CURED (CU Risk Estimator for Discharge): Development and Internal Validation of a Machine Learning Tool for Safe Early Patient Discharge During Hospital Surge

Jason Persoff

jason.persoff@cuanschutz.edu

University of Colorado Anschutz Medical Campus

Guannan Shen

University of Colorado Anschutz Medical Campus

Yaxu Zhuang

University of Colorado Anschutz Medical Campus

Jessica Ryder

University of Colorado Anschutz Medical Campus

Pranav Shetty

University of California, Davis

Kevin Yeskey

MDB, Inc.

Kaitlin Rainwater-Lovett

Uniformed Services University of the Health Sciences

Jeffrey Freeman

Uniformed Services University of the Health Sciences

Charles Little

University of Colorado Anschutz Medical Campus

Angela Keniston

University of Colorado Anschutz Medical Campus

Article

Keywords:

Posted Date: May 11th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9502083/v1

Additional Declarations: No competing interests reported.

Abstract

Introduction

A large-scale combat operation (LSCO) returning 1,000 casualties a day for at least 100 days to the US homeland would rapidly saturate domestic hospital capacity and potentially trigger crisis standards of care (CSC). Early patient discharge (EPD), the deliberate identification and release of lower-acuity inpatients before their expected discharge, is among the most actionable throughput strategies available during surge. However, no validated, data-driven tool currently exists to support safe EPD decisions at scale.

Objective

To develop and internally validate a machine learning model (MLM) for identifying inpatients at low risk for composite 30-day adverse outcomes—mortality, hospital readmission, and emergency department revisit—as candidates for EPD during hospital surge.

Design, Setting, and Participants / Statistical Analysis

This multi-site retrospective cohort study included 139,166 adult inpatients and observation patients (182,456 encounters) across 12 UCHealth hospitals in Colorado from January 2023 through December 2024. Candidate predictors included demographics, Elixhauser comorbidity index (ECI), vital signs, laboratory values, functional status scores (AM-PAC 6-Clicks, JH-HLM), Glasgow Coma Scale, length of stay, and prior hospitalization history. Four daily snapshot datasets were constructed at discharge (day 0) and 1–3 days prior. Ridge logistic regression, random forest, generalized additive models, and LightGBM were trained for each of three outcomes and four snapshot days (48 total models) using 10-fold cross-validation with PR-AUC-guided hyperparameter optimization. Final EPD classification required all selected models across all three outcomes to predict no adverse event (all-negative consensus rule).

Results

The model classified 20.7–24.7% of patients as safe for EPD. Among this group, 30-day mortality was 0.3–1.0% (negative predictive value [NPV] 99.1–99.7%), representing a 62.9–83.3% relative reduction from the overall test cohort mortality of 1.8–2.6%. Thirty-day readmission rates were 6.4–9.4% (NPV 90.6–93.6%; 45.9–55.0% relative reduction) and ED revisit rates were 5.4–6.5% (NPV 93.5–94.6%; 49.6–59.0% relative reduction). The composite adverse event rate in the predicted-safe group was 11.1–14.1%, compared with 23.0–27.2% in the overall cohorts — a 48–53% relative reduction.

Conclusions

CURED, the CU Risk Estimator for Discharge, is an internally validated machine learning decision support tool that identifies inpatients at substantially reduced risk for 30-day adverse outcomes as candidates

for EPD during hospital surge. These findings support further external validation and prospective clinical study in surge exercise settings.

Key Points

Question: Can a machine learning tool reliably identify hospitalized patients at sufficiently low risk for 30-day adverse outcomes to support safe early discharge during hospital surge under crisis standards of care?

Findings: In this multi-site retrospective cohort study of 182,456 encounters across 12 hospitals, CURED—an internally validated ensemble of 42 machine learning models—classified 20.7–24.7% of patients as safe for early discharge, achieving 30-day mortality negative predictive values of 99.1–99.7% and reducing composite adverse outcome rates by approximately 48–53% relative to the overall cohort.

Meaning: CURED demonstrates the feasibility of data-driven early patient discharge decision support during large-scale casualty surge, warranting external validation and prospective evaluation in surge exercise settings.

Importance

In fiscal year 2020, the United States (US) Congress tasked the Department of Defense to evaluate interoperable partnerships, enhance interoperability, and expand capability and capacity of the military and civilian health systems that may be required to provide casualty care during a large-scale combat operation (LSCO) through the National Disaster Medical System (NDMS) Pilot Program. The return of up to 100,000 ill and injured military personnel to the US homeland over 100 days would saturate domestic hospital capacity—particularly burn and intensive care unit (ICU) beds—within two weeks, likely necessitating a declaration of a Public Health Emergency or invocation of the Stafford Act as acute care demand far exceeds available resources.[1–3] Unlike conventional mass casualty events, a LSCO scenario involves a sustained and geographically distributed surge across the civilian hospital system, with no single epicenter and no rapid resolution—demanding coordinated, data-driven strategies for managing inpatient throughput at scale.

Four broad strategies can expand hospital throughput under surge: extending care to non-traditional spaces and increasing staff capacity; using medical screening exams to redirect non-emergency patients away from the ED while directing borderline patients normally kept for observation to outpatient clinics for early follow-up; canceling elective admissions and surgeries; and facilitating early patient discharge (EPD), the deliberate discharge of current inpatients earlier than would otherwise occur, typically by hours to days.[4–6] EPD is the least operationalized of the four and is most closely linked to crisis standards of care (CSC) when hospital capacity is extremely constrained. Seminal work by Kelen and colleagues demonstrated that reverse triage (essentially EPD) of hospitalized adults at low risk for critical interventions could increase gross surge capacity by 77–103% across representative US hospital

types, with reverse triage accounting for the majority of reclaimed beds.[6] Simulation studies of acute mass casualty events suggest EPD could increase available bed space by 20–59% within 48 hours.[3] Despite this potential, few real-world surge events—including the SARS-CoV-2 pandemic—implemented systematic EPD programs, in part because no validated, scalable clinical decision support tool existed to identify candidates safely and efficiently.

