitney U test; categorical variables are counts with percentages compared with Fisher's exact test. Univariable and multivariable associations were estimated with logistic regression and are reported as odds ratios with 95% confidence intervals. Collinearity was assessed with variance inflation factors. Linearity of the two continuous terms on the logit scale was checked by adding a restricted cubic spline term with three knots and comparing nested models with a likelihood-ratio test.

Two analyses addressed repeat encounters by the same patient. First, the model was refitted with standard errors clustered by patient. Second, one encounter per patient was selected (the earliest by recorded vital-sign time; where that time was unavailable, the lowest encounter identifier), yielding a first-encounter cohort of 399 patients in which the model was refitted with multiple imputation and in complete cases.

Internal validation was performed on all 424 encounters with the imputation step nested inside the resampling, so that no held-out encounter was imputed with its own outcome.[18] In each training set, missing saturation was imputed by predictive mean matching from a regression that included the outcome (five imputations per fold), and the model was fitted to each imputed training set; in the corresponding held-out set, missing saturation was imputed from a regression fitted on the training set without the outcome, as would be done at the point of care, and the predictions from the five training models were averaged. Stratified 10-fold cross-validation was repeated five times, and the out-of-fold predictions provided the cross-validated AUC, calibration slope and intercept (calibration-in-the-large, estimated with the linear predictor as an offset), Brier score, calibration plot by decile, risk tertiles, and decision curve.[19,20] Bootstrap optimism correction (500 resamples) used the same nesting: a model was fitted to each resample with outcome-informed imputation and evaluated both on that resample and on the original data, with outcome-free imputation at prediction in each case; optimism was the mean difference, subtracted from the apparent value, and the optimism-corrected Brier score was the apparent score plus the mean of (resample Brier − original-data Brier). Because the calibration slope and intercept of a logistic model evaluated on its own development data are 1 and 0 by construction, only resampled values are reported as evidence of calibration. The same bootstrap and cross-validation procedures were also applied to the complete-case subset, in which no imputation is required, as a secondary check (Additional file 1: Table S2).

The decision curve compares the net benefit of using the model to select encounters for observation against observing every patient and observing none, across threshold probabilities from 5% to 60%; these thresholds define the range examined and are not proposed as validated decision cut-points. ICU admission (71 events) and recurrent epinephrine (158 events) were modeled with the same seven predictors and the same imputation procedure. The life-support outcome (33 events) was analyzed as an exploratory model with four clinically chosen predictors (age, asthma, hypertension, food trigger) using Firth's penalized likelihood to reduce small-sample bias; because this model does not include saturation, it was fitted to all encounters without imputation.[21] Food-triggered and other encounters were compared on baseline characteristics and on each outcome component, with risk differences and Newcombe confidence intervals. As a descriptive safety indicator, return to the same ED within 72 hours and 7 days and in-hospital death were tabulated among encounters discharged from the ED.

Analyses were performed in Python 3.12 using statsmodels 0.15 and scikit-learn 1.9. Two-sided p-values below 0.05 were considered statistically significant; no adjustment was made for multiple comparisons, and secondary and exploratory analyses are interpreted accordingly.

Results

Cohort and outcome frequency

The 424 epinephrine-treated encounters had a median age of 34 years (IQR 23–50); 66.7% were female and 25.9% arrived by ambulance. A food trigger was recorded in 10.4%, a drug trigger in 13.0%, and an

insect trigger in 1.4%. Epinephrine was given a median of 8 min (IQR 6–17) after arrival.

Higher-intensity care occurred in 185 encounters (43.6%). Recurrent epinephrine was the most frequent component (158, 37.3%), followed by ICU admission (71, 16.7%), mechanical ventilation (21, 5.0%), and vasopressor use (18, 4.2%) (Fig. 1). All ventilated and all vasopressor-treated encounters were among the ICU admissions; the life-support outcome occurred in 33 encounters (7.8%). There were no in-hospital deaths. 92 encounters (21.7%) were admitted to hospital.

Baseline characteristics

Encounters that progressed to higher-intensity care involved older patients who more often arrived by ambulance, more often had asthma, hypertension, or coronary artery disease, more often had a food or drug trigger, and had slightly lower initial oxygen saturation (Table 1). Blood pressure, heart rate, respiratory rate, shock index, and triage acuity did not differ meaningfully between groups.

Table 1
Baseline characteristics at the initial emergency department assessment, by occurrence of higher-intensity care.

Characteristic	Uncomplicated (n = 239)	Higher-intensity care (n = 185)	p
Age, years, median (IQR)	30 (22–46)	40 (26–55)	< 0.001
Female sex, n (%)	170 (71.1%)	113 (61.1%)	0.04
Ambulance arrival, n (%)	48 (20.1%)	62 (33.5%)	0.002
Initial SBP, mmHg, median (IQR)	132 (120–144)	128 (118–144)	0.24
Initial heart rate, bpm, median (IQR)	90 (80–102)	89 (75–100)	0.16
Initial respiratory rate, /min, median (IQR)	18 (16–20)	18 (16–20)	0.85
Initial SpO_2, %, median (IQR)	100 (98–100)	99 (97–100)	0.007
Shock index, median (IQR)	1 (1–1)	1 (1–1)	0.42
Hypotension, n (%)	9 (3.8%)	15 (8.1%)	0.06
Hypoxemia, n (%)	1 (0.4%)	10 (5.4%)	0.001
Asthma, n (%)	7 (2.9%)	23 (12.4%)	< 0.001
Hypertension, n (%)	12 (5.0%)	38 (20.5%)	< 0.001
Coronary artery disease, n (%)	1 (0.4%)	6 (3.2%)	0.05
Food trigger, n (%)	15 (6.3%)	29 (15.7%)	0.002
Drug trigger, n (%)	23 (9.6%)	32 (17.3%)	0.03
Insect trigger, n (%)	2 (0.8%)	4 (2.2%)	0.41
High triage acuity (ESI 1–2), n (%)	214 (89.5%)	169 (91.4%)	0.62

ESI, Emergency Severity Index; IQR, interquartile range; SBP, systolic blood pressure; SpO_2, peripheral oxygen saturation. Vital-sign rows use available measurements (missingness 6–11%; Table 4). p-values from Mann–Whitney U or Fisher's exact test.

Univariable associations

Age, ambulance arrival, lower initial saturation, documented hypoxemia, asthma, hypertension, and food or drug triggers were each associated with the outcome, whereas initial blood pressure, heart rate, respiratory rate, shock index, and triage acuity were not (Table 2). The very wide interval for hypoxemia reflects its rarity.

Table 2
Univariable associations with higher-intensity care (all 424 encounters; vital-sign rows use available values).

