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.

Furosemide initiation during norepinephrine infusion and proximal hemodynamic destabilization: a target trial emulation in MIMIC-IV with external validation in eICU-CRD

Lujun Shao
The Affiliated Taizhou People's Hospital of Nanjing Medical University
Cheng Xu
The Affiliated Taizhou People's Hospital of Nanjing Medical University
Huiyu Tai
The Affiliated Taizhou People's Hospital of Nanjing Medical University
Yue Li
nicolasphelpsly@163.com

The Affiliated Taizhou People's Hospital of Nanjing Medical University

Research Article

Keywords: furosemide, norepinephrine, target trial emulation, causal inference, heterogeneous treatment effects, hemodynamic instability, acute kidney injury, critical care

Posted Date: July 2nd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9845399/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.

Intravenous furosemide is frequently administered to intensive care unit (ICU) patients receiving norepinephrine, but the proximal hemodynamic consequences of initiating diuresis during vasopressor exposure remain inadequately characterized. Prior work has focused on efficacy and mortality endpoints.

Methods.

We conducted a target trial emulation in the Medical Information Mart for Intensive Care IV (MIMIC-IV) database (derivation, 2008–2022) and replicated the analysis in the eICU Collaborative Research Database (eICU-CRD; validation, 2014–2015). Adults receiving ≥ 120 minutes of norepinephrine were eligible. Time zero was the first furosemide dose during active norepinephrine (treated) or a sampled norepinephrine-on hour (control). The primary 24-hour composite outcome was sustained norepinephrine dose escalation (≥ 1.3 times the index dose) or new vasopressor initiation. Analyses combined 1:2 propensity-score matching, exact-plus-propensity sensitivity matching, stabilized inverse probability weighting, causal forest, depth-2 policy tree, and E-value sensitivity analyses.

Results.

The MIMIC-IV cohort comprised 6,759 patients (1,667 treated, 5,092 control); 3,836 were matched. Furosemide was associated with the primary composite (odds ratio [OR] 2.03, 95% confidence interval [CI] 1.77 to 2.34), driven by norepinephrine dose escalation (OR 2.66, 2.27 to 3.12; E-value 3.75) but not new vasopressor initiation (OR 1.15). The doubly robust average treatment effect was an absolute risk difference of 0.161 (0.135 to 0.188), corresponding to a number needed to harm of approximately 6; all 6,759 patient-level causal-forest estimates were positive, and the depth-2 policy tree identified no benefit-favorable subgroup. Acute kidney injury (AKI) was null in the primary analysis (OR 1.03) but positive on exact-plus-propensity sensitivity (OR 1.25, 1.10 to 1.43); 28-day mortality showed no association (OR 1.05). eICU-CRD validation (n = 7,137; 2,307 matched) directionally confirmed the primary findings: primary composite OR 1.71 (1.40 to 2.09), AKI 2.01 (1.65 to 2.44), in-hospital mortality null (1.10).

Conclusions.

In two independent intensive care databases, initiation of furosemide during active norepinephrine infusion was associated with proximal hemodynamic destabilization without a detectable short-term mortality signal. No patient subgroup was predicted to derive net hemodynamic benefit: causal-forest

estimates were positive for all 6,759 patients, and the depth-2 policy tree identified only relative-cost ordering. Findings support careful consideration of timing in vasopressor-dependent patients pending prospective evaluation.

Background

Hemodynamic instability requiring vasopressor support is a defining feature of critical illness, arising most commonly from distributive shock due to sepsis but also encompassing cardiogenic, postoperative vasoplegic, and hypovolemic states. Sepsis and septic shock are the most frequent precipitating conditions and remain among the leading causes of intensive care admission and death, with reported in-hospital mortality rates ranging from 25 to 40 percent depending on case mix and care era [1, 2]. Aggressive intravenous fluid resuscitation has been a historical cornerstone of early hemodynamic management across these shock states [3, 4], yet observational and interventional evidence accumulated over the past two decades has consistently linked cumulative positive fluid balance to adverse outcomes including increased mortality, prolonged ventilator dependence, and exacerbation of acute kidney injury [5–7]. Acute kidney injury (AKI) complicates up to 57 percent of intensive care admissions [8] and, when staged using the Kidney Disease: Improving Global Outcomes classification [9], carries independent associations with short- and long-term mortality and progression to chronic kidney disease. The recognition that fluid overload itself contributes to organ dysfunction has motivated a paradigm shift toward more conservative resuscitation strategies and, in stable patients, active deresuscitation [10–12].

Loop diuretics, principally furosemide, are the most frequently prescribed agents for fluid removal in critically ill patients and are administered to roughly half of all ICU admissions [13, 14]. Their natriuretic action at the Na-K-2Cl cotransporter of the thick ascending limb of the loop of Henle is well characterized pharmacologically [15], but their net clinical effect in critically ill patients remains contested. Observational studies have reported associations between diuretic use and both increased and decreased mortality after AKI [16–18]; randomized trials have shown that diuretics increase urine output without consistently improving renal recovery or survival [19, 20]; and meta-analyses have concluded that loop diuretics shorten the duration of renal replacement therapy but do not modify mortality [21]. A distinct but related clinical question — the hemodynamic safety of administering diuretics during ongoing vasopressor support — has received considerably less rigorous attention. Norepinephrine, the recommended first-line vasopressor for septic shock and a commonly used agent across other vasopressor-requiring states [4, 22, 23], maintains arterial pressure largely through alpha-adrenergic reduction of venous capacitance and consequent recruitment of preload; abrupt diminution of effective intravascular volume by concurrent diuresis may therefore compromise the dose–response coupling that sustains perfusion in this state.

The largest prior study to address this specific question — Bandak and colleagues' propensity-matched analysis of 939 Mayo Clinic ICU patients — compared mean hourly vasopressor dose in the six hours preceding and following furosemide administration and reported no significant difference, but did not

anchor the comparison in a target trial framework with rigorous time-zero alignment, did not employ an event-based outcome capable of detecting clinically meaningful escalation in an affected subset, and did not characterize treatment effect heterogeneity across patient phenotypes [24]. Recent ultrasound-based work has characterized the prevalence of systemic venous congestion in critically ill patients and examined its potential association with AKI, with results that have not consistently confirmed this association in general ICU cohorts, motivating broader clinical interest in whether systemic congestion may identify subgroups in whom decongestive therapy is differentially indicated [25]. Resolving these questions requires both a causally interpretable framework to address indication bias inherent to non-random treatment assignment [26] and methods capable of estimating and characterizing patient-level heterogeneity in treatment effect [27–29].

We therefore conducted a target trial emulation in the Medical Information Mart for Intensive Care IV (MIMIC-IV) database to estimate the average and conditional treatment effects of intravenous furosemide on proximal hemodynamic destabilization, on AKI progression, and on 28-day mortality among adults receiving active norepinephrine infusion [30]. We further employed a causal forest to characterize heterogeneity across baseline covariates and a depth-2 policy tree to translate that heterogeneity surface into an interpretable decision rule. To assess the generalizability of the primary findings beyond the single-center derivation cohort, we additionally conducted an external validation of the average treatment effect estimates in the eICU Collaborative Research Database (eICU-CRD), an independent multicenter critical care database encompassing 208 United States hospitals [31].

Methods

Study design and data source

We conducted a retrospective cohort study using the target trial emulation framework proposed by Hernán and Robins [26], in which the components of a hypothetical pragmatic randomized trial were prespecified before analysis to anchor the observational comparison and minimize common sources of bias in critical care database studies. The target trial protocol, specifying eligibility, treatment strategies, assignment procedure, start and end of follow-up, outcomes, and causal contrast of interest, is summarized in Supplementary Table S1.