Validated tools already in clinical use—including the National Early Warning Score 2 (NEWS2),[7] which aggregates physiologic parameters into a deterioration risk score, and the HOSPITAL score,[8,9] which predicts potentially avoidable 30-day readmission, may offer relevant signals for EPD. These instruments were designed for distinct purposes—NEWS2 to trigger rapid response escalation, HOSPITAL to guide post-discharge care transitions—and neither was developed or evaluated as a safe-discharge decision tool, and neither integrates functional status nor the multimodal electronic health record (EHR) data that characterizes the full trajectory of an inpatient stay. A machine learning approach that synthesizes these and additional predictors across multiple clinically relevant outcome domains—30-day mortality, readmission, and ED re-presentation—and across multiple pre-discharge time points could provide the comprehensive clinical decision support needed for EPD in a sustained medical surge environment.

We developed and internally validated CURED (the CU Risk Estimator for Discharge), a machine learning decision support tool built from 182,456 inpatient and observation encounters across 12 hospitals in the UCHealth system. CURED integrates demographics, comorbidity burden, vital signs, laboratory values, functional status, and clinical trajectory data to identify patients at sufficiently low risk across all three 30-day adverse outcome domains—mortality, readmission, and ED revisit—to be considered candidates for EPD under CSC. The model is designed to function as one component of a broader EPD strategy and is intended for use exclusively under CSC conditions, where the ethical and clinical framework for altered standards of care has been formally activated.

Methods

Setting and Participants

This study was a multi-site retrospective observational cohort study conducted across 12 hospitals belonging to a large Colorado hospital system. Adult inpatient or observation patients with known sex, aged 18 years or older, and not on an obstetric service, admitted on or after January 1, 2023, and discharged on or before December 31, 2024, were included in this analysis. Patients who were only ever cared for in the emergency department and never admitted to the hospital, patients who were discharged to hospice, and patients who died before discharge were excluded. The first eligible encounter within any 30-day interval for each patient was retained for analysis. Clinical data collected as part of patients' routine care were queried from the UCHealth Epic data warehouse and the Colorado Death Registry by Health Data Compass (healthdatacompass.org), a multi-institutional data warehouse funded by UCHealth, Children's Hospital Colorado, CU Medicine, and the University of Colorado School of

Medicine. This study was approved as exempt research by the Colorado Multiple Institution Review Board (COMIRB # 25–0131). Additional details regarding the methods are available in the Supplement.

Predictors

Candidate predictors, selected based on existing literature and inclusion in other validated tools like NEWS2 and the HOSPITAL score, were assembled from patient demographics, encounter characteristics, diagnoses, procedures, flowsheet data, and laboratory data with a focus on identifying parameters that could indicate medical illness or frailty. The following covariates were included: age at encounter, sex, self-reported race, primary language, insurance type, admission type, patient class, discharge hospital, discharge service category, ICU admission during encounter, stepdown unit admission during encounter, number of prior hospitalizations in the previous year, length of hospital stay in days, presence of any ICD-10-PCS procedure code, Elixhauser comorbidity index for mortality risk (ECI, mortality), Elixhauser comorbidity index for readmission (ECI, readmission), vital signs (systolic blood pressure, pulse, respiration rate, oxygen saturation, temperature, and supplemental oxygen status), the AM-PAC "6-Clicks" Inpatient Basic Mobility Short Form (AMPAC6), the Johns Hopkins Highest Level of Mobility scale (JH-HLM), Glasgow Coma Scale (GCS), and hemoglobin and sodium levels.[10–16]

Outcome variables

The primary outcome for this analysis was all-cause mortality from the Colorado Death Registry within 30 days of discharge with secondary outcomes of all-cause readmission to the hospital (either as observation or admission status) or ED re-presentation to a UCHealth hospital within 30 days of discharge. Thirty-day mortality was derived from discharge date and death date and coded as positive when death occurred within 30 days after discharge. A composite adverse outcome was defined as the occurrence of at least one of the following within 30 days of discharge: all-cause mortality, all-cause hospital readmission, or ED revisit.

Statistical methods

Daily Snapshot Framework. To simulate prospective clinical use, four daily snapshot datasets were constructed for each encounter. The index day (day 0) corresponded to the discharge date, and three preceding days (day − 1, day − 2, day − 3) represented 1, 2, and 3 calendar days before discharge, respectively. For each snapshot, time-varying predictors (length of stay, vital signs, laboratory values, functional status scores) reflected only information available up to and including that calendar day. Static predictors (demographics, comorbidity indices, admission characteristics) were identical across all four snapshots. Separate models were trained and evaluated for each snapshot day, producing four independent prediction time points per outcome. Because each snapshot day constitutes an independent test dataset, patient counts differ across rows in Table 3, and these numbers are not

additive across days. The day − 3, −2, and − 1 cohorts are smaller than the day 0 cohort because patients with shorter hospital stays do not contribute to earlier daily snapshots. Additionally, missingness in time-varying predictors decreases as snapshots approach the discharge date.

Missing Data Handling. For time-varying measures (respiration rate, hemoglobin, sodium, GCS, AMPAC6, and JH-HLM), values were first ordered by date within each patient encounter and then imputed using last observation carried forward (LOCF) within that encounter. This approach used only information available up to the current day and avoided incorporation of future measurements. For GCS specifically, if a patient had no GCS recorded at any point during the encounter, GCS was imputed as 15 (i.e., normal); otherwise, intermittent missing days were imputed using encounter-level LOCF.

After LOCF, residual missing predictor values were handled in a model-specific manner.

Predictor Screening. Predictors were specified dynamically for each outcome (30-day mortality, 30-day readmission, and 30-day ED visit) and each daily snapshot (day 0 through day − 3). Within each training dataset, candidate predictors were screened using univariate logistic regression with likelihood-ratio testing. Predictors with global likelihood-ratio p-values less than 0.20 were retained for multivariable model development. This screening process was repeated independently for each snapshot dataset so that the predictor set could vary by prediction day and by outcome.

Data Splitting. The cohort was randomly partitioned into a training set (90%) and an internal held-out test set (10%). To account for heterogeneity across hospitals and the low prevalence of the primary outcome (30-day mortality), the split was stratified by the mortality outcome variable within each hospital using a grouped splitting procedure. This approach ensured that class imbalance and hospital-specific case mixes were preserved in both partitions.

Model Training and Internal Validation. All model development steps, including data preprocessing, feature engineering, and hyperparameter tuning, were conducted exclusively within the training set. Hyperparameter tuning used 10-fold cross-validation, stratified by the target outcome. Model performance during tuning was guided primarily by precision-recall area under the curve (PR-AUC), with ROC-AUC retained as a secondary discrimination metric. The ANOVA racing procedure was then applied to efficiently explore the hyperparameter space by adaptively eliminating poorly performing configurations early during the search.