Predictor	Odds ratio (95% CI)	p
Age (per 10 years)	1.28 (1.14–1.45)	< 0.001
Female sex	0.64 (0.42–0.96)	0.03
Ambulance arrival	2.01 (1.29–3.11)	0.002
Initial SBP (per 10 mmHg)	0.93 (0.85–1.03)	0.17
Initial heart rate (per 10 bpm)	0.92 (0.83–1.02)	0.12
Initial respiratory rate (per breath/min)	0.99 (0.92–1.06)	0.82
Initial SpO_2 (per 1% lower)	1.19 (1.06–1.33)	0.003
Shock index (per 0.1)	0.98 (0.88–1.09)	0.74
Hypotension	2.25 (0.96–5.27)	0.06
Hypoxemia	13.60 (1.72–107.23)	0.01
Asthma	4.71 (1.97–11.23)	< 0.001
Coronary artery disease	7.98 (0.95–66.85)	0.06
Hypertension	4.89 (2.47–9.67)	< 0.001
Food trigger	2.78 (1.44–5.35)	0.002
Drug trigger	1.96 (1.11–3.49)	0.02
Insect trigger	2.62 (0.47–14.46)	0.27
High acuity (ESI 1–2)	1.23 (0.64–2.39)	0.53

CI, confidence interval. For SpO_2 the odds ratio is per 1% lower saturation, in the same direction as Table 3. The hypoxemia estimate is unstable because the flag was present in only 11 encounters.

Multivariable model and sensitivity to missing-data handling

In the prespecified multiply imputed model (Table 3), older age (1.16 (1.02–1.33) per decade), asthma (3.65 (1.45–9.23)), hypertension (3.45 (1.64–7.29)), and a food-related trigger (2.50 (1.22–5.12)) were independently associated with higher-intensity care; ambulance arrival was at the margin of significance (1.62 (1.00–2.63)), and lower saturation showed a consistent but non-significant trend (1.11 (0.98–1.27) per 1%). Variance inflation factors were all below 1.3, and spline tests gave no evidence of non-linearity for age (p = 0.43) or saturation (p = 0.74).

Table 3
Adjusted odds ratios (95% CI) for higher-intensity care from the prespecified model under alternative handling of missing data and repeat encounters.

Predictor	Multiple imputation (primary; n = 424)	Complete case (n = 387)	Without SpO_2 term (all 424; no imputation)	Patient-clustered SE (n = 387)	First encounter, multiple imputation (n = 399)
Age (per 10 years)	1.16 (1.02–1.33)	1.18 (1.02–1.36)	1.18 (1.04–1.36)	1.18 (1.02–1.36)	1.16 (1.00–1.33)
Female sex	0.79 (0.50–1.23)	0.77 (0.48–1.22)	0.75 (0.48–1.16)	0.77 (0.48–1.23)	0.76 (0.48–1.20)
Ambulance arrival	1.62 (1.00–2.63)	1.53 (0.93–2.52)	1.73 (1.07–2.78)	1.53 (0.91–2.56)	1.60 (0.98–2.63)
Initial SpO_2 (per 1% lower)	1.11 (0.98–1.27)	1.11 (0.98–1.26)	—	1.11 (0.98–1.26)	1.12 (0.98–1.27)
Asthma	3.65 (1.45–9.23)	2.99 (1.14–7.84)	3.66 (1.45–9.22)	2.99 (1.14–7.86)	4.94 (1.72–14.17)
Hypertension	3.45 (1.64–7.29)	2.90 (1.34–6.24)	3.43 (1.63–7.19)	2.90 (1.29–6.52)	3.35 (1.53–7.32)
Food-related trigger	2.50 (1.22–5.12)	1.71 (0.77–3.80)	2.43 (1.19–4.95)	1.71 (0.79–3.74)	2.16 (1.00–4.66)
Events / encounters	185 / 424	161 / 387	185 / 424	161 / 387	168 / 399
Apparent AUC	0.703	0.677	0.697	0.677	0.701

AUC, area under the curve; SE, standard error. Multiple imputation: chained equations, 20 imputations, Rubin's rules; AUC averaged across imputed datasets. The complete-case and clustered models use identical rows; the clustered model differs only in the standard errors. The first-encounter cohort retains one encounter per patient.

The complete-case model (387 encounters, 161 events) gave similar estimates for age, asthma, and hypertension but a smaller and non-significant estimate for the food trigger (1.71 (0.77–3.80)). Table 4 shows why. The 37 encounters without a recorded initial saturation differed systematically from the rest: a food trigger was recorded in 29.7% of them compared with 8.5% of complete encounters ($p < 0.001$); 64.9% reached the primary outcome compared with 41.6% ($p = 0.009$); and 45.9% were admitted to the ICU compared with 14.0% ($p < 0.001$). 12 of the 33 life-support events fell in this group. Excluding these encounters therefore removes a disproportionate share of severe food-triggered presentations. Consistent with this, the crude association between food trigger and the outcome is OR 2.78 ($p = 0.002$) in all 424 encounters, in which neither variable has missing values, and OR 2.03 ($p = 0.06$) in complete cases. A model without the saturation term, fitted to all 424 encounters without any imputation, reproduced the imputed estimates (food trigger 2.43 (1.19–4.95); asthma 3.66 (1.45–9.22);

hypertension 3.43 (1.63–7.19)). The difference between the complete-case and imputed results thus arises from informative exclusion of encounters rather than from the imputation procedure.

Repeat encounters

Clustering standard errors by patient changed the confidence intervals only marginally (Table 3). In the first-encounter cohort (399 patients, 168 events), the imputed model gave estimates of 1.16 (1.00–1.33) for age, 4.94 (1.72–14.17) for asthma, 3.35 (1.53–7.32) for hypertension, and 2.16 (1.00–4.66) for food trigger, the last lying at the margin of significance. Within-patient correlation therefore does not account for the principal associations.

Model performance

The apparent AUC of the multiply imputed model, averaged over the imputed datasets, was 0.703 (Table 3); because those imputations draw on the outcome, this value overstates performance at the point of use. With the imputation step nested inside resampling on all 424 encounters, repeated 10-fold cross-validation gave an AUC of 0.682 (95% CI 0.632–0.735; range across repeats 0.671–0.687), a calibration slope of 0.85 (95% CI 0.59–1.11), an intercept of 0.00, and a Brier score of 0.219; mean cross-validated predicted risk (43.6%) matched the observed rate (43.6%) (Figs. 2 and 3). Nested bootstrap gave an optimism-corrected AUC of 0.687 (apparent 0.704; optimism 0.016), a corrected slope of 0.90 (95% CI 0.67–1.17), a corrected intercept of -0.00, and a Brier score of 0.212 apparent and 0.204 corrected. In the complete-case subset (387 encounters) the apparent AUC was 0.677 (95% CI 0.622–0.733), the bootstrap-corrected AUC 0.653, and the cross-validated AUC 0.642 with a slope of 0.78; the lower values in this subset are consistent with its exclusion of the most severe presentations (Additional file 1: Table S2). A slope below 1 in every analysis indicates mild overfitting, with predictions at the extremes slightly too extreme.