Data were extracted from the MIMIC-IV (version 3.1) [30], a publicly available critical care database covering adult intensive care unit admissions at Beth Israel Deaconess Medical Center between 2008 and 2022. The database provides hour-resolution physiological monitoring data, structured laboratory results, and time-stamped medication administration records. This study was a secondary analysis of two publicly available, fully de-identified critical care databases (MIMIC-IV and the eICU-CRD) and was conducted in accordance with the Declaration of Helsinki. The establishment of MIMIC-IV was approved by the Institutional Review Boards of the Beth Israel Deaconess Medical Center (Boston, MA, USA) and the Massachusetts Institute of Technology (Cambridge, MA, USA), which granted a waiver of individual informed consent; the eICU-CRD is an Institutional Review Board–exempt de-identified database.

Because all records were de-identified, additional institutional review board approval and informed consent were not required for this analysis. The authors accessed both databases as credentialed users after completing the mandated human-subjects research training and signing the PhysioNet data use agreement. Reporting follows the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement [32] and its REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) extension [33] for studies using routinely collected health data.

Study population

Eligibility criteria for the emulated trial were age 18 years or older, one ICU stay per patient (selecting the chronologically first qualifying ICU stay per subject_id; each subject contributes exactly one observation to the analytic cohort, ensuring patient-level independence), ICU length of stay of at least 24 hours, and cumulative norepinephrine infusion duration of at least 120 minutes to ensure sustained vasopressor dependence rather than brief peri-operative use. Patients with end-stage kidney disease (identified by International Classification of Diseases (ICD)-9 codes 585.6, V45.11, V56.x and ICD-10 codes N18.6, Z99.2), those receiving renal replacement therapy within the first 24 hours of ICU admission, and those receiving intravenous furosemide before initiation of norepinephrine were excluded. The cohort assembly is summarized in Fig. 1.

Treatment strategies and time zero

Two treatment strategies were compared: (i) initiation of intravenous furosemide during an active norepinephrine infusion at any time during the ICU stay, versus (ii) no intravenous furosemide administration while on norepinephrine. Time zero, the start of follow-up, was set at the index time, defined as follows. For patients assigned to the furosemide strategy (treated arm), index time was the timestamp of the first intravenous furosemide dose recorded while within an active norepinephrine infusion window. For patients assigned to the no-furosemide strategy (control arm), an hourly grid of candidate index times spanning all on-norepinephrine ICU hours was generated, from which one candidate index time per patient was selected at random; the control arm includes patients who never received intravenous furosemide during any norepinephrine infusion episode, and the index time is therefore necessarily prior to any later furosemide exposure that occurred after norepinephrine had been discontinued (a small minority of patients). This eligibility–exposure–time-zero alignment ensures that baseline covariates are measured at the same clinical reference point in both arms and that assignment is not informed by post-time-zero data, addressing immortal time bias that arises when these elements are decoupled.

Norepinephrine infusion rates were harmonized to micrograms per kilogram per minute using patient weight from the derived first-day weight table. Records lacking unit information or patient weight could not be unit-normalized and were treated as missing for dose calculation. Full item identifier mapping is described in Supplementary Methods S1 and drug unit harmonization in Supplementary Methods S2; the

random sampling procedure for control index times is described above in the Treatment strategies and time zero subsection.

Outcomes

The primary outcome was a composite indicator of hemodynamic destabilization within 24 hours of time zero, defined as the occurrence of either (a) sustained norepinephrine dose escalation to at least 1.3 times the index dose lasting at least 120 minutes, or (b) initiation of a second vasoactive agent (vasopressin, epinephrine, phenylephrine, dopamine, or dobutamine) that was not active at time zero. These two components were prespecified as a single composite because each represents a clinically equivalent expression of failure of in-place hemodynamic support: in both cases, the current pressor regimen has proven insufficient to maintain perfusion targets, prompting either upward titration of the active agent or addition of a new agent. A third candidate component, sustained mean arterial pressure below 65 mmHg, was prespecified but reclassified to a sensitivity outcome on methodological grounds: inspection of the marginal (cohort-level, treatment-blinded) event-count distribution showed that this indicator yielded event prevalence exceeding 50 percent across the full cohort, a rate more consistent with a baseline severity proxy in this vasopressor-dependent population than with a treatment-induced hemodynamic event, and at such high baseline prevalence the indicator is statistically underpowered to detect incremental treatment-attributable change. The reclassification was made on the basis of this marginal distribution alone; by-arm mean arterial pressure (MAP)-derived event rates were examined only after the reclassification decision had been finalized. We disclose this reclassification as not fully prespecification-blinded, since the marginal cohort distribution itself constitutes a data-derived observation; the corresponding by-arm MAP-derived event rates in the unmatched analytic cohort, and the by-arm event rates for the MAP component are reported in Supplementary Methods S3 for transparency. Full specification of all outcome components is also provided in Supplementary Methods S3. Secondary outcomes were progression of AKI within 7 days of time zero, defined as any increase in maximum post-index Kidney Disease: Improving Global Outcomes (KDIGO) stage [9, 34] relative to the maximum stage in the 24 hours preceding time zero, and all-cause mortality within 28 days of time zero.

Covariates

Baseline covariates measured at time zero comprised demographics (age, sex, weight), ICU timing variables, norepinephrine dynamics (current dose, dose at 3 and 6 hours prior, peak in the prior 6 hours, and a categorical trajectory class coded as rising, stable, falling, or unknown; the algorithmic definition of the trajectory class is provided in Supplementary Methods S2b), concomitant vasoactive agents, hemodynamic parameters (mean arterial pressure, heart rate, central venous pressure, peripheral oxygen saturation), laboratory values (creatinine, lactate, albumin, chloride, bicarbonate, potassium, arterial pH, blood urea nitrogen), cumulative fluid intake, output, and balance, the Sequential Organ Failure Assessment (SOFA) score [35] (computed over the 24 hours preceding time zero using the mimiciv_3_1_derived.sofa table, abbreviated in variable form as sofa_total_24h), invasive mechanical

ventilation status, the Charlson Comorbidity Index [36], and binary indicators for major comorbidities (with collinearity between the aggregate Index and individual component indicators tolerated, as both forest splits and propensity score estimation handle correlated covariates without estimator bias). Vital signs used a 6-hour lookback window and laboratory values a 24-hour lookback. Missing values were handled using a missing-indicator approach combined with median imputation for variables with substantial missingness, and silent median imputation for variables with low missingness; full specifications are reported in Supplementary Methods S4 and exact missingness rates in Supplementary Table S4. The 49 evaluated covariates correspond to 48 numeric columns of the design matrix used by the propensity score and causal forest models (model.matrix absorbs the trajectory-unknown level as the reference category), so balance and variable-importance counts are reported as 49 evaluated covariates (45 of 49 balanced; equivalent to 44 of 48 design-matrix columns excluding the absorbed reference).