Four model classes (Ridge Logistic Regression (RLR), Random Forest (RF), Generalized Additive Model (GAM), and LightGBM) were trained for each combination of outcome and snapshot day, yielding a grid of models across 3 outcomes × 4 days × 4 model types.

Threshold Optimization. For each trained model, an outcome-specific classification threshold was derived to maximize the F1 score based on out-of-fold (OOF) predictions from 10-fold cross-validation. The F1-optimized threshold was then used to convert continuous predicted probabilities into binary class labels ("Yes" or "No") for each outcome. The final model was then refitted on the entire

development set using the best hyperparameter configuration, and the evaluation was performed on the 10% held-out test set. Validation metrics include PR-AUC, ROC-AUC, accuracy, F1, sensitivity, and specificity (Appendix Table 3).

Model Selection and Ensemble Decision Rule. Model selection proceeded in two stages. First, for each combination of outcome and snapshot day, the best-performing run per model type was identified by ranking on test-set PR-AUC (primary), test-set ROC-AUC (secondary), and training PR-AUC (tertiary). LightGBM models with degenerate F1-optimized thresholds (≤ 0.02) were excluded. The top N distinct model types were then retained per outcome and snapshot day. The final safe-discharge decision system aggregated predictions from all selected models across all three outcomes using an all-negative consensus rule. At deployment, a single current-day EHR snapshot was scored by all selected models trained at day 0, day − 1, day − 2, and day − 3. For a given patient on a given snapshot day, each selected model produced a binary prediction at its F1-optimized threshold. The patient was classified as safe for early discharge if and only if every model for every outcome predicted "No" (i.e., no adverse event). If any single model for any outcome predicted "Yes," the patient was flagged as potentially unsafe. This conservative, maximum-risk veto approach was designed to prioritize high sensitivity and negative predictive value for adverse outcomes, consistent with a clinical rule-out application where the cost of a false negative (discharging a patient who subsequently experiences an adverse event) substantially exceeds the cost of a false positive (retaining a patient who would have been safe). Final model performance was assessed exclusively on the independent 10% held-out test set, which was not exposed to the model during the training or tuning process.

Semantic matching was implemented in Python (version 3.11.9) using the transformers library. SAS version 9.4 (SAS Institute, Cary, NC) was used for Elixhauser comorbidity index calculation. All other analyses were implemented in R (version 4.5.0) using the tidymodels (version 1.4.1) framework (rsample, recipes, parsnip, tune, yardstick, workflows), with the bonsai, ranger, glmnet, and mgcv packages providing model engines.[17–23]

The TRIPOD (Transparent Reporting of a multivariable prediction model for

Individual Prognosis Or Diagnosis) + AI guidelines were applied to ensure transparent reporting of this analysis.[24]

Results

Across 12 UCHealth hospitals, 139,166 adult patients, accounting for 182,456 inpatient or observation encounters made up the data set (Appendix Fig. 1). Demographics, clinical characteristics, and outcomes, including 30-day mortality, 30-day readmission, and 30-day ED revisit, for the entire cohort are provided in Tables 1 and 2. Comparisons of demographics, clinical characteristics, and outcomes between the development (90% training) and evaluation (10% test) by daily snapshot are provided in Appendix Table 1.

We first evaluated whether NEWS2 and the HOSPITAL score could be used as primary predictors for early patient discharge, adjusting for demographics, encounter characteristics, ECI and functional status scores (Supplement, Appendix Table 2). Performance was modest across all three outcomes, motivating development of a broader predictor set. Following evaluation of published risk scores, we screened candidate predictors for inclusion in multivariable model development using univariate logistic regression with a global likelihood-ratio test (LRT) (Supplement, Appendix Table 3). Discrimination and classification metrics for all 42 constituent models, each evaluated on the held out 10% test set at its corresponding training snapshot day, are reported in Appendix Table 4. The optimal hyperparameters and cross-validation precision-recall AUC (PR-AUC) for all 42 models are provided in Appendix Table 5. Applying the aggregated all-negative safe discharge rule with outcome- and model-specific F1-optimized thresholds across four daily snapshot test datasets, the model classified 20.7% to 24.7% of patients as safe for discharge (Table 3).

Among patients predicted safe, the observed 30-day mortality rate ranged from 0.3% to 0.9%, with corresponding negative predictive values (NPVs) of 99.1% to 99.7%. Relative to the overall test cohorts, this represented an approximate 62.9% to 83.3% reduction in mortality risk. The predicted-safe group also had lower 30-day readmission rates (6.4% to 9.4%; NPV 90.6% to 93.6%), corresponding to an approximate 45.9% to 55.0% relative reduction compared with the overall cohort, and lower 30-day emergency department visit rates (5.4% to 6.5%; NPV 93.5% to 94.6%), corresponding to an approximate 49.6% to 59.0% relative reduction. For the composite adverse outcome, the event rate among patients predicted safe was 11.1% to 14.1%, compared with 23.0% to 27.2% in the overall test cohorts, representing an approximate 48% to 53% relative reduction in adverse-event prevalence within the group identified as safe for discharge.

Table 1
Patient-level characteristics

Characteristic	**N = 139,166[1]**
Age at encounter /years	64.0 (47.0, 75.0)
Sex	
Male	69,616 (50%)
Female	69,550 (50%)
Race	
White	107,880 (78%)
Other	16,834 (12%)
Black or African American	8,401 (6.0%)
Asian	2,503 (1.8%)
More Than One Race	2,174 (1.6%)
American Indian or Alaska Native	993 (0.7%)
Native Hawaiian or Other Pacific Islander	381 (0.3%)
Ethnicity	
Non-Hispanic	115,392 (83%)
Hispanic	21,860 (16%)
Unknown	1,914 (1.4%)
Primary language	
English	130,774 (94%)
Not English	8,392 (6.0%)
[1] n (%); Median (Q1, Q3)	