Risk strata

Observed event rates rose across tertiles of nested cross-validated predicted risk in the full cohort, from 30.3% (43/142) in the lowest tertile to 35.5% (50/141) and 65.2% (92/141) in the highest (Fig. 4). The lowest tertile still experienced the outcome in roughly three of every ten encounters. None of the examined risk strata had an event rate low enough to support routine immediate discharge; this statement applies to the strata examined and does not exclude the existence of lower-risk groups defined by other means.

Decision-curve analysis

Using nested cross-validated predictions for all 424 encounters, the model yielded a higher net benefit than observing all patients across threshold probabilities from about 0.26 to 0.60, and a higher net benefit than observing none throughout (Fig. 5). Below a threshold of about 0.26 the curves for the model and for observing everyone coincide, which means that a clinician unwilling to accept a missed-escalation risk above roughly one in four would gain nothing from the model and should observe all

patients. The thresholds were examined for research purposes and are not proposed as clinical decision rules. The net benefit was estimated in the development data and does not establish that model-guided disposition would reduce adverse outcomes in practice.

Outcome-specific models

The predictors behaved differently across outcome components (Table 5). For ICU admission, asthma (7.29 (2.87–18.49)), hypertension (10.12 (4.62–22.16)), food trigger (5.83 (2.45–13.89)), and age (1.38 (1.14–1.68)) had larger effects than for the primary composite, and discrimination was higher (apparent AUC 0.877; bootstrap-corrected 0.858; corrected calibration slope 0.90). For recurrent epinephrine, only ambulance arrival (1.79 (1.12–2.84)) was clearly associated; asthma (1.55 (0.70–3.42)) and hypertension (1.07 (0.55–2.10)) were not, and discrimination was poor (AUC 0.643). In the exploratory life-support model (33 events), hypertension (6.07 (2.39–15.41)), food trigger (8.94 (3.49–22.88)), and age (1.34 (1.05–1.70)) were associated with ventilation or vasopressor use; asthma was not (2.54 (0.89–7.27)). The comorbidity and trigger signals in the composite therefore derive mainly from ICU-level escalation rather than from the need for a second dose of epinephrine.

Table 5
Adjusted odds ratios (95% CI) for the individual outcome components.

Predictor	ICU admission (MI; 71 events)	Recurrent epinephrine (MI; 158 events)	Life support, exploratory (Firth; 33 events)
Age (per 10 years)	1.38 (1.14–1.68)	1.12 (0.98–1.28)	1.34 (1.05–1.70)
Female sex	0.82 (0.42–1.61)	0.86 (0.55–1.33)	—
Ambulance arrival	1.27 (0.62–2.58)	1.79 (1.12–2.84)	—
Initial SpO_2 (per 1% lower)	1.08 (0.90–1.30)	1.12 (0.99–1.27)	—
Asthma	7.29 (2.87–18.49)	1.55 (0.70–3.42)	2.54 (0.89–7.27)
Hypertension	10.12 (4.62–22.16)	1.07 (0.55–2.10)	6.07 (2.39–15.41)
Food-related trigger	5.83 (2.45–13.89)	1.94 (0.99–3.81)	8.94 (3.49–22.88)
Apparent AUC	0.877 (corrected 0.858)	0.643	0.887

MI, multiple imputation (20 imputations, all 424 encounters). The life-support model used four prespecified clinical predictors and Firth's penalized likelihood on all 424 encounters (no saturation term, hence no missing values) and is exploratory. Every ventilated or vasopressor-treated encounter was also an ICU admission, so ICU admission is identical to a composite of ICU admission, ventilation, and vasopressor use.

Food-triggered compared with other encounters

Patients in food-triggered encounters (44, 10.4%) were younger (median 26 vs 34 years, p = 0.04), more ambulance arrivals (47.7% vs 23.4%, p < 0.001), and more asthma (20.5% vs 5.5%, p = 0.002), with a similar prevalence of hypertension (11.4% vs 11.8%) (Table 6). The primary outcome occurred in 65.9% of food-triggered and 41.1% of other encounters (risk difference 24.9 percentage points, 95% CI 9.3 to 38.0). ICU admission (40.9% vs 13.9%; risk difference 27.0, 95% CI 13.2 to 42.0) and vasopressor use (22.7% vs 2.1%; risk difference 20.6, 95% CI 10.5 to 34.9) accounted for most of the excess. These comparisons rest on 44 food-triggered encounters and on 10 vasopressor events in that group, and the intervals are correspondingly wide.

Table 6
Food-triggered compared with other epinephrine-treated encounters.

Characteristic / outcome	Food trigger (n = 44)	Other trigger (n = 380)	Risk difference, % (95% CI)	p
Age, years, median (IQR)	26 (20–48)	34 (24–50)	—	0.04
Female sex	26 (59.1%)	257 (67.6%)	-8.5 (-23.9 to 5.5)	0.31
Ambulance arrival	21 (47.7%)	89 (23.4%)	24.3 (9.6 to 39.2)	< 0.001
Asthma	9 (20.5%)	21 (5.5%)	14.9 (5.2 to 29.1)	0.002
Hypertension	5 (11.4%)	45 (11.8%)	-0.5 (-7.8 to 12.5)	1.00
Initial SpO_2, %, median (IQR)	99 (99–100)	99 (98–100)	—	0.94
Higher-intensity care (primary)	29 (65.9%)	156 (41.1%)	24.9 (9.3 to 38.0)	0.002
Recurrent epinephrine	24 (54.5%)	134 (35.3%)	19.3 (4.0 to 33.8)	0.02
ICU admission	18 (40.9%)	53 (13.9%)	27.0 (13.2 to 42.0)	< 0.001
Mechanical ventilation	4 (9.1%)	17 (4.5%)	4.6 (-1.5 to 16.8)	0.26
Vasopressor use	10 (22.7%)	8 (2.1%)	20.6 (10.5 to 34.9)	< 0.001
Life support (ventilation or vasopressor)	12 (27.3%)	21 (5.5%)	21.7 (10.5 to 36.4)	< 0.001

Risk differences (food minus other) with Newcombe 95% confidence intervals; p-values from Fisher's exact test or Mann–Whitney U test. Saturation medians use available values.

Post-discharge return

Of the 424 encounters, 332 (78.3%) were discharged from the ED (327 home; 5 eloped or left against advice) and 92 were admitted. Among discharged encounters, 10 (3.0%) returned to the same ED within 72 hours and 12 (3.6%) within 7 days. In the lowest tertile of cross-validated risk, 5 of 134 discharged encounters (3.7%) returned within 72 hours. Presentations to other hospitals and deaths outside hospital are not captured, so these figures describe same-institution returns only and cannot be read as evidence that early discharge was safe.