Statistical analysis

Primary causal contrast

The primary effect estimate was the average treatment effect on the treated (ATT) of furosemide on the composite primary outcome in the matched cohort. Propensity score matching (PSM) used nearest-neighbor matching on the logit propensity [37–39] with a target ratio of up to 1:2, caliper 0.2 standard deviations of the logit propensity, and without replacement; the caliper truncation reduced the achieved average ratio to approximately 1:1.52 (1,521 treated patients matched to 2,315 controls). The propensity score model included all baseline covariates as main effects with no interaction or spline terms. Covariate balance was assessed using absolute standardized mean differences (SMDs), with values below 0.10 indicating acceptable balance. The propensity score distribution by treatment group is shown in Supplementary Fig. S2. Treatment effects on binary outcomes were estimated by logistic regression on the matched cohort using only the treatment indicator as a predictor, with MatchIt-returned matching weights (each treated patient weight = 1; each control receives weight 1/k for k matched controls in the same set, preserving within-set weight sum). Odds ratios (ORs) with 95 percent confidence intervals (CIs) were obtained from a generalized linear model with cluster-robust standard errors at the matched-set level. The weighted estimates may differ modestly from crude 2×2 contingency-table odds ratios: for example, the secondary AKI outcome in the exact-plus-propensity matched cohort has a raw 2×2 odds ratio of approximately 1.11 versus a weighted estimate of 1.25 (95% CI 1.10 to 1.43).

Sensitivity analyses

Two prespecified sensitivity analyses assessed robustness. First, four time-related covariates retained residual imbalance after primary matching (hours since first norepinephrine, hours since ICU admission, cumulative fluid balance, cumulative output volume); we performed exact matching on quartiles of two of these (hours since first norepinephrine and cumulative fluid balance — the principal index within each

of two tightly correlated pairs in our cohort) combined with PSM on all remaining covariates (exact-plus-propensity matching). Imposing exact matching on all four would have over-restricted the matched set without addressing additional independent residual confounding. Second, we re-estimated the treatment effect using stabilized inverse probability of treatment weighting with weights truncated at 10. To quantify the magnitude of unmeasured confounding required to fully explain the observed associations, E-values [40] were computed on the risk-ratio scale (Supplementary Methods S8): each odds ratio was converted to a risk ratio (RR) at the control-arm event rate ($RR = OR / (1 - p_0 + p_0 \cdot OR)$), then $E = RR + \sqrt{(RR \times (RR - 1))}$ was applied. This RR-based approach is appropriate because the outcomes analyzed here are not rare (control event rates 11.3 to 51.5 percent). Point and lower-confidence-limit E-values are reported.

Heterogeneous treatment effect analysis

To investigate whether the treatment effect varied across patient phenotypes, we estimated conditional average treatment effects (CATEs) using a causal forest [27, 28] with 4,000 trees, honesty splitting, and automated hyperparameter tuning, fit on the full unmatched cohort. The covariate set comprised 49 evaluated baseline variables; the four-level norepinephrine trajectory class entered as indicator columns with the unknown level absorbed as the reference, yielding a final design matrix of 6,759 × 48 numeric columns. In addition to the population-average treatment effect (ATE), we report the average treatment effect on the treated (ATT) and on the controls (ATC) from the same model. Heterogeneity was formally tested using the calibration test of Athey, Tibshirani, and Wager [28], in which a coefficient on the differential forest prediction significantly greater than zero indicates heterogeneity. Variable importance, computed as the depth-weighted frequency with which each covariate was selected as a splitting variable, characterized the relative contribution of each covariate domain. To translate CATE estimates into an interpretable decision rule, a depth-2 policy tree [29] was fit using doubly robust scores [41] from the causal forest, with a prespecified set of six clinically interpretable covariates as candidate splitters. Causal forest hyperparameters and policy tree specifications are detailed in Supplementary Methods S5–S6.

Prespecified subgroup and phenotype analyses

Three subgroup analyses were prespecified: by norepinephrine trajectory class (rising, stable, falling); by quartile of cumulative fluid balance; and by quartile of SOFA score. Subgroup-conditional treatment effects were estimated as the within-stratum mean of patient-level CATE predictions from the primary causal forest. Ninety-five percent CIs were constructed using the doubly robust influence-function approach: variance of the augmented inverse probability-weighted scores was estimated at the cohort level ($\sigma^2_\Gamma = n \cdot SE^2_{ATE}$) and scaled to the subgroup sample size ($SE_S = \sqrt{(\sigma^2_\Gamma / n_S)}$), under the working assumption of homoscedastic score variance across strata. In a separate exploratory phenotype analysis, the cohort was stratified by clinically meaningful cut points for blood urea nitrogen, serum albumin, chloride, potassium, lactate, and central venous pressure, with mean CATE compared with the population-average mean CATE within each stratum. Cut points and complete stratum-level

results are reported in Supplementary Methods S7 and Supplementary Table S5. The primary outcome was a single prespecified composite; secondary outcomes, sensitivity analyses, and stratified subgroup analyses are reported without formal multiplicity adjustment and are interpreted as exploratory.

External validation cohort

To assess the generalizability of the primary findings, we replicated the propensity score-matched analysis in the eICU-CRD, a multicenter critical care database covering 208 United States hospitals between 2014 and 2015 [31]. Eligibility criteria mirrored the primary analysis: adult age, first ICU stay per patient, ICU length of stay at least 24 hours, cumulative norepinephrine infusion duration at least 120 minutes, and exclusion of end-stage kidney disease, renal replacement therapy within the first 24 hours of ICU admission, and intravenous furosemide administration before norepinephrine initiation.

Norepinephrine doses were harmonized from the unit-labeled drug names in the infusionDrug table: records labeled mcg/min were normalized by patient weight, records labeled mcg/kg/min were used directly, and records labeled ml/hr were converted assuming the standard 4 mg in 250 mL concentration. Intravenous furosemide events were drawn from both the medication table (bolus, predominant route) and the infusionDrug table (continuous infusion). The primary composite outcome and the new vasopressor initiation component were operationalized as in the primary analysis, with the norepinephrine dose escalation component adapted to the hourly snapshot structure of eICU infusion records (at least three consecutive post-index snapshots at or above 1.3 times the index dose, spanning at least 120 minutes). AKI progression at 7 days was operationalized using the serum creatinine component of KDIGO alone, because urine output ascertainment in eICU is less reliable than in MIMIC-IV. Mortality was restricted to in-hospital death because eICU lacks systematic out-of-hospital death tracking; the 28-day mortality endpoint used in the primary analysis could not be replicated. PSM used the same 1:2 nearest-neighbor specification with caliper of 0.2 standard deviations of the logit propensity, applied to a covariate set parallel to the primary analysis with Acute Physiology and Chronic Health Evaluation IV (APACHE IV) substituting for SOFA (the latter is not available as a derived table in eICU). Full pipeline specifications, harmonization procedures, and outcome adaptations are detailed in Supplementary Methods S9.

Software

Analyses were performed in R version 4.5.2 using the MatchIt [42] and cobalt packages for matching and balance, WeightIt for inverse probability weighting, grf for causal forest estimation, and policytree for policy tree fitting. Data extraction was performed in Google BigQuery. The complete extraction and analysis pipeline, together with a data dictionary, is openly available at Zenodo (https://doi.org/10.5281/zenodo.20394738).

Results

Cohort assembly and baseline characteristics

From an initial pool of 52,447 adult ICU stays meeting the basic length-of-stay criterion in MIMIC-IV v3.1, sequential exclusion of patients with end-stage kidney disease or early renal replacement therapy (n = 2,689 excluded), insufficient norepinephrine exposure or unresolvable infusion units (n = 42,003 excluded), and prior furosemide administration before norepinephrine initiation (n = 965 excluded) produced an eligible cohort of 6,790 patients. Among these, 1,667 (24.6 percent) received intravenous furosemide while on norepinephrine and were classified as treated; the remaining 5,123 patients constituted the control pool, from which one index time per patient was randomly sampled to yield 5,092 controls with complete index-time covariates, for a final analytic dataset of 6,759 patients (Fig. 1). No treated patient was excluded for incomplete index-time covariates because, by definition, the treated index time coincides with the first intravenous furosemide administration and is therefore necessarily observed; 31 candidate control patients were excluded at this step. Median age in the full analytic cohort was 68 years (interquartile range 57–78), median SOFA score at time zero was 7 (5–9), and median norepinephrine dose at time zero was 0.07 μg/kg/min (0.04–0.14), characteristics consistent with a moderate-to-severe vasopressor-dependent ICU population. Detailed baseline characteristics of the unmatched analytic cohort by treatment assignment are provided in Supplementary Table S2.