Table 2
Encounter-level characteristics and 30-day outcomes

Characteristic	**N = 182,456[1]**
Patient class	
Inpatient	139,361 (76%)
Observation	43,095 (24%)
Insurance	
Medicare	93,151 (51%)
Commercial	39,880 (22%)
Medicaid	32,259 (18%)
Other	9,613 (5.3%)
Self-Pay	7,553 (4.1%)
ICU admission	34,418 (19%)
Stepdown unit admission	27,709 (15%)
Admission type	
Emergent	153,539 (84%)
Elective	28,917 (16%)
Discharge hospital service	
Metro Denver	67,902 (37.2%)
Colorado Springs / Southern Colorado	50,350 (27.6%)
Northern Colorado	59,338 (32.5%)
Southern Colorado / Pueblo	4,866 (2.7%)
Service category	
Medicine	128,267 (70%)
Surgery	42,394 (23%)
Orthopedics	7,861 (4.3%)
Oncology	3,934 (2.2%)
Number of hospitalizations (prior year)	0.0 (0.0, 1.0)

[1] *n (%); Median (Q1, Q3)*

Characteristic	N = 182,456[1]
Patient class	
Length of stay /days	3.0 (2.0, 6.0)
Length of stay /hours	69.0 (40.0, 123.0)
Procedure (any)	95,213 (52%)
ECI, readmission	5.0 (0.0, 12.0)
(Missing)	9
ECI, mortality	0.0 (-5.0, 11.0)
(Missing)	9
Mortality within 30 days	3,361 (1.8%)
Readmission within 30 days	24,878 (14%)
ED visit within 30 days	23,416 (13%)
[1] n (%); Median (Q1, Q3)	

Table 3
CURED Model Performance: Safe Early Discharge Outcomes

Sample Size			
Snapshot Day	**Test Cohort (n)**	**Predicted Safe (n)**	**Predicted Safe (%)**
Day 0 (day of discharge)	18,252	3,862	21.2%
Day − 1	16,207	3,996	24.7%
Day − 2	11,995	2,669	22.3%
Day − 3	8,641	1,792	20.7%
30-Day Mortality			
Snapshot Day	**Test Prevalence**	**Predicted-Safe Prevalence**	**NPV / Relative Reduction**
Day 0 (day of discharge)	0.5%	99.5%	75.1%
Day − 1	0.3%	99.7%	83.3%
Day − 2	0.7%	99.3%	69.0%
Day − 3	1.0%	99.1%	62.9%
30-Day Hospital Readmission			
Snapshot Day	**Test Prevalence**	**Predicted-Safe Prevalence**	**NPV / Relative Reduction**
Day 0 (day of discharge)	6.6%	93.4%	51.1%
Day − 1	6.4%	93.6%	55.0%
Day − 2	7.3%	92.7%	55.0%
Day − 3	9.4%	90.6%	45.9%
30-Day Emergency Department Visit			
Snapshot Day	**Test Prevalence**	**Predicted-Safe Prevalence**	**NPV / Relative Reduction**
Day 0 (day of discharge)	6.5%	93.5%	49.6%
Day − 1	5.6%	94.4%	55.6%
Day − 2	5.7%	94.3%	56.9%
Day − 3	5.4%	94.6%	59.0%

Sample Size			
Snapshot Day	**Test Cohort (n)**	**Predicted Safe (n)**	**Predicted Safe (%)**
Composite Adverse Outcome			
Snapshot Day	**Test Prevalence**	**Predicted-Safe Prevalence**	**NPV / Relative Reduction**
Day 0 (discharge date)	12.1%	87.9%	47.2%
Day − 1	11.1%	88.9%	53.0%
Day − 2	12.3%	87.7%	52.6%
Day − 3	14.1%	85.9%	48.3%

To characterize how the number of models in the consensus rule influenced the safety-coverage tradeoff, we evaluated four nested configurations of the all-negative rule: top 1 (12 models), top 2 (24 models), top 3 (36 models), and the full system of 42 models (Fig. 2). A random discharge baseline, in which all patients are discharged regardless of predicted risk, was included as a reference.

Under random discharge, the composite adverse outcome rate equaled the overall test-cohort prevalence of 23.0% to 27.2% across the four snapshot datasets. As the number of models in the consensus rule increased from 12 (top 1) through 24, 36, and 42, coverage narrowed and the composite adverse outcome rate in the predicted-safe group fell progressively, reaching 11.1% to 14.1% under the full 42-model system — a 47.2% to 53.0% relative reduction (Table 3; Fig. 2).

The largest incremental safety gains occurred between the top-1 and top-2 tiers for mortality, and between the top-2 and top-3 tiers for readmission, ED revisit, and the composite outcome. Mortality discrimination was largely established by the top-2 tier with diminishing marginal returns thereafter, whereas readmission and ED revisit discrimination improved more gradually across all tiers (Fig. 3).

Under the full 42-model system, the predicted-safe group demonstrated consistent reductions in all three adverse outcomes relative to the overall cohort across all four snapshot days (Fig. 1). Observed 30-day mortality among predicted-safe patients was 0.3% to 0.9% compared with 1.8% to 2.6% in the overall cohort (NPV 99.1% to 99.7%). Thirty-day readmission rates were 6.4% to 9.4% versus 13.5% to 17.3% (NPV 90.6% to 93.6%), and 30-day ED revisit rates were 5.4% to 6.5% versus 12.7% to 13.3% (NPV 93.5% to 94.6%). These reductions were observed at all snapshot time points, with day 0 and day − 1 models generally achieving the lowest adverse event rates in the predicted-safe group.

Discussion

A large-scale combat operation returning 100,000 casualties to the US homeland over 100 days would rapidly exhaust domestic hospital capacity, with modeling suggesting that burn and intensive care unit

beds could reach saturation within two weeks.[1] Early patient discharge is among the most actionable throughput strategies available to hospital systems during surge.

Despite longstanding recognition that EPD is a viable surge strategy, the field has lacked a rigorous, data-driven tool for operationalizing it safely. Prior approaches have relied primarily on clinical judgment or coarse scoring systems.[6] CURED advances this approach by integrating clinical scores (NEWS2 and HOSPITAL) alongside functional status measures, laboratory indices, and detailed EHR phenotyping within an ensemble machine learning framework, training 42 models across four model classes, three outcomes, and four pre-discharge time points. The result is a system optimized to rule out adverse outcomes by applying a conservative all-negative consensus rule that classifies a patient as safe for EPD only if every constituent model across all three outcomes predicts no adverse event.

The decision to evaluate the full 42-model consensus rule rather than a smaller subset reflects an asymmetric view of the safety-coverage tradeoff. On the safety axis, there is no ceiling: the lower the adverse event rate among patients recommended for EPD, the stronger the clinical and ethical justification for acting on the recommendation. On the coverage axis, however, only a limited proportion of the discharge-eligible population needs to be flagged in any given operational scenario.