Discussion

Principal findings

In 424 ED encounters treated with epinephrine, 43.6% progressed to higher-intensity care. Four features recorded at presentation (older age, asthma, hypertension, and a food-related trigger) were associated with that progression, and the associations survived complete-case analysis, patient-level clustering, and restriction to first encounters, with the partial exception of the food trigger, whose estimate depends on how missing saturation values are handled. The model discriminated modestly and, after correction for optimism, was mildly overfitted. Its most useful property may be a negative one: none of the risk strata examined had an event rate low enough to justify immediate discharge without observation.

Separating the composite into its components clarified what the predictors measure. Asthma, hypertension, and food trigger were weakly related to recurrent epinephrine but strongly related to ICU admission and, in an exploratory analysis, to ventilation or vasopressor use. Ambulance arrival behaved in the opposite way. The composite therefore mixes two signals, a frequent but softer need for a second dose that is partly clinician-dependent, and a rarer escalation to organ support; the comorbidity and trigger variables track the latter.

Why complete-case and imputed estimates differed

A methodological review of an earlier draft noted that the food-trigger odds ratio fell from about 2.5 in the imputed analysis to about 1.7 in complete cases and asked whether imputation had inflated it. The present analysis shows the reverse. Encounters with no recorded initial saturation were not a random subset: they were three to four times more likely to be food-triggered, more likely to reach the outcome, and three times more likely to be admitted to the ICU, and they contained 12 of the 33 life-support events. An initial saturation is most likely to go unrecorded when a patient is taken directly to resuscitation, so complete-case analysis preferentially removes the sickest food-triggered presentations. Because both the trigger and the outcome are fully observed, the association can be estimated on the full cohort without imputation by dropping the saturation term, and that estimate matches the imputed one. The pattern is compatible with data missing at random given observed variables, the assumption under which multiple imputation is valid and complete-case analysis is biased.[17] It does not prove that assumption: a mechanism in which missingness also depends on the unrecorded saturation itself (missing not at random) cannot be excluded with these data, so the explanation offered here is the most plausible one rather than a demonstrated one. We therefore present all four analyses, so that readers can see how much of any estimate depends on this choice.

Interpretation of the predictors

Older age, asthma, and hypertension are plausible markers of reduced physiological reserve and of airway involvement, and they recur as risk factors in the biphasic-reaction literature.[3,7] Hypertension in this setting is probably a proxy for cardiovascular comorbidity rather than a cause of escalation. The

food-trigger association is biologically credible, since ingested allergen continues to be absorbed for hours and food is the trigger most consistently linked to biphasic reactions, and the subgroup comparison shows that it is not explained by a difference in hypertension.[6,7,12] Even so, the estimate rests on a small number of food-triggered encounters and a handful of vasopressor events, its magnitude varies with the handling of missing data, and we regard it as a hypothesis for external testing rather than a settled finding. Drug and insect triggers were associated with the outcome only in univariable analysis.

Initial oxygen saturation showed a consistent gradient that did not reach significance. Documented hypoxemia was present in 11 encounters and could not be modeled stably; saturation at presentation should not be discounted on that account.

Clinical implications

The findings support a graded approach to observation rather than a binary rule. Patients in the highest risk tertile had an event rate near 65.2% and are reasonable candidates for prolonged monitored observation with a low threshold for ICU consultation, particularly when asthma, hypertension, or a food trigger is present. Patients in the lowest tertile still had an event rate of about 30.3%, which argues for a defined observation period with explicit return instructions rather than discharge on the basis of presentation features alone.[22,23] These implications concern initial features associated with higher-intensity care; they do not describe deterioration after documented improvement, which this dataset cannot identify. The decision curve indicates that model-guided observation offers net benefit only for clinicians willing to accept a threshold probability above about 0.26; below that, observing everyone is the better policy. The costs of the two errors are not symmetric: unnecessary observation consumes time and beds, whereas premature discharge of a patient who then deteriorates may cause serious harm, and the thresholds examined here should be read with that asymmetry in mind.

Comparison with prior work

Previous ED studies have identified severe initial presentation, delayed or repeated epinephrine, and food triggers as correlates of biphasic reactions.[6,7,12,13] Our study differs in anchoring the outcome to escalation of care, in restricting predictors to what is known at presentation, in quantifying usefulness with decision curves, and in reporting how the results change under alternative assumptions about missing data and repeat encounters. Pediatric work has approached the same disposition problem by identifying children at low risk of inpatient therapy;[11] in this adult cohort no comparably low-risk stratum emerged.

Strengths and limitations

Strengths include a prespecified model, exclusion of post-treatment variables, multiple internal-validation strategies, and transparent reporting of sensitivity analyses that were prompted by independent review. Several limitations apply.

First, the extract contains no time stamps for ICU admission, intubation, or vasopressor initiation. Recurrent epinephrine is necessarily subsequent to the first dose, but for the other components we cannot exclude that escalation began before or at the same time as the first dose; the predictors should therefore be understood as markers of a severe course in epinephrine-treated encounters, not as predictors of deterioration after documented improvement. Second, the database does not record symptom resolution, so clinical improvement after the first dose was not verified. Third, ICU admission is a decision influenced by bed availability and institutional practice as well as physiology; the life-support outcome is more objective but had only 33 events. Fourth, anaphylaxis could not be distinguished from other epinephrine-treated allergic reactions with the available fields, so the predictors may partly reflect initial severity. Fifth, the data come from a single center with a critical-care orientation, and the model has not been validated externally; its discrimination is modest and its calibration slope below 1, and it should not be deployed as a decision rule. Sociodemographic fairness beyond sex was not examined. Sixth, the unit of analysis is the encounter, although sensitivity analyses limited to first encounters and with clustered standard errors gave consistent results. Finally, return-visit data capture only the index institution, and the subgroup estimates for food-triggered encounters and for life support rest on small numbers.

Conclusions

Among adult ED encounters treated with epinephrine for anaphylaxis or an allergic reaction, asthma, hypertension, older age, and a food-related trigger identified patients at increased risk of higher-intensity care, and particularly of ICU-level escalation. The associations held across analyses of repeat encounters and of missing-data handling, with the food-trigger estimate most sensitive to the latter for reasons that are now understood. None of the examined risk strata was at sufficiently low risk to justify routine immediate discharge. External validation, ideally with time-stamped escalation events, is needed before these findings inform practice.

Abbreviations

AUC
area under the receiver operating characteristic curve
CI
confidence interval
ED
emergency department
ICU
intensive care unit
IQR
interquartile range
MIMIC-IV
Medical Information Mart for Intensive Care IV

OR
odds ratio
SpO_2
peripheral oxygen saturation
TRIPOD + AI
Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis,artificial-intelligence extension
VIF
variance inflation factor.

Declarations

Ethics approval and consent to participate:

The study used MIMIC-IV, a de-identified, publicly available database. Its creation was approved by the Institutional Review Boards of Beth Israel Deaconess Medical Center and the Massachusetts Institute of Technology, with a waiver of individual consent. The authors completed the required training and signed the data-use agreement. No additional approval was required for this secondary analysis, which conformed to the Declaration of Helsinki.

Consent for publication:

Not applicable.

Funding:

None.