PSM matched 1,521 of 1,667 treated patients (91.2 percent) to 2,315 controls (matched cohort n = 3,836; achieved average matching ratio 1:1.52, with the caliper of 0.2 SD on the logit propensity excluding the remaining 146 treated patients who lacked sufficiently similar controls). After matching, 45 of 49 evaluated covariates (equivalent to 44 of 48 design-matrix columns excluding the absorbed trajectory-unknown reference level; see Methods Covariates) achieved an absolute SMD below 0.10; four time-related covariates retained marginal residual imbalance (hours since first norepinephrine, hours since ICU admission, cumulative fluid balance, cumulative output volume). A prespecified sensitivity analysis addressed this using exact matching on quartiles of two of these covariates (hours since first norepinephrine and cumulative fluid balance), the principal index within each of two tightly correlated pairs among the imbalanced variables (Methods Sensitivity analyses). Baseline characteristics of the matched cohort are shown in Table 1; full balance diagnostics are presented in Supplementary Fig. S1.

Table 1
Baseline characteristics of the propensity score–matched cohort (n = 3,836).

Characteristic	Diuretic (n = 1,521)	Control (n = 2,315)
Demographics		
Age, years	69 (59, 78)	69 (59, 80)
Male sex, n (%)	902 (59)	1,377 (59)
Weight, kg	83 (70, 99)	82 (69, 99)
ICU and vasopressor context		
Hours since ICU admission	30 (15, 62)	21 (10, 44)
Hours since first norepinephrine	22 (9, 52)	14 (5, 32)
Norepinephrine dose at index, μg/kg/min	0.08 (0.04, 0.15)	0.07 (0.04, 0.14)
Norepinephrine trajectory class, n (%)		
Unknown	491 (32)	788 (34)
Rising	236 (16)	341 (15)
Stable	357 (23)	511 (22)
Falling	437 (29)	675 (29)
Vasopressin co-administration, n (%)	328 (22)	437 (19)
Invasive mechanical ventilation, n (%)	1,028 (68)	1,557 (67)
Disease severity		
SOFA score	8 (6, 10)	8 (6, 10)
Charlson Comorbidity Index	5 (3, 7)	5 (3, 7)
Congestive heart failure, n (%)	719 (47)	1,073 (46)
Chronic kidney disease, n (%)	346 (23)	517 (22)
Severe liver disease, n (%)	124 (8.2)	188 (8.1)
Hemodynamics and laboratory values		
Mean arterial pressure, mmHg	72 (65, 79)	72 (65, 79)
Central venous pressure, mmHg	13 (10, 17)	12 (9, 16)
Lactate, mmol/L	1.70 (1.20, 2.50)	1.80 (1.30, 2.60)
Creatinine, mg/dL	1.30 (0.80, 2.00)	1.20 (0.80, 2.00)

Characteristic	Diuretic (n = 1,521)	Control (n = 2,315)
Demographics		
Albumin, g/dL	2.80 (2.30, 3.20)	2.80 (2.40, 3.30)
Fluid status (cumulative to index time)		
Cumulative fluid balance, mL	−2,265 (− 5,145, − 835)	−1,555 (− 3,465, − 660)
Outcomes (matched cohort)		
Primary composite (24 h), n (%)	593 (39.0)	535 (23.1)
AKI progression (7 d), n (%)	767 (50.4)	1,192 (51.5)
28-day mortality, n (%)	508 (33.4)	713 (30.8)

Continuous variables are reported as median (Q1, Q3); categorical variables as n (%). Descriptive statistics are computed on observed values prior to imputation; the regression and causal forest analyses used median imputation with missing-indicator coding for variables with substantial missingness (Supplementary Methods S4). In matched-cohort reporting, between-arm balance is assessed by absolute standardized mean differences (SMDs) rather than by hypothesis tests, because conventional p-values in large matched samples can flag clinically negligible differences as statistically significant. SMD diagnostics for all covariates are presented in Supplementary Fig. S1; 45 of 49 evaluated covariates achieved post-matching SMD < 0.10 (acceptable balance), with the four time-related covariates retaining marginal residual imbalance addressed in the exact-plus-PSM sensitivity analysis (Methods; Table 2*). Abbreviations: ICU, intensive care unit; SOFA, Sequential Organ Failure Assessment; AKI, acute kidney injury.*

Primary outcome and average treatment effect

In the primary propensity score-matched analysis, furosemide administration during active norepinephrine infusion was associated with a substantially higher risk of hemodynamic destabilization within 24 hours: the primary composite outcome occurred in 593 of 1,521 (39.0 percent) treated patients versus 535 of 2,315 (23.1 percent) matched controls, yielding an odds ratio of 2.03 (95% CI, 1.77 to 2.34; p < 0.001) (Table 2). Decomposition of the composite revealed that the signal was driven principally by the norepinephrine dose escalation component, with treated patients experiencing sustained escalation of at least 1.3 times the index dose in 474 of 1,521 (31.2 percent) of cases versus 322 of 2,315 (13.9 percent) of controls, yielding an odds ratio of 2.66 (95% CI, 2.27 to 3.12; p < 0.001). The new vasopressor initiation component showed a smaller and statistically non-significant difference (13.1 percent versus 11.3 percent; odds ratio 1.15, 95% CI 0.95 to 1.40; p = 0.15).

Table 2
Treatment effects of furosemide administration during active norepinephrine infusion, across matching and weighting strategies.

Outcome	Diuretic, n (%)	Control, n (%)	Odds ratio (95% CI)	p-value
Primary analysis — PSM 1:2 (n = 3,836)				
Primary composite outcome (24 h)	**593 (39.0)**	**535 (23.1)**	**2.03 (1.77–2.34)**	**< 0.001**
Norepinephrine dose escalation	474 (31.2)	322 (13.9)	2.66 (2.27–3.12)	< 0.001
New vasopressor initiation	200 (13.1)	262 (11.3)	1.15 (0.95–1.40)	0.15
AKI progression (7 d)	767 (50.4)	1,192 (51.5)	1.03 (0.91–1.17)	0.63
All-cause mortality (28 d)	508 (33.4)	713 (30.8)	1.05 (0.91–1.20)	0.51
Sensitivity analysis 1 — exact match on key time-related covariates + PSM (1,439 treated, 2,227 controls, total n = 3,666)				
Primary composite outcome (24 h)	553 (38.4)	534 (24.0)	1.92 (1.67–2.22)	< 0.001
Norepinephrine dose escalation	442 (30.7)	345 (15.5)	2.33 (1.99–2.73)	< 0.001
AKI progression (7 d)	733 (50.9)	1,074 (48.2)	1.25 (1.10–1.43)	< 0.001
All-cause mortality (28 d)	491 (34.1)	730 (32.8)	1.01 (0.88–1.16)	0.91
Sensitivity analysis 2 — stabilized IPTW, full cohort (n = 6,759)				
Primary composite outcome (24 h)	—	—	2.05 (1.81–2.31)	< 0.001