The performance of CURED in the held-out test cohort is clinically meaningful in the context of EPD. Among the 20.7–24.7% of patients classified as safe for EPD across snapshot days, 30-day mortality was 0.3–1.0%, representing a 62.9–83.3% relative reduction compared with the overall test cohort prevalence of 1.8–2.6%. The negative predictive values (NPV) for 30-day mortality (99.1–99.7%) are consistent with those reported in analogous ML-based discharge safety studies and support CURED's application as a rule-out tool in which minimizing missed adverse events is the primary safety criterion.[25]

Readmission and emergency department revisit rates in the predicted-safe group were similarly reduced by approximately 46–55% and 50–59%, respectively. Crucially, the composite adverse outcome rate among patients flagged as safe (11.1–14.1%) was roughly half that of the overall cohort (23.0–27.2%), meaning that CURED concentrates lower-risk patients for discharge while directing clinical attention toward those who should remain hospitalized. This kind of stratification, not a binary safe/unsafe designation but a probabilistic enrichment of a lower-risk subgroup, is precisely what a CSC decision support tool should provide.

The daily snapshot framework, modeling outcomes at discharge (day 0) and up to three days prior to actual discharge (days − 1 through − 3), is a design strength with direct operational implications. In a surge environment, clinicians will not wait until the morning of planned discharge to begin identifying EPD candidates; the ability to flag patients up to 24–72 hours in advance enables parallel workstreams for care coordination, family notification, and post-discharge resource arrangement. The consistency of CURED’s NPVs across snapshot days suggests the model is stable over this window and does not substantially degrade with earlier prediction. The model’s constituent components also span multiple domains: patient functional status (AM-PAC 6-Clicks, JH-HLM), physiologic stability (NEWS2

components, GCS), laboratory markers (hemoglobin, sodium), and comorbidity burden (ECI scores), reflecting the multidimensional nature of discharge readiness in a diverse inpatient population. This integration across data types advances beyond prior single-score approaches and aligns with the broader machine learning literature demonstrating superior discrimination when multimodal EHR data are combined.

Several important caveats should frame how CURED is interpreted and deployed. The model is designed as a clinical decision support tool to be used within a declared CSC, functioning as one component of a broader EPD strategy that necessarily includes clinician judgment, social circumstances assessment, and post-discharge resource availability. Patients flagged as safe for EPD by CURED are not unconditionally low-risk: a composite adverse outcome rate of 11–14% in this group reflects meaningful residual risk, particularly relevant if community-level support infrastructure is strained during a concurrent disaster. A prospective pilot study embedding CURED into clinical workflow during a planned surge exercise would be the logical next step, allowing real-time evaluation of its operational feasibility, clinician adoption, and impact on throughput.

As a development-stage tool, CURED has not yet been evaluated for clinician usability, workflow integration, or user trust; these assessments are appropriately reserved for prospective surge exercise settings where operational feasibility can be evaluated alongside model performance.

Limitations

Several limitations warrant consideration. First, CURED was developed and validated within a large, multi-site academic health system with 182,456 encounters across 12 hospitals in Colorado. Despite this geographic distribution, the cohort represents a particular patient mix, care culture, and level of EHR data completeness that may not generalize to other systems. Readmission and ED revisit outcomes were ascertained only within the UCHealth network and the Colorado Death Registry; adverse events occurring at outside facilities were not captured, likely underestimating true event rates and potentially inflating model NPVs. Second, the model has not been validated prospectively. All predictions reflect patterns learned from historical discharge decisions made under routine clinical conditions, not under CSC. Third, the daily snapshot framework captures data at fixed pre-discharge intervals (day 0 through day − 3), which may not reflect the full dynamic trajectory of every patient's hospitalization. Fourth, CURED does not incorporate social determinants of health, including access to medications, home health services, caregiver availability, and transportation, all of which are integral to the safety of EPD in practice and which may be particularly compromised during disaster conditions. A composite risk score that integrates clinical and social risk factors would more fully characterize EPD candidacy. Fifth, missing data required model-specific imputation strategies, including last-observation-carried-forward and complete-case analysis; if patients with missing data differ systematically from those with complete records, estimates may be biased. Finally, EPD under CSC raises distinct ethical considerations, including equity of discharge decisions across demographic groups, that prospective studies must

address explicitly and that future iterations of CURED should evaluate through fairness analyses stratified by race, ethnicity, insurance status, and language preference.

Conclusions and Relevance

CURED is an internally validated, multi-outcome machine learning clinical decision support tool designed to identify inpatients as candidates for early discharge during hospital surge under crisis standards of care. Trained on 182,456 encounters across 12 hospitals and evaluated on an independent 10% holdout, the model classified 20.7–24.7% of patients as safe for EPD—a group in whom 30-day mortality NPVs exceeded 99% and the composite adverse outcome rate was reduced by approximately 48–53% relative to the overall cohort. The all-negative consensus architecture, which requires every constituent model across all three outcome domains to predict no adverse events before flagging a patient as safe, is deliberately conservative and reflects the asymmetric cost of a false negative in a CSC discharge decision. Applied as one component of a comprehensive EPD strategy, alongside clinician judgment, social circumstances assessment, and active post-discharge follow-up, CURED has the potential to support meaningful increases in hospital throughput during large-scale casualty surge events. External validation across diverse health systems, prospective evaluation embedded in surge exercises, and expansion of the predictor set to include social determinants of health are the necessary next steps toward clinical deployment.

Abbreviations

AM-PAC 6-Clicks	Activity Measure for Post-Acute Care “6-Clicks” Inpatient Basic Mobility Short Form
ANOVA	Analysis of variance
CSC	Crisis standards of care
CURED	CU (University of Colorado) Risk Estimator for Discharge
ECI	Elixhauser comorbidity index
ED	Emergency department
EHR	Electronic health record
EPD	Early patient discharge
F1	Harmonic mean of precision and recall (F1 score)
GAM	Generalized additive model
GCS	Glasgow Coma Scale
HOSPITAL score	Hemoglobin, discharge from an Oncology service, Sodium level, Procedure during index admission, Index admission Type, Any Admissions in the previous year, Length of stay
ICD-10-CM	International Classification of Diseases, Tenth Revision, Clinical Modification
ICD-10-PCS	International Classification of Diseases, Tenth Revision, Procedure Coding System
ICU	Intensive care unit
JH-HLM	Johns Hopkins Highest Level of Mobility scale
LightGBM	Light Gradient-Boosting Machine
LOCF	Last observation carried forward
LSCO	Large-scale combat operation
LRT	Likelihood-ratio test
ML	Machine learning
NDMS	National Disaster Medical System
NEWS2	National Early Warning Score 2
NPV	Negative predictive value
OOF	Out-of-fold (cross-validation predictions)
PR-AUC	Precision-recall area under the curve
REML	Restricted maximum likelihood

RF	Random forest
RLR	Ridge logistic regression
ROC-AUC	Receiver operating characteristic area under the curve
SpO2	Peripheral oxygen saturation
TRIPOD+AI	Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis, with extension for artificial intelligence

Declarations

This study was approved as exempt research by the Colorado Multiple Institution Review Board (COMIRB #25-0131).