Author Contribution

R.S.A. conceived and designed the study, developed the analysis plan, performed the statistical analyses, and drafted the manuscript. Ö.A. contributed to the conception of the study, validated the clinical definitions, and interpreted the findings from an allergy and immunology perspective. Ö.O. extracted and curated the data and contributed to the software, formal analysis, and visualization. M.Ş.B. and N.A. contributed to the clinical interpretation of the results from an emergency medicine perspective and to the revision of the manuscript at several stages of its development. All authors critically revised the manuscript for important intellectual content, read and approved the final version, and agree to be accountable for all aspects of the work.

Acknowledgements:

Not applicable.

Data Availability

The data supporting the findings of this study were obtained from the Medical Information Mart for Intensive Care IV Emergency Department (MIMIC-IV-ED) database, available through PhysioNet (https://physionet.org/content/mimic-iv-ed/). Access to the database is granted to qualified researchers who complete the required credentialing process and agree to the applicable data use agreement. Due to these data use restrictions, the authors cannot publicly redistribute individual-level patient data. The analysis code is provided as Additional file 2 to facilitate reproducibility.

References

1. Golden DBK, Wang J, Waserman S, Akin C, Campbell RL, Ellis AK, et al. Anaphylaxis: a 2023 practice parameter update. Ann Allergy Asthma Immunol. 2024;132(2):124–76. 10.1016/j.anai.2023.09.015.
2. Cardona V, Ansotegui IJ, Ebisawa M, El-Gamal Y, Fernandez Rivas M, Fineman S, et al. World Allergy Organization anaphylaxis guidance 2020. World Allergy Organ J. 2020;13(10):100472. 10.1016/j.waojou.2020.100472.
3. Shaker MS, Wallace DV, Golden DBK, Oppenheimer J, Bernstein JA, Campbell RL, et al. Anaphylaxis—a 2020 practice parameter update, systematic review, and Grading of Recommendations, Assessment, Development and Evaluation (GRADE) analysis. J Allergy Clin Immunol. 2020;145(4):1082–123. 10.1016/j.jaci.2020.01.017.
4. Wang J, Lieberman JA, Wallace DV, Waserman S, Golden DBK. Anaphylaxis in practice: a guide to the 2023 practice parameter update. J Allergy Clin Immunol Pract. 2024;12(9):2325–36. 10.1016/j.jaip.2024.06.036.
5. Weller KN, Hsieh FH. Anaphylaxis: highlights from the practice parameter update. Cleve Clin J Med. 2022;89(2):106–11. 10.3949/ccjm.89a.21076.
6. Grunau BE, Li J, Yi TW, Stenstrom R, Grafstein E, Wiens MO, et al. Incidence of clinically important biphasic reactions in emergency department patients with allergic reactions or anaphylaxis. Ann Emerg Med. 2014;63(6):736–44. 10.1016/j.annemergmed.2013.10.017.
7. Lee S, Bellolio MF, Hess EP, Campbell RL. Predictors of biphasic reactions in the emergency department for patients with anaphylaxis. J Allergy Clin Immunol Pract. 2014;2(3):281–7. 10.1016/j.jaip.2014.01.012.
8. Campbell RL, Li JTC, Nicklas RA, Sadosty AT. Emergency department diagnosis and treatment of anaphylaxis: a practice parameter. Ann Allergy Asthma Immunol. 2014;113(6):599–608. 10.1016/j.anai.2014.10.007.
9. Lieberman P, Nicklas RA, Randolph C, Oppenheimer J, Bernstein D, Bernstein J, et al. Anaphylaxis—a practice parameter update 2015. Ann Allergy Asthma Immunol. 2015;115(5):341–84. 10.1016/j.anai.2015.07.019.

10. Short HB, Walters B, Fabi M, Ravida N, Boehmer S, Reyes L. Evaluating practice patterns of observation periods following epinephrine administration for anaphylaxis among pediatric patients. Cureus. 2024;16(9):e69419. 10.7759/cureus.69419.
11. Dribin TE, Michelson KA, Monuteaux MC, Stack AM, Farbman KS, Schneider LC, et al. Identification of children with anaphylaxis at low risk of receiving acute inpatient therapies. PLoS ONE. 2019;14(2):e0211949. 10.1371/journal.pone.0211949.
12. Tole JW, Lieberman P. Biphasic anaphylaxis: review of incidence, clinical predictors, and recommendations for observation. Immunol Allergy Clin North Am. 2007;27(2):309–26. 10.1016/j.iac.2007.03.011.
13. Douglas DM, Sukenick E, Andrade WP, Brown JS. Biphasic systemic anaphylaxis: an inpatient and outpatient study. J Allergy Clin Immunol. 1994;93(6):977–85. 10.1016/S0091-6749(94)70044-3.
14. Simard D, Bouchard V, Plourde A, Lefebvre S, Herman-Lemelin A, Lapointe S, et al. Factors influencing emergency department observation time following anaphylaxis: a systematic review. CJEM. 2021;23(4):480–93. 10.1007/s43678-021-00112-z.
15. Johnson AEW, Bulgarelli L, Shen L, Gayles A, Shammout A, Horng S, et al. MIMIC-IV, a freely accessible electronic health record dataset. Sci Data. 2023;10(1):1. 10.1038/s41597-022-01899-x.
16. Collins GS, Moons KGM, Dhiman P, Riley RD, Beam AL, Van Calster B, et al. TRIPOD + AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. BMJ. 2024;385:e078378. 10.1136/bmj-2023-078378.
17. Sterne JAC, White IR, Carlin JB, Spratt M, Royston P, Kenward MG, et al. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. BMJ. 2009;338:b2393. 10.1136/bmj.b2393.
18. Steyerberg EW, Harrell FE Jr, Borsboom GJJM, Eijkemans MJC, Vergouwe Y, Habbema JDF. Internal validation of predictive models: efficiency of some procedures for logistic regression analysis. J Clin Epidemiol. 2001;54(8):774–81. 10.1016/S0895-4356(01)00341-9.
19. Van Calster B, McLernon DJ, van Smeden M, Wynants L, Steyerberg EW. Calibration: the Achilles heel of predictive analytics. BMC Med. 2019;17(1):230. 10.1186/s12916-019-1466-7.
20. Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. Med Decis Mak. 2006;26(6):565–74. 10.1177/0272989X06295361.
21. Firth D. Bias reduction of maximum likelihood estimates. Biometrika. 1993;80(1):27–38. 10.1093/biomet/80.1.27.
22. Campbell RL, Luke A, Weaver AL, St Sauver JL, Bergstralh EJ, Li JT, et al. Prescriptions for self-injectable epinephrine and follow-up referral in emergency department patients presenting with anaphylaxis. Ann Allergy Asthma Immunol. 2008;101(6):631–6. 10.1016/S1081-1206(10)60227-X.
23. Pumphrey RSH. Lessons for the management of anaphylaxis from a study of fatal reactions. Clin Exp Allergy. 2000;30(8):1144–50. 10.1046/j.1365-2222.2000.00864.x.