Effect estimates are odds ratios (OR) with 95% CIs from logistic regression on the matched (PSM and exact + PSM) or weighted (IPTW) cohort. Because event rates are not rare (control-arm rates 11.3 to 51.5 percent), the corresponding risk ratios are appreciably closer to the null than the ORs (OR-to-RR conversion given in Table 3*). PSM, PSM (1:2 nearest-neighbor, caliper 0.2 SD on the logit propensity). Exact match was performed on quartiles of hours since first norepinephrine and cumulative fluid balance; the exact-plus-PSM cohort comprised 1,439 treated and 2,227 controls. IPTW, stabilized inverse probability of treatment weighting with weights truncated at 10, applied to the primary composite only (weighted estimates; event counts not reported). The primary composite was norepinephrine dose escalation to ≥ 1.3× the index dose for ≥ 120 minutes, or initiation of a second vasoactive agent, within 24 hours of time zero. AKI, acute kidney injury.*

Risk-difference estimates from the causal forest were consistent with the matched-cohort logistic regression: the doubly robust population-average treatment effect (ATE) on the primary outcome was 0.161 (95% CI, 0.135 to 0.188), corresponding to approximately 16 additional events per 100 patients had the entire cohort been exposed (Table 3). The average treatment effect on the treated (ATT) was 0.147 (95% CI, 0.119 to 0.175) and the average treatment effect on the controls (ATC) was 0.167 (95% CI, 0.132 to 0.202). The point-estimate ordering ATT < ATE < ATC is suggestive of clinicians having preferentially administered furosemide to patients in whom the hemodynamic cost was anticipated to be lower, although the three CIs overlap substantially and the difference between ATT and ATC is not statistically distinguishable; the overlap also supports the validity of the population-average treatment effect (ATE) despite the asymmetric distribution of propensity scores between arms.

Table 3
Causal forest estimates, heterogeneity test, and E-values for unmeasured confounding.

Parameter	Estimate	Standard error	95% CI
Causal forest estimates (risk-difference scale)			
ATE (entire cohort)	0.161	0.014	0.135 to 0.188
ATT (treated subset)	0.147	0.015	0.119 to 0.175
ATC (control subset)	0.167	0.018	0.132 to 0.202
Heterogeneity calibration test			
Mean forest prediction coefficient	1.037	0.098	$p < 0.001$
Differential forest prediction coefficient	1.414	0.379	$p < 0.001$
E-values for unmeasured confounding (primary composite outcome, RR-based)			
Primary analysis (PSM)	2.66	—	lower CI: 2.37
Exact + PSM	2.52	—	lower CI: 2.23
IPTW	2.69	—	lower CI: 2.42
E-values for unmeasured confounding (norepinephrine dose escalation, strongest signal, RR-based)			
Primary analysis (PSM)	3.75	—	lower CI: 3.27
Exact + PSM	3.27	—	lower CI: 2.84

ATE, average treatment effect (entire cohort); ATT, average treatment effect on the treated; ATC, average treatment effect on the controls; all on the absolute risk-difference scale. The heterogeneity calibration test follows Athey, Tibshirani, and Wager [28]: *the mean prediction coefficient ideally equals 1 when the average forest prediction matches the population ATE (here 1.037, consistent with a well-calibrated mean); the differential prediction coefficient is positive in the presence of treatment effect heterogeneity, with a value of 1 corresponding to perfect within-sample calibration of heterogeneity, values greater than 1 suggesting that the spread of patient-level CATE predictions modestly underestimates the true heterogeneity (here 1.414, indicating clear within-sample heterogeneity captured by the forest). E-values are reported on the risk-ratio scale per VanderWeele and Ding* [40]: *the closed-form expression $E = RR + \sqrt{(RR \times (RR - 1))}$ was*

applied after converting each odds ratio to its corresponding risk ratio using the control-arm event rate (RR = OR / (1 − p_0 + p_0·OR)), because the outcomes analyzed in this study are not rare (control event rates 11.3 to 51.5 percent across outcomes) and direct application of the OR-based formula would overstate the E-value. The E-value represents the minimum strength of association on the risk-ratio scale that an unmeasured confounder would need to have with both treatment and outcome, conditional on the measured covariates, to fully explain away the observed association.

Secondary outcomes

Furosemide administration was not associated with progression of AKI at 7 days in the primary analysis (50.4 percent versus 51.5 percent; odds ratio 1.03, 95% CI 0.91 to 1.17; p = 0.63), nor with 28-day all-cause mortality (33.4 percent versus 30.8 percent; odds ratio 1.05, 95% CI 0.91 to 1.20; p = 0.51) (Table 2). The dissociation between a clearly positive proximal hemodynamic signal and null distal outcomes is discussed below.

Sensitivity and robustness analyses

The primary association was robust across all prespecified sensitivity analyses (Fig. 2). Exact matching on quartiles of hours since first norepinephrine and cumulative fluid balance, combined with PSM on remaining covariates (1,439 treated and 2,227 controls, total n = 3,666), produced an odds ratio for the primary composite of 1.92 (95% CI, 1.67 to 2.22), and for norepinephrine escalation alone of 2.33 (95% CI, 1.99 to 2.73), confirming that the residual time-related imbalance after primary matching could not account for the observed effect. In this more strictly matched cohort, AKI progression at 7 days was positively associated with furosemide exposure (odds ratio 1.25, 95% CI 1.10 to 1.43; p < 0.001), in contrast with the null result from the primary matching strategy; the implications of this between-analysis divergence are addressed in the Discussion. Stabilized inverse probability of treatment weighting on the full cohort yielded an odds ratio of 2.05 (95% CI, 1.81 to 2.31) for the primary outcome, fully consistent with the matched-cohort estimates.

E-values on the risk-ratio scale were uniformly high across all three matching strategies, ranging from 2.52 to 2.69 for the primary composite outcome at the point estimate, with corresponding lower-confidence-limit E-values from 2.23 to 2.42 (Table 3). For norepinephrine dose escalation specifically, the strongest signal in the dataset, the E-value reached 3.75 in the primary PSM analysis (lower CI 3.27) and 3.27 in the exact-plus-PSM sensitivity analysis (lower CI 2.84), indicating that an unmeasured confounder would need to be associated with both furosemide administration and norepinephrine escalation by a RR of at least approximately 3, conditional on the measured covariates, to fully explain away the observed association in the primary analysis.

Heterogeneous treatment effects

Across the 6,759 patients, conditional average treatment effect (CATE) predictions from the causal forest ranged from 0.038 to 0.275 with a unimodal distribution; the arithmetic mean of patient-level predictions was 0.154, while the doubly robust population-average ATE was 0.161 (Fig. 3, Table 3), with

the modest difference reflecting the two estimators' distinct weighting of patient-level scores. All 6,759 individual CATE predictions were positive, indicating qualitative absence of effect sign reversal across the cohort. The calibration test of Athey, Tibshirani, and Wager [28] confirmed statistically significant heterogeneity in effect magnitude across patients (coefficient on the differential forest prediction 1.41, $p < 0.001$; Table 3).

Variable importance analysis (Fig. 4; full ranking of the top 25 covariates in Supplementary Table S3) identified current norepinephrine dose at index time as the single most important driver of treatment effect heterogeneity (importance 0.163, more than twice the next variable). The next five most important covariates spanned three domains: cumulative fluid balance (rank 2, importance 0.075), hours since first norepinephrine (rank 3, 0.059), congestive heart failure (rank 4, 0.057), blood urea nitrogen (rank 5, 0.053), and rising norepinephrine trajectory class (rank 6, 0.045). Aggregated by category, indication- and decision-related variables (current and prior norepinephrine doses, peak dose in the preceding 6 hours, hours since first norepinephrine and since ICU admission, and trajectory class) collectively accounted for 41 percent of total variable importance; biological phenotype variables 32 percent; fluid management 11 percent; hemodynamic state 12 percent; and missing-indicator and composite severity variables the remaining 4 percent.