Competing interests: No authors had any financial or non-financial competing interests to report.

Disclosures: Dr. Keniston reports funding from the Agency of Healthcare Research and Quality, the American Medical Association, and the National Institute for Occupational Safety and Health, unrelated to this work. Dr. Keniston contributed to the development of GrittyWork, a digital workforce application and trademark of the University of Colorado. The opinions and assertions expressed herein are those of the author(s) and do not reflect the official policy or position of the Uniformed Services University of the Health Sciences or the Department of Defense. The contents and opinions in this presentation are the sole responsibility of the author(s) and do not necessarily reflect the views, opinions or policies of The Henry M. Jackson Foundation for the Advancement of Military Medicine, Inc.

AI Use Disclosure: The authors used Claude Sonnet 4.6 (Anthropic, model version claude-sonnet-4-6) to assist with editing and revising original author content for clarity, logical structure, and readability. Authors take full responsibility for the integrity of all content.

Role of the Funding Source: Funded by the Uniformed Services University of the Health Sciences Award to the Henry M. Jackson Foundation for the Advancement of Military Medicine, Inc. Agreement HU00012420023, subcontract 390.01. There were no restrictions placed on the program's evaluation.

Protocol Registration and Data Availability: This study was not prospectively registered, as it was a retrospective observational analysis conducted under an IRB-approved exempt protocol (COMIRB #25-0131). The study protocol has not been deposited in a public repository. Individual patient-level data cannot be shared due to data use agreement restrictions with UCHealth and the sponsoring agency. The analytic code and trained model may be made available to qualified investigators upon reasonable request, subject to approval by the project team and sponsor.

Author Contributions: J.P., G.S., Y.Z., and A.K. conceptualized and designed the study. G.S. and Y.Z. developed and implemented the analytical code and machine learning models. J.P. and A.K. secured funding and supervised the project. J.P. drafted the original manuscript. J.R., P.S., K.Y., K.R-L., J.D.F., and

C.L. contributed to interpretation of findings in the context of clinical and operational surge medicine and participated in critical review and revision of the manuscript. All authors have approved the submitted version and agree to be accountable for their own contributions and to ensure that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Primary Funding Source

The Uniformed Services University of the Health Sciences (USU), 4301 Jones Bridge Rd., A1040C, Bethesda, MD 20814-4799 is the awarding and administering office under award HU00012420023, subcontract 390.01, unrestricted.

Data Availability

The datasets generated and analyzed during this study are not publicly available due to data use agreement restrictions with UCHealth and the sponsoring agency but are available from the corresponding author on reasonable request and with permission of UCHealth and the sponsor.

Code Availability

The underlying code for this study is not publicly available but may be made available to qualified researchers on reasonable request from the corresponding author, subject to approval by the project team and sponsor.

References

1. Weber E, Buckler D, Petrozzo K, et al. Financial Impacts of Receiving Combat Casualties during a Large-Scale Combat Operation on Civilian Hospitals in the NDMS Pilot Study. *Disaster medicine and public health preparedness.* 2026;20.
2. Section 319 of the Public Health Service Act, U.S.C., § 247d (2024).
3. Robert T. Stafford Disaster Relief and Emergency Assistance Act, U.S.C., §§§ 5121–5207 (2024).
4. Eraso D, Wright B. Surge Planning. In: Shiver J, ed. *Critical Care of COVID-19 in the Emergency Department*: Springer Cham; 2021.
5. Jacobs-Wingo JL, Cook HA, Lang WH. Rapid Patient Discharge Contribution to Bed Surge Capacity During a Mass Casualty Incident: Findings From an Exercise With New York City Hospitals. *Qual Manag Health Care.* 2018;27(1):24–29.
6. Kelen G, McCarthy M, Kraus C, et al. Creation of surge capacity by early discharge of hospitalized patients at low risk for untoward events. *Disaster Med Public Health Prep.* 2009;3(2 Suppl):S10–16.
7. National Early Warning Score (NEWS) 2. *RCP Quality Improvement and Patient Safety* https://www.rcplondon.ac.uk/projects/outputs/national-early-warning-score-news-2. Accessed December 22, 2021.

8. Donze J, Aujesky D, Williams D, Schnipper JL. Potentially avoidable 30-day hospital readmissions in medical patients: derivation and validation of a prediction model. *JAMA Intern Med.* 2013;173(8):632–638.
9. Burke RE, Schnipper JL, Williams MV, et al. The HOSPITAL Score Predicts Potentially Preventable 30-Day Readmissions in Conditions Targeted by the Hospital Readmissions Reduction Program. *Med Care.* 2017;55(3):285–290.
10. Cummings-Vaughn LA, Chavakula NN, Malmstrom TK, Tumosa N, Morley JE, Cruz-Oliver DM. Veterans Affairs Saint Louis University Mental Status examination compared with the Montreal Cognitive Assessment and the Short Test of Mental Status. *J Am Geriatr Soc.* 2014;62(7):1341–1346.
11. Jette DU, Stilphen M, Ranganathan VK, Passek SD, Frost FS, Jette AM. AM-PAC "6-Clicks" functional assessment scores predict acute care hospital discharge destination. *Phys Ther.* 2014;94(9):1252–1261.
12. Warren M, Knecht J, Verheijde J, Tompkins J. Association of AM-PAC "6-Clicks" Basic Mobility and Daily Activity Scores With Discharge Destination. *Phys Ther.* 2021;101(4).
13. Hoyer EH, Young DL, Klein LM, et al. Toward a Common Language for Measuring Patient Mobility in the Hospital: Reliability and Construct Validity of Interprofessional Mobility Measures. *Phys Ther.* 2018;98(2):133–142.
14. Elixhauser A, Steiner C, Harris D, Coffey R. Comorbidity measures for use with administrative data. *Medical care.* 1998;36(1).
15. Elixhauser Comorbidity Software Refined for ICD-10-CM. https://hcup-us.ahrq.gov/toolssoftware/comorbidityicd10/comorbidity_icd10.jsp. Accessed April 4, 2026.
16. Teasdale G, Jennett B. Assessment of coma and impaired consciousness. A practical scale. *Lancet.* 1974;2(7872):81–84.
17. Kuhn M, Wickham H. Tidymodels: a collection of packages for modeling and machine learning using tidyverse principles. 2020; https://www.tidymodels.org/. Accessed April 21, 2026.
18. Friedman J, Hastie T, Tibshirani R. Regularization Paths for Generalized Linear Models via Coordinate Descent. *J Stat Softw.* 2010;33(1):1–22.
19. Wright M, Ziegler A. ranger: A fast implementation of random forests for high dimensional data in C + + and R. *J Stat Softw.* 2017;77:1–17.
20. Wood S. *Generalized Additive Models: An Introduction with R*. 2nd ed: Chapman and Hall/CRC; 2017.
21. Wolf T, Debut L, Sanh V, et al. Transformers: State-of-the-Art Natural Language Processing. Paper presented at: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations2026.
22. Ke G, Meng Q, Finley T, et al. Lightgbm: A highly efficient gradient boosting decision tree. *Advances in neural information processing systems.* 2017;30.