Figures

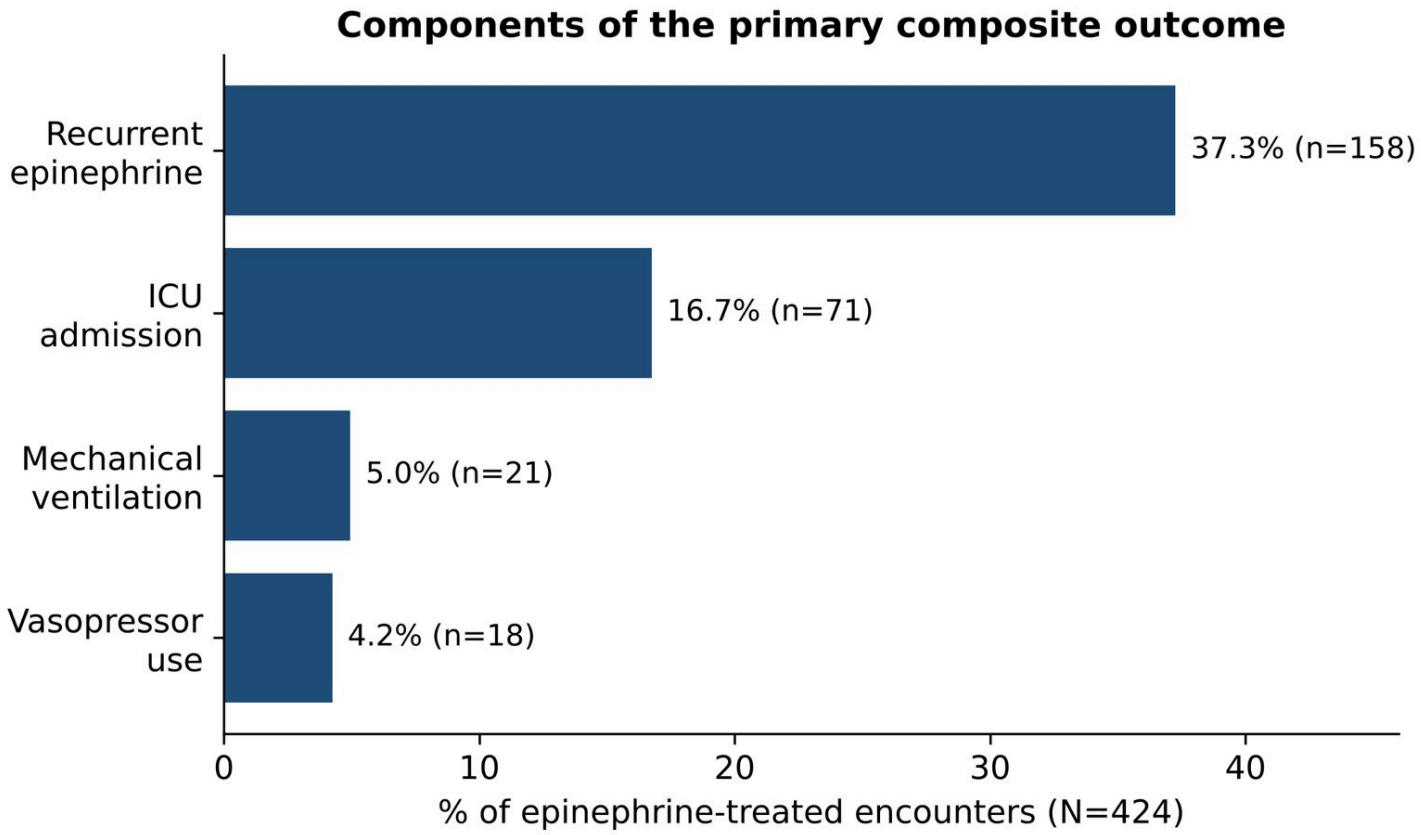


Figure 1

Components of the primary composite outcome among 424 epinephrine-treated encounters. Components overlap; percentages refer to the full cohort. All ventilated and vasopressor-treated encounters were admitted to the ICU.