A depth-2 policy tree fit to the doubly robust scores from the causal forest yielded an interpretable depth-2 decision structure with three splitting variables: among patients with a low six-hour-prior norepinephrine dose (less than 0.06 μg/kg/min), the algorithm's split depended on current norepinephrine dose; among those with higher prior dose, the split depended on time since ICU admission. Because all CATE predictions were positive, the leaf labels rank subgroups by predicted hemodynamic cost rather than benefit (Supplementary Methods S6 and Fig. S3); the policy tree is interpreted as a hypothesis-generating tool for relative-risk ordering, not a definitive clinical rule.

Prespecified subgroup and phenotype analyses

Across the three prespecified subgroup analyses (Fig. 5), all within-stratum mean conditional average treatment effects (CATEs) from the primary causal forest were positive, ranging from 0.108 (norepinephrine trajectory rising subgroup) to 0.172 (least-negative cumulative fluid balance quartile). The sample-size–weighted average of within-stratum means returned to the population-average mean CATE of 0.154 (Table 3, Fig. 3) by construction when all levels of the stratification variable are included; for the norepinephrine trajectory analysis this identity holds across all four classes (rising, stable, falling, unknown), with the unknown class (n = 2,862, 42.3 percent of the analytic cohort, substantially overlapping with the subset for whom the six-hour-prior norepinephrine dose was unobserved) excluded from display because it is not clinically interpretable. Within the three classifiable trajectory subgroups (n = 3,897, 57.7 percent), the within-stratum mean CATE was lowest in the rising class (0.108, 95% CI 0.034 to 0.182; n = 867) and increased monotonically through stable (0.138, 0.077 to 0.199; n = 1,278) and falling (0.158, 0.106 to 0.210; n = 1,752), counter to the prior expectation that patients with increasing vasopressor requirements would experience the largest treatment effect. By cumulative fluid balance quartile (n ≈ 1,690 per quartile; overall median − 1,541 mL, range − 6,458 to − 234 mL across

quartile medians; the predominantly negative values reflect underrecording of oral and enteral intake in MIMIC-IV's inputevents table), the within-stratum mean CATE increased monotonically from 0.131 (95% CI 0.078 to 0.184) in the most negative quartile through 0.150 (0.097 to 0.203), 0.162 (0.109 to 0.215), and 0.172 (0.119 to 0.225) in the least negative. By SOFA quartile (n = 1,833, 1,824, 1,468, and 1,634; unequal sizes reflect tied values at discrete-score cut points), CATE decreased monotonically with illness severity from 0.164 (95% CI 0.113 to 0.215) through 0.157 (0.106 to 0.208), 0.149 (0.092 to 0.206), and 0.142 (0.088 to 0.196). All eleven CIs excluded zero, consistent with the absence of effect sign reversal in the patient-level CATE distribution.

Stratification by biological phenotype variables revealed minimal within-variable variation in CATE: across blood urea nitrogen quartiles, mean CATE ranged from 0.145 to 0.159; across the three albumin strata (< 2.5, 2.5–3.5, > 3.5 g/dL), from 0.147 to 0.158; across chloride strata, from 0.152 to 0.158; across potassium strata, from 0.153 to 0.155; across lactate strata (n = 5,471 with recorded values), from 0.151 to 0.154; and across patients with recorded central venous pressure (CVP) (n = 2,376), from 0.143 to 0.151 between lowest and highest CVP strata. Stratum-level details are presented in Supplementary Table S5. The narrow range of CATE within biological phenotype strata, in contrast with the wider range across vasopressor-related strata, indicates that while biological variables contribute meaningfully to the variable-importance ranking, their stratum-level effect modification is small relative to current norepinephrine dose and the time-course of vasopressor dependence.

External validation in eICU-CRD

Of 200,859 ICU stays in the eICU-CRD, 7,137 patients met all primary-analysis eligibility criteria after sequential exclusion and norepinephrine dose unit harmonization; 775 (10.9 percent) received intravenous furosemide during norepinephrine infusion and were classified as treated, with the remaining 6,362 forming the control pool. The lower treated proportion compared with MIMIC-IV (10.9 versus 24.6 percent) is consistent with eICU's heterogeneous community-hospital case mix and broader institutional variation in deresuscitation practice. Cohort assembly is detailed in Supplementary Methods S9; baseline characteristics appear in Supplementary Table S6.

PSM with the same 1:2 nearest-neighbor specification matched 772 of 775 treated patients (99.6 percent) to 1,535 controls (matched cohort n = 2,307; achieved ratio 1:1.99; better propensity score overlap than in MIMIC-IV). All 42 of 42 evaluated covariates achieved an absolute SMD below 0.10 after matching (Supplementary Table S7). On the matched cohort, the primary composite outcome occurred in 29.7 percent of treated patients versus 19.8 percent of controls (odds ratio 1.71, 95% CI 1.40 to 2.09; $p < 0.001$; RR approximately 1.50 at the control-arm event rate). Decomposition gave odds ratios of 1.62 (95% CI 1.28 to 2.06) for sustained norepinephrine dose escalation and 1.65 (1.27 to 2.15) for new vasopressor initiation (both $p < 0.001$). AKI progression at 7 days was positively associated with furosemide exposure (odds ratio 2.01, 95% CI 1.65 to 2.44; $p < 0.001$). In-hospital mortality showed no significant association (33.5 versus 31.3 percent; odds ratio 1.10, 95% CI 0.91 to 1.33; $p = 0.32$) (Table 4, Fig. 6).

Table 4
External validation: treatment effects in the MIMIC-IV derivation and eICU-CRD validation cohorts.

Outcome	Database	Matched n	Treated, %	Control, %	OR (95% CI)	p value
Primary composite (24 h)	MIMIC-IV	3,836	39.0	23.1	2.03 (1.77–2.34)	< 0.001
	eICU-CRD	2,307	29.7	19.8	1.71 (1.40–2.09)	< 0.001
Norepinephrine dose escalation	MIMIC-IV	3,836	31.2	13.9	2.66 (2.27–3.12)	< 0.001
	eICU-CRD	2,307	19.2	12.8	1.62 (1.28–2.06)	< 0.001
New vasopressor initiation	MIMIC-IV	3,836	13.1	11.3	1.15 (0.95–1.40)	0.15
	eICU-CRD	2,307	14.9	9.6	1.65 (1.27–2.15)	< 0.001
AKI progression (7 d) †	MIMIC-IV	3,836	50.4	51.5	1.03 (0.91–1.17)	0.63
	eICU-CRD	2,307	33.2	19.9	2.01 (1.65–2.44)	< 0.001
Mortality ‡	MIMIC-IV	3,836	33.4	30.8	1.05 (0.91–1.20)	0.51
	eICU-CRD	2,307	33.5	31.3	1.10 (0.91–1.33)	0.32

OR, odds ratio; CI, confidence interval; AKI, acute kidney injury. PSM used 1:2 nearest-neighbor matching on the logit propensity, caliper 0.2 standard deviations, without replacement; treatment effects estimated by logistic regression on the matched cohort with cluster-robust standard errors at the matched-set level. Direction concordance is defined as both odds ratios on the same side of 1.0, irrespective of statistical significance.