23. Falbel D, Damiani A, Hogervorst R, Kuhn M, Couch S, Hvitfeldt E. bonsai: Model Wrappers for Tree-Based Models. 2025; R package version 0.4.0:https://bonsai.tidymodels.org/. Accessed April 22, 2026.
24. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): the TRIPOD statement. *Ann Intern Med.* 2015;162(1):55–63.
25. Tschoellitsch T, Maletzky A, Moser P, et al. Machine learning prediction of unexpected readmission or death after discharge from intensive care: A retrospective cohort study. *J Clin Anesth.* 2024;99(111654).

Figures

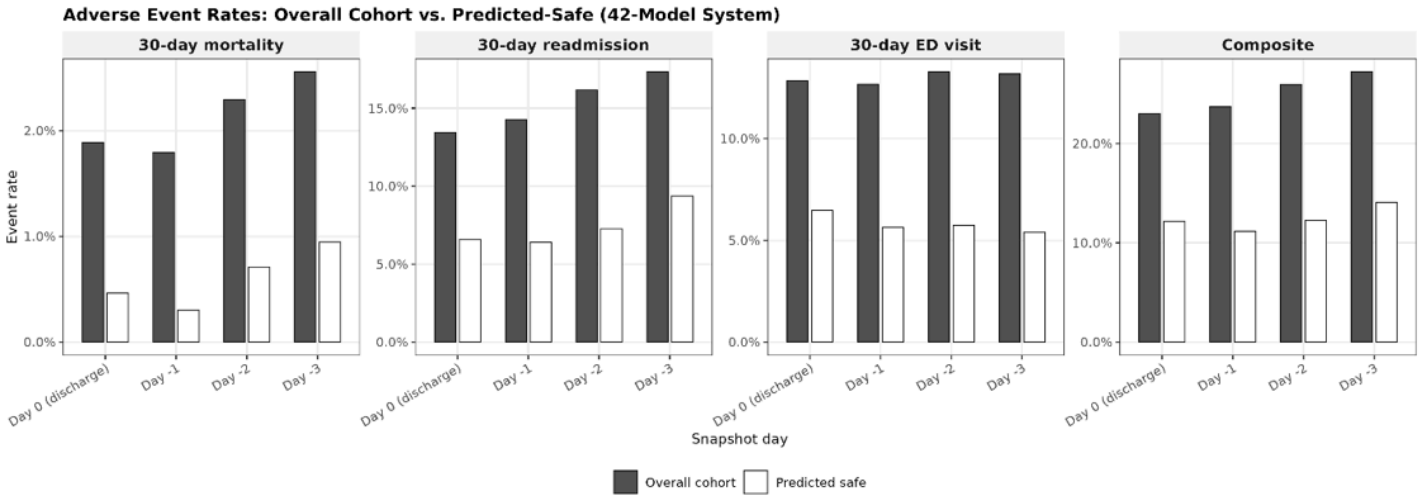

Figure 1

Adverse event rates in the overall test cohort versus the predicted-safe group under the 42-model all-negative consensus rule. Grouped bars compare the observed 30-day adverse event rates in the full test cohort (dark bars) with the rates among patients classified as safe for early patient discharge by the complete 42-model all-negative consensus rule (white bars). Panels display 30-day mortality, 30-day hospital readmission, 30-day emergency department revisit, and the composite adverse outcome. Results are shown for each of the four daily snapshot test datasets (day 0 through day -3). Error bars are omitted because all rates are computed on the full held-out test set without resampling.