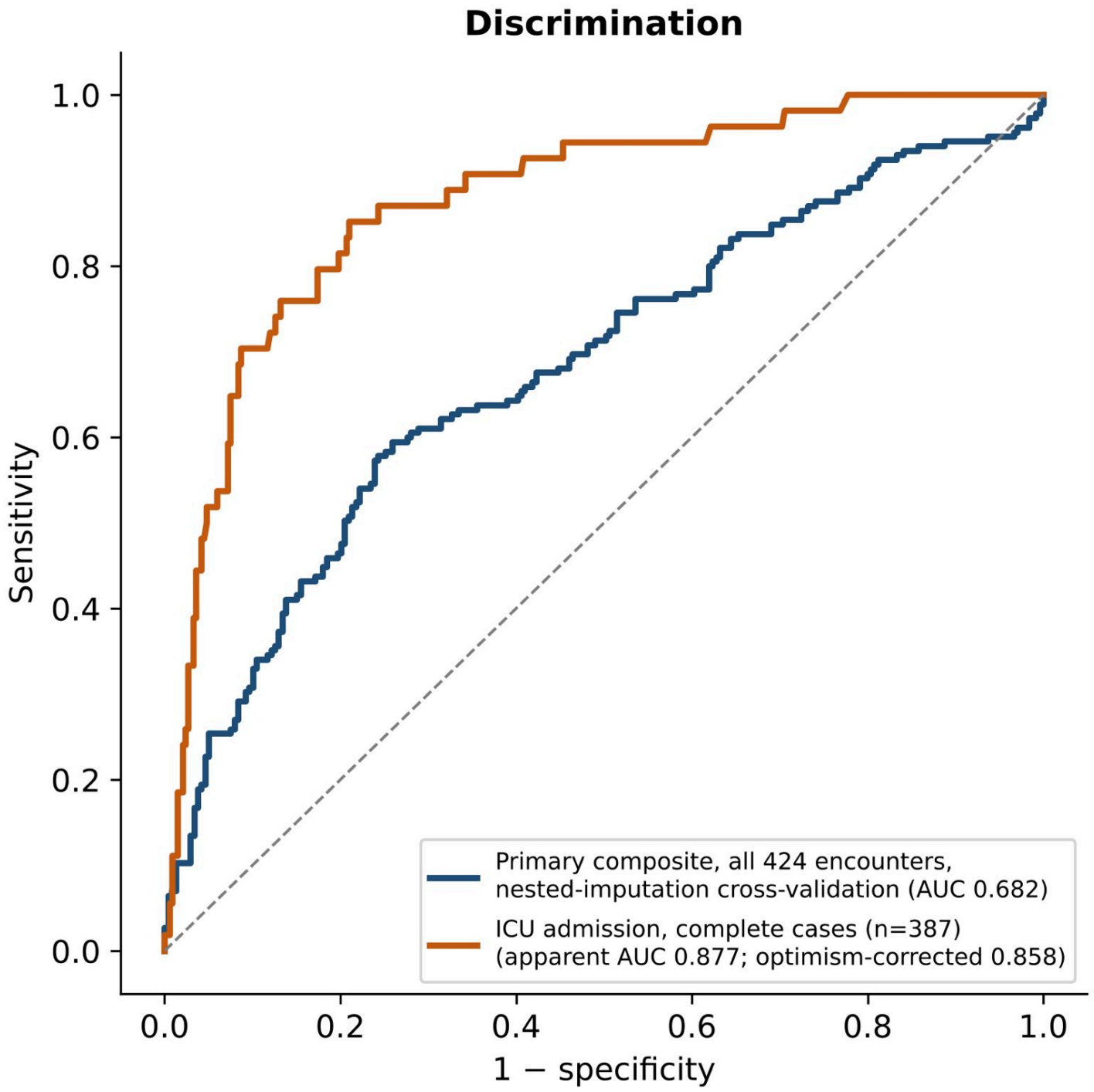


Figure 2

Receiver operating characteristic curves for the prespecified seven-variable model. The primary-composite curve is based on nested-imputation cross-validated predictions for all 424 encounters; the ICU-admission curve is based on the complete-case cohort (n=387), with apparent and bootstrap optimism-corrected AUCs shown.

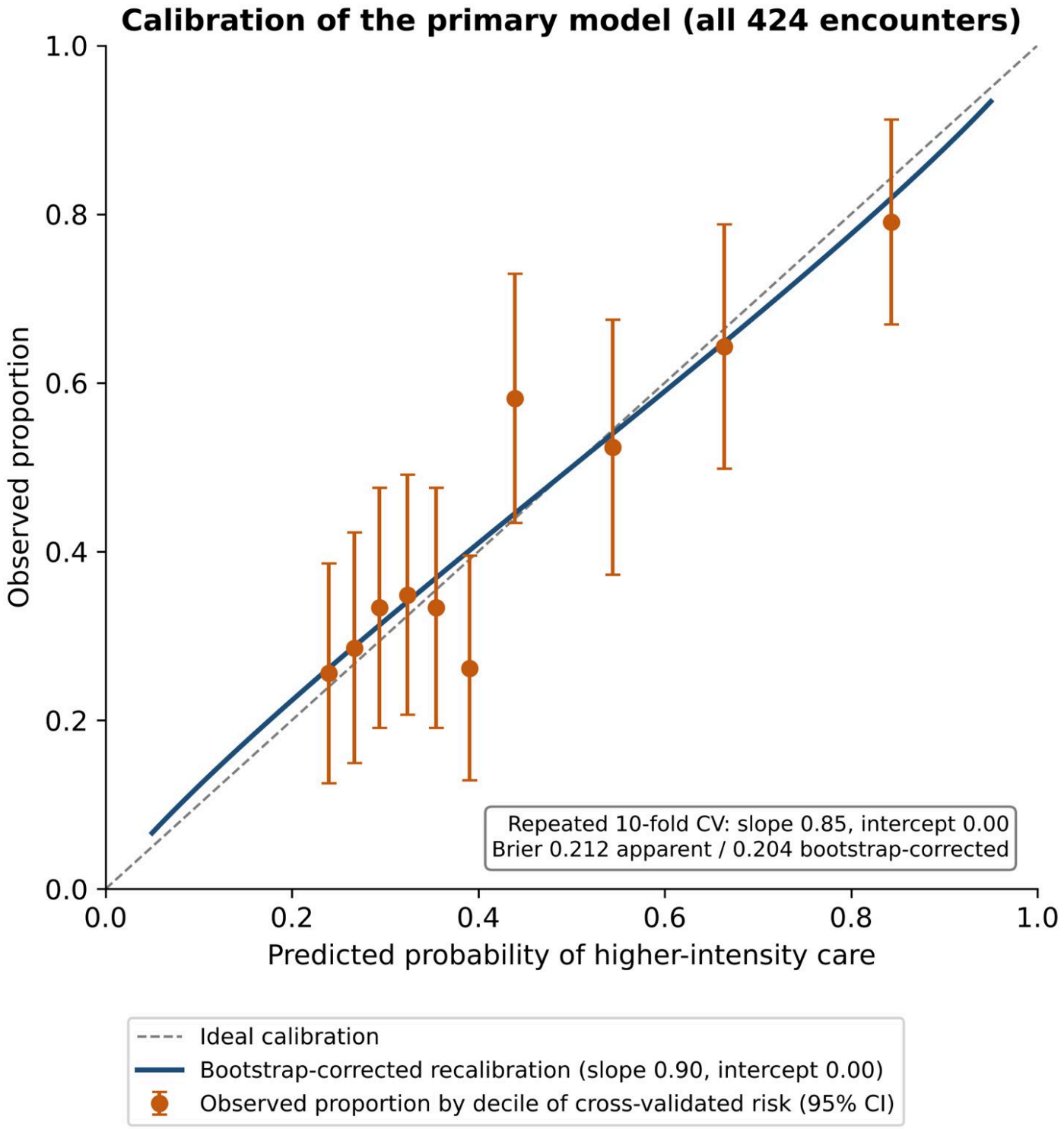


Figure 3

Calibration of the primary model in all 424 encounters. Points show observed event proportions by decile of nested-imputation cross-validated predicted risk with 95% confidence intervals; the solid line is the bootstrap-corrected recalibration line (500 nested resamples); the dashed line is ideal calibration. The optimism-corrected Brier score equals the apparent score plus the mean of (resample Brier − original-data Brier).