† MIMIC-IV uses the mimiciv_3_1_derived.kdigo_stages table integrating the serum creatinine and urine output components of KDIGO 2012; eICU-CRD uses the serum creatinine component alone (post-index serum creatinine at least 1.5× baseline or at least 0.3 mg/dL above baseline within 7 days), because the eICU intakeOutput table has substantially incomplete coverage of catheterized urine output. The two definitions are not directly comparable on the magnitude scale.

‡ MIMIC-IV reports 28-day all-cause mortality combining in-hospital records with external death index linkage; eICU-CRD reports in-hospital mortality because the database lacks systematic out-of-hospital death tracking. The two definitions are not directly comparable on the magnitude scale but converge to a null association in both cohorts.

External validation forest plot comparing treatment effects in the MIMIC-IV derivation cohort (navy, n = 3,836 matched) and the eICU-CRD validation cohort (red, n = 2,307 matched) across the five prespecified outcomes; within each row the MIMIC-IV estimate is plotted above the eICU-CRD estimate. Horizontal lines are 95% CIs; the dotted line at OR = 1 is the null. Odds ratios (95% CI) are annotated at right. Direction of effect was concordant for all five outcomes, with attenuated magnitude in the multicenter validation cohort. AKI and mortality definitions differ between databases (see Table 4 *footnotes).*

The direction of effect was concordant across all five outcomes between the two databases (Table 4, Fig. 6). The magnitude of the primary composite and dose escalation odds ratios in eICU was attenuated relative to MIMIC-IV (primary 1.71 versus 2.03; norepinephrine escalation 1.62 versus 2.66), consistent with both the broader case mix of a multicenter community-hospital sample and the coarser hourly snapshot operationalization of dose escalation in eICU. The risk-ratio-scale E-value for the primary composite in the eICU-matched cohort was 2.36 at the point estimate (1.92 at the lower confidence limit); for norepinephrine dose escalation the corresponding E-values were 2.37 and 1.78. The new vasopressor initiation and AKI components reached statistical significance in eICU while remaining null in the primary MIMIC-IV propensity-matched analysis; the eICU AKI signal (odds ratio 2.01) is directionally consistent with the positive estimate (1.25) observed in the MIMIC exact-plus-propensity sensitivity analysis, compatible with the interpretation that the association reflects a real effect attenuated by residual confounding in the primary MIMIC analysis. In-hospital mortality was null in both databases, consistent with the proximal-distal dissociation observed in the derivation cohort.

Discussion

In this target trial emulation of 6,759 vasopressor-dependent ICU patients from MIMIC-IV, intravenous furosemide initiation during active norepinephrine infusion was associated with a sustained norepinephrine dose escalation event in the matched cohort with an odds ratio of 2.66 (95% CI 2.27 to 3.12; relative risk approximately 2.16 at the control-arm event rate); the doubly robust population-average treatment effect on the primary composite outcome was an absolute risk difference of 0.161 (95% CI 0.135 to 0.188), approximately 16 percentage points (corresponding to a number needed to harm of approximately 6). The signal was robust across alternative matching strategies and stabilized inverse probability of treatment weighting, with risk-ratio-scale E-values reaching 3.75 for the dose-escalation signal in the primary PSM analysis and 2.52 to 2.69 for the composite outcome — indicating that an unmeasured confounder would need a RR of at least 2.52 with both exposure and outcome to fully explain the observed association. The proximal hemodynamic effect did not translate into a measurable association with 28-day mortality, while the secondary AKI outcome was null in the primary analysis but positive (OR 1.25) under exact-plus-propensity matching. Treatment effect heterogeneity, formally established by the calibration test of Athey, Tibshirani, and Wager [28], was uniformly positive across all 6,759 patient-level CATE estimates, indicating that no patient subgroup was predicted to derive net hemodynamic benefit; current norepinephrine dose at index time emerged as the most important driver. External validation in the eICU-CRD (n = 7,137 across 208 United States hospitals) directionally

confirmed both the primary composite (odds ratio 1.71, 95% CI 1.40 to 2.09) and dose escalation (1.62, 1.28 to 2.06) signals.

Mechanistic interpretation

Several physiologic considerations support the observed proximal–distal dissociation. Furosemide reduces effective intravascular volume by promoting renal sodium and water excretion. In norepinephrine-dependent patients, venous capacitance is reduced by alpha-adrenergic vasoconstriction, and many lie on the ascending limb of the Frank–Starling curve where small reductions in preload produce disproportionate decrements in stroke volume. Compensatory escalation of catecholamine support is the expected acute response, and the binary norepinephrine escalation outcome captures precisely this physiologic signature. The within-composite divergence is informative: the strong signal on titration of the in-place agent (odds ratio 2.66 for norepinephrine dose escalation) but the null signal on initiation of a second agent (odds ratio 1.15 for new vasopressor) is consistent with clinician preference for titrating the established infusion rather than introducing a new drug class, a pattern of medical inertia aligned with bedside practice. The null effects on AKI in the primary analysis and on 28-day mortality, juxtaposed with the strong proximal signal, are consistent with the inference that ICU clinicians detect and correct the acute hemodynamic perturbation in real time, preventing translation into persistent organ injury or excess late mortality.

The direction of unmeasured confounding by indication cannot be determined a priori: clinicians may have selectively administered furosemide to patients in whom hemodynamic cost was anticipated to be lower (consistent with the point-estimate ordering ATT < ATC, although CIs overlap), or alternatively to patients on a deteriorating trajectory in whom subsequent escalation reflects underlying disease progression. Residual indication bias in either direction cannot be excluded, and the magnitude of the E-values (2.52 to 3.75) constrains — but does not eliminate — the plausibility of such confounding. The depth-2 policy tree partitions the cohort on prior and current norepinephrine dose and time since ICU admission, indicating that effect magnitude tracks with the operational state of vasopressor support rather than the laboratory variables interrogated in stratified analyses; we treat the tree as a hypothesis-generating tool for relative-risk ordering and discuss its limitations in the Clinical implications section below.

Comparison with prior literature

Bandak and colleagues (2020) [24], in a propensity-matched single-center analysis of 939 vasopressor-dependent ICU patients, compared mean hourly vasopressor dose before and after intravenous furosemide administration and reported no significant difference. In the same matched cohort, Bandak et al. also reported higher 90-day mortality in the diuretic-exposed group (45.2 versus 36.9 percent, p = 0.009), although the design did not include explicit time-zero alignment, did not employ quantitative bounds on unmeasured confounding, and the authors interpreted the mortality difference cautiously given likely residual confounding by indication. Our short-term mortality estimates fall within a follow-up

window (28 days in MIMIC-IV; in-hospital in eICU-CRD) shorter than Bandak's primary mortality endpoint, and the inference of no detectable short-term mortality association is supported by concordant null findings across both databases and by E-value quantification of unmeasured confounding required to explain the observed associations. The two frameworks are complementary rather than contradictory: mean-dose averaging dilutes a heterogeneous population-level signal when clinically meaningful escalations concentrated in an affected subset are averaged against stable or de-escalating doses in the larger remainder, whereas an event-based binary outcome with a clinically relevant threshold isolates the affected subset. The present analysis additionally demonstrates that no prespecified subgroup defined by baseline covariates yielded a mean CATE consistent with predicted benefit on the primary composite outcome.