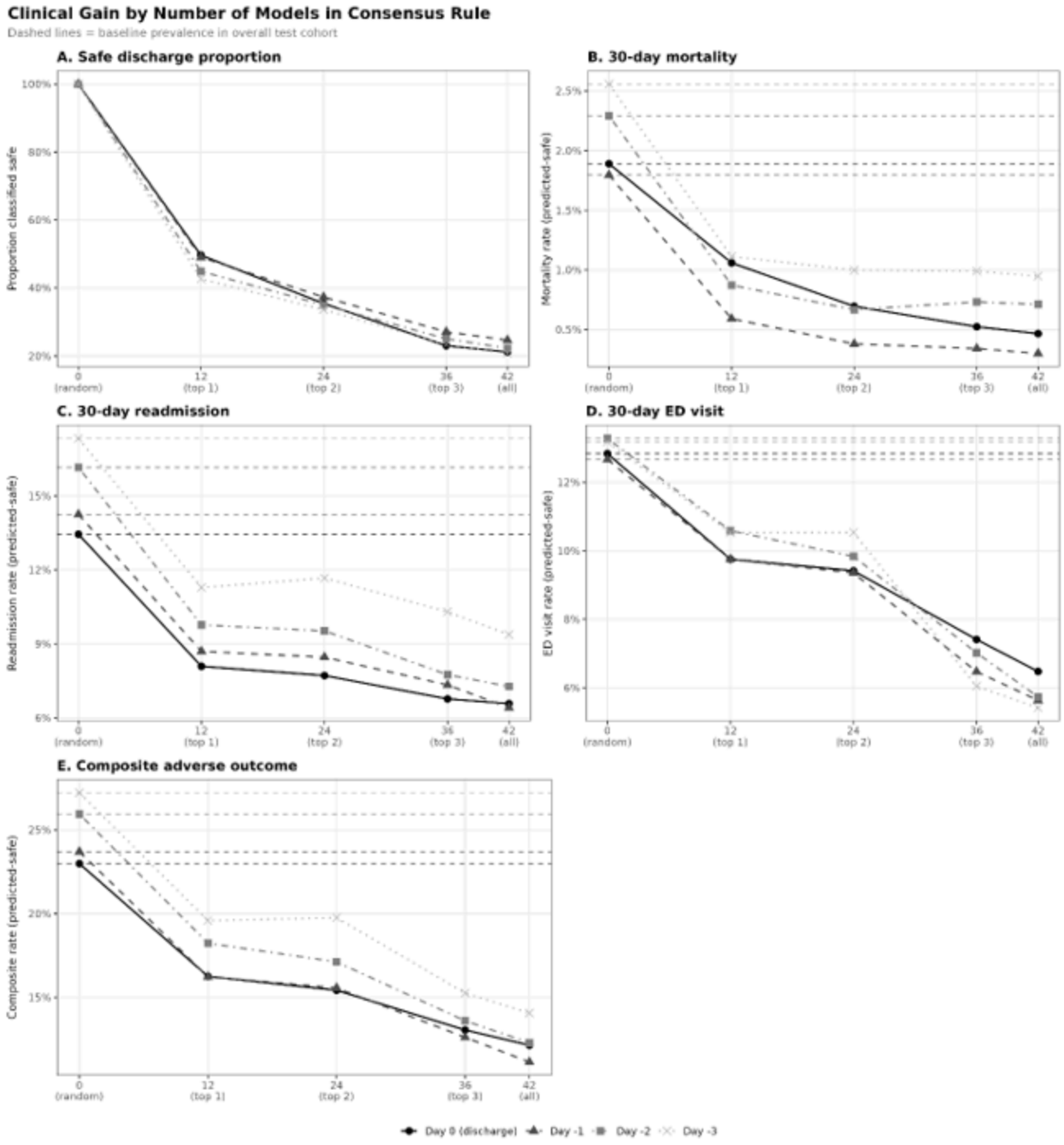

Figure 2

Clinical gain by number of models in the consensus rule. Panel A shows the proportion of test-set patients classified as safe for early patient discharge (EPD) as the number of models in the all-negative consensus rule increases from 0 (random discharge, equivalent to discharging all patients) through 12 (top 1 model per outcome-day combination), 24 (top 2), 36 (top 3), and 42 (all selected models). Panels B through E show the corresponding adverse event rates among the predicted-safe group for 30-day mortality (B), 30-day readmission (C), 30-day emergency department revisit (D), and the composite adverse outcome (E). Each line represents one of four daily snapshot test datasets: discharge day (day 0) and 1, 2, and 3 days prior. Dashed horizontal lines indicate the baseline prevalence for each outcome

in the overall test cohort. As additional models are incorporated into the consensus rule, both the proportion of patients classified as safe and the adverse event rate within the predicted-safe group decrease, reflecting the tradeoff between coverage and safety.

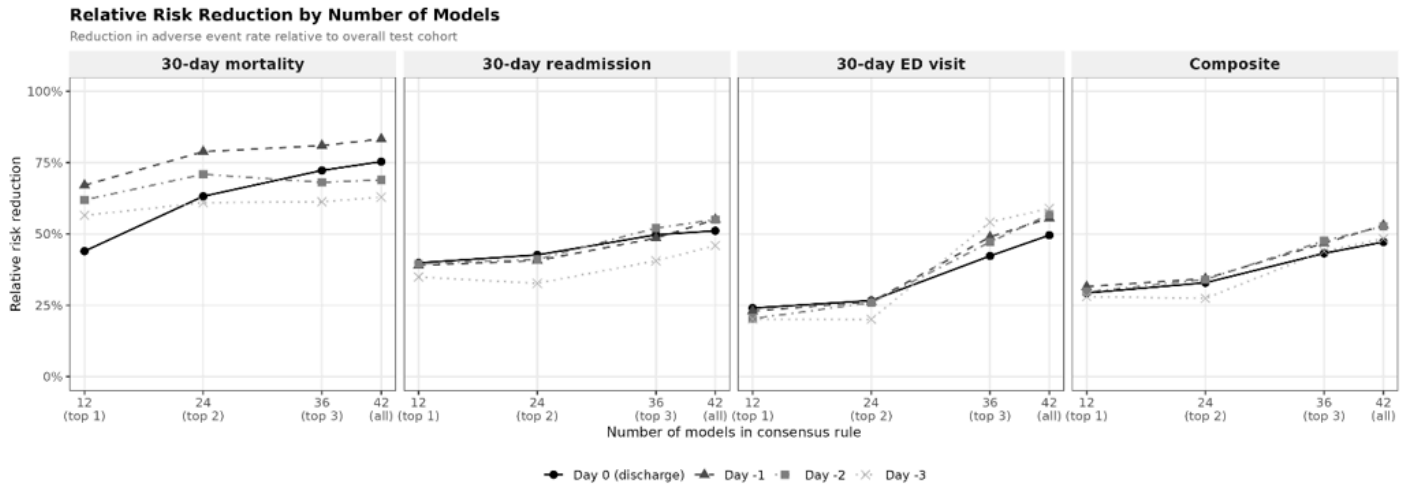

Figure 3

Relative risk reduction by number of models in the consensus rule. Each panel shows the relative reduction in adverse event rate among patients classified as safe for early patient discharge compared with the overall test cohort, plotted against the number of models in the all-negative consensus rule (12, 24, 36, and 42). Panels display 30-day mortality, 30-day readmission, 30-day emergency department revisit, and the composite adverse outcome. Each line represents one daily snapshot test dataset (day 0 through day -3). For 30-day mortality, relative risk reduction increased steeply from the top 1 tier to the top 2 tier and continued to rise incrementally through the full 42-model system. Gains for readmission, ED revisit, and the composite outcome were more gradual at lower tiers, with the largest incremental improvements occurring between the top 2 (24 models) and top 3 (36 models) configurations.

Supplementary Files

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

- Supplement.docx