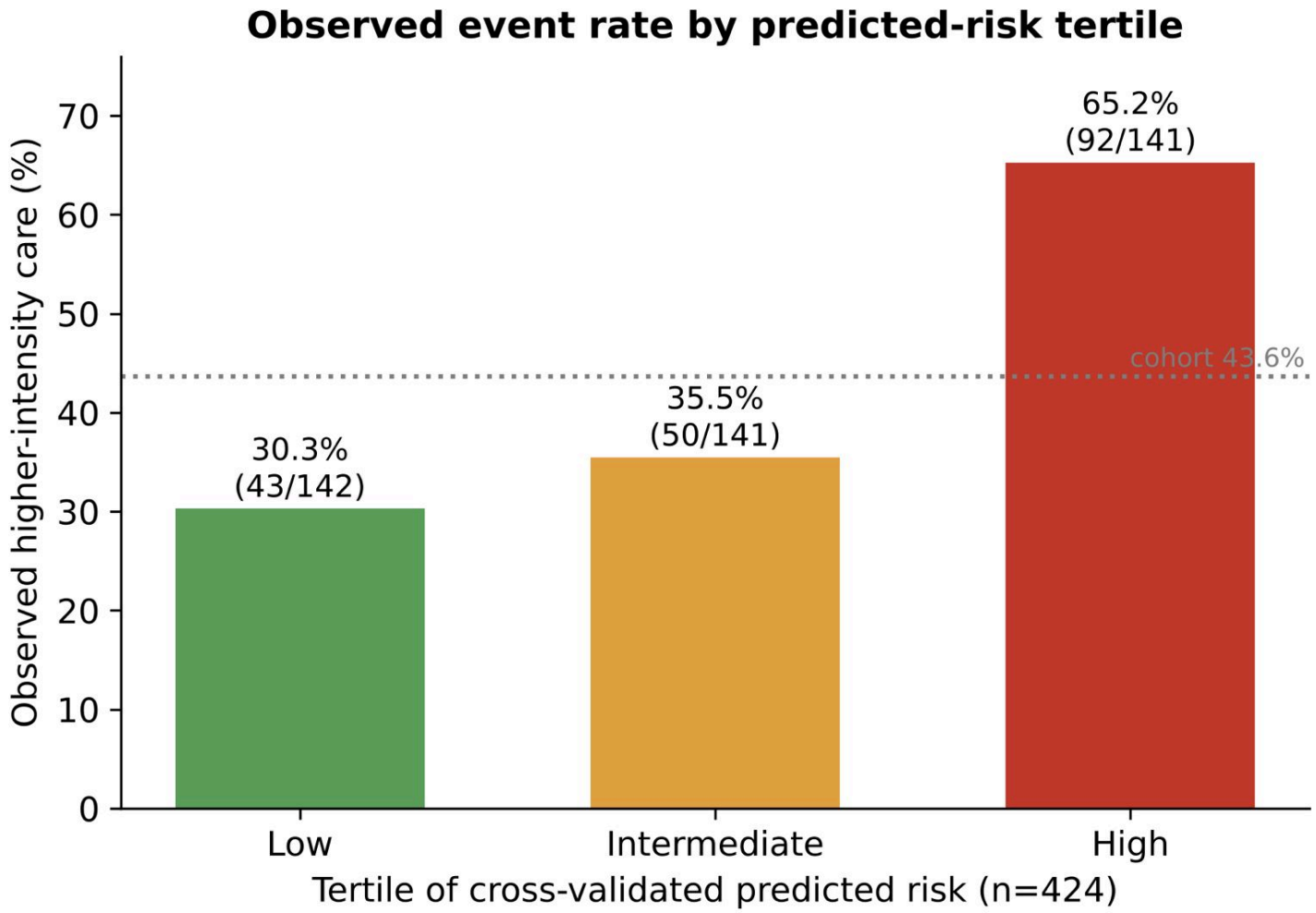


Figure 4

Observed rate of higher-intensity care by tertile of nested-imputation cross-validated predicted risk (all 424 encounters). The dotted line marks the cohort rate.

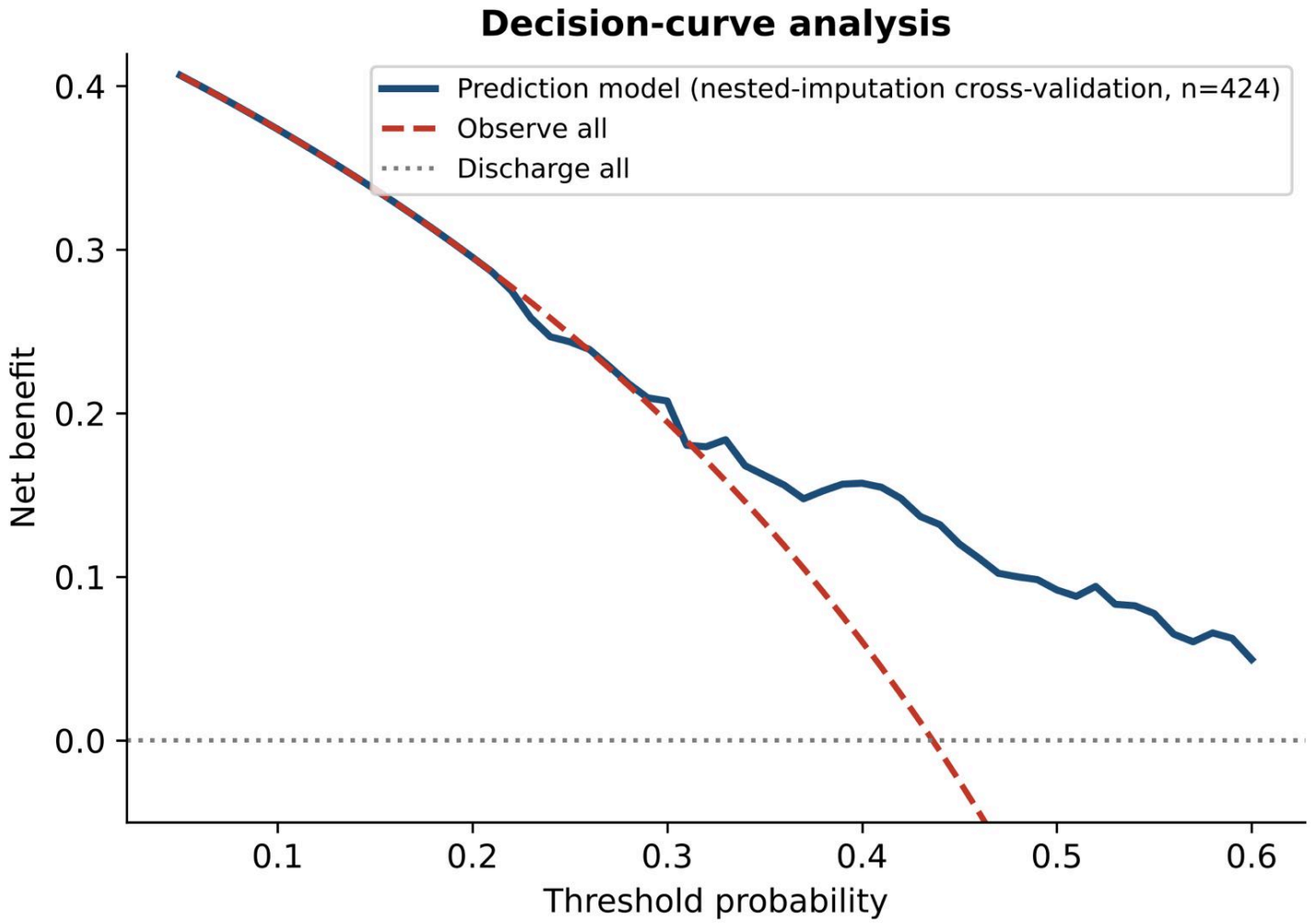


Figure 5

Decision-curve analysis based on nested-imputation cross-validated predictions for all 424 encounters, comparing the net benefit of model-guided observation with the strategies of observing all patients and observing none across threshold probabilities of 0.05 to 0.60. Thresholds were examined for research purposes and are not proposed as clinical decision rules.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Additionalfile1.docx