A noteworthy internal divergence concerns the secondary AKI outcome. In the primary propensity-matched analysis the association was null (odds ratio 1.03, 95% CI 0.91 to 1.17), but in the exact-plus-PSM sensitivity analysis — which additionally controlled for residual imbalance in time-related covariates (hours since first norepinephrine and cumulative fluid balance) — the association reached an odds ratio of 1.25 (95% CI 1.10 to 1.43). Two interpretations are compatible. The first is that exact matching removed residual confounding by ICU course duration, unmasking a modest true effect attenuated in the primary analysis. The second is that the additional constraint reduced the sample to a more homogeneous subset in which residual differences favor an AKI association, producing a finding specific to that restricted population. The directional concordance of the dose-escalation effect across both matching strategies (2.66 versus 2.33 in exact-plus-PSM) argues against systematic distortion by exact matching. A 25 percent increase in the odds of AKI progression at 7 days cannot be excluded; the eICU-CRD validation cohort provided further support, with an odds ratio of 2.01 (95% CI 1.65 to 2.44) observed using a serum creatinine-based KDIGO definition. The convergence of the MIMIC exact-plus-PSM result with the independent eICU estimate is compatible with a real AKI signal attenuated in the primary analysis by residual confounding, though differing AKI definitions (KDIGO with urine output in MIMIC-IV, creatinine alone in eICU-CRD) preclude direct quantitative comparison.

External validation

The eICU-CRD validation cohort provides three substantive observations beyond directional concordance of the primary signal. First, the attenuation in magnitude (primary composite 1.71 in eICU versus 2.03 in MIMIC-IV; norepinephrine dose escalation 1.62 versus 2.66) is consistent with two non-exclusive mechanisms: the broader case mix in the multicenter community-hospital sample relative to the single-center academic derivation cohort, and the coarser operationalization of dose escalation under eICU's hourly snapshot infusion structure. Both mechanisms attenuate estimates toward the null without distorting direction, and persistence of statistical significance under both attenuating influences supports robustness of the underlying physiologic association. Second, the new vasopressor initiation component reached statistical significance in eICU (odds ratio 1.65, 95% CI 1.27 to 2.15) while remaining null in MIMIC-IV (1.15, 0.95 to 1.40); this divergence may reflect institutional heterogeneity in second-agent prescribing thresholds, with some hospitals favoring early addition of a second vasoactive agent

over upward titration of the established infusion — a pattern of cross-institutional variability that a single-center derivation cohort cannot capture. Third, in-hospital mortality in the eICU validation cohort was null (odds ratio 1.10, 95% CI 0.91 to 1.33), converging with the null 28-day mortality estimate in MIMIC-IV despite differing operationalizations; the convergence of two independent null findings on a hard distal endpoint, juxtaposed with the strong concordant proximal signal, supports the inference of an acute hemodynamic perturbation detected and corrected in real time without translating into excess late mortality.

Clinical implications

Two observations emerge that bear on bedside hypothesis generation. First, the conditional treatment effect varies meaningfully across patient strata without any single covariate domain exclusively governing heterogeneity. Variable-importance analysis (Supplementary Table S3) identifies current norepinephrine dose at index time as the dominant individual predictor (importance 0.163, more than twice the next variable), with cumulative fluid balance, congestive heart failure status, and blood urea nitrogen contributing meaningfully at lower ranks. Among the 57.7 percent of patients with a classifiable norepinephrine trajectory, patients with rising doses displayed the smallest within-stratum mean CATE (0.108), likely because the underlying disease course was already producing dose escalation independently of diuretic exposure; patients with stable (0.138) and falling (0.158) pressor requirements showed larger effects where furosemide-attributable perturbation had greater room to manifest. Cumulative fluid balance modified the effect in a monotonically increasing pattern across quartiles (0.131 through 0.172), while SOFA score showed an inverse pattern (0.164 in the lowest quartile through 0.142 in the highest). All eleven subgroup CIs excluded zero, consistent with the absence of effect sign reversal in the patient-level CATE distribution.

Second, the depth-2 policy tree (Supplementary Fig. S3) partitions the cohort on prior and current norepinephrine dose and time since ICU admission, ordering subgroups by predicted hemodynamic cost. The root split on six-hour-prior norepinephrine dose is influenced by median imputation for the 42.3 percent of patients in whom this value was unobserved; an observation-only sensitivity analysis (n = 3,897; Supplementary Results S.R.4) does not select this variable as the root split. The tree is therefore hypothesis-generating only, and specific dose thresholds should not be carried forward without external validation.

Taken together, these findings support the inference that the state of vasopressor support, encompassing current dose and the trajectory of dependence, is a candidate clinical dimension for timing diuretic administration to minimize hemodynamic cost, with specific operational thresholds requiring further validation in cohorts where the relevant pre-index dose history is fully observed.

Limitations

As an observational analysis within the target trial emulation framework, residual unmeasured confounding cannot be eliminated; the risk-ratio-scale E-values (2.52 to 3.75 across primary analyses) provide a quantitative bound on the association strength an unmeasured confounder would need to fully

explain the observed effects. The cohort encompasses heterogeneous shock states, with congestive heart failure present in 33 percent of the unmatched cohort and 46 to 47 percent of the matched cohorts; the mechanism of diuretic-induced hemodynamic perturbation may differ across distributive, cardiogenic, and postoperative states, and shock-type stratified analyses were not performed. The primary analysis derives from a single United States academic center spanning 2008 to 2022, during which fluid management paradigms evolved; era effects were not separately modeled. The multicenter eICU-CRD validation partially addresses generalizability by confirming directional concordance across 208 hospitals, although that dataset is itself restricted to 2014 to 2015 and to United States populations, and further replication in recent international cohorts remains warranted.

Several measurement-related considerations apply. AKI was defined using KDIGO with both creatinine and urine output in MIMIC-IV but the creatinine component alone in eICU-CRD, precluding direct quantitative comparison of AKI magnitude across databases; the pharmacologic interaction between loop diuretics and the urine-output component of KDIGO may further contribute to cross-database differences in measured AKI rates. Additionally, the eICU-CRD validation cohort's reliance on serum creatinine alone may partly reflect transient hemoconcentration from furosemide-induced volume contraction rather than true renal injury; the concordant positive estimate from the MIMIC-IV exact-plus-PSM analysis (OR 1.25), which incorporates both urine-output and creatinine criteria, provides partial counter-evidence that the eICU signal is not purely artifactual. Mortality in eICU was restricted to in-hospital deaths, so the 28-day endpoint of the derivation cohort was not directly replicable. Norepinephrine doses recorded in volume-rate units in eICU were converted using an assumed standard concentration of 4 mg in 250 mL; deviations would introduce dose-misclassification noise that biases odds ratios toward the null without affecting direction, because the 1.3-times escalation threshold is invariant under patient-level multiplicative scaling (Supplementary Results S.R.5). Cumulative fluid balance is systematically negative across all quartiles because the MIMIC-IV inputevents table omits oral and enteral intake; the monotonic effect modification is therefore conditional on this measurement scale. Four time-related covariates retained marginal residual imbalance after primary matching, partially addressed by the exact-plus-propensity sensitivity analysis. The control arm was anchored at a single randomly sampled index time per patient; while three matching strategies yielded consistent results, sensitivity under alternative sampling realizations was not tested. All secondary outcomes are exploratory and were not adjusted for multiplicity.

Conclusions

Among vasopressor-dependent adults in the MIMIC-IV derivation cohort, initiation of intravenous furosemide during active norepinephrine infusion was associated with a substantial increase in proximal hemodynamic destabilization within 24 hours, driven principally by sustained norepinephrine dose escalation; this primary association was directionally confirmed in the multicenter eICU-CRD validation cohort (odds ratio 1.71, 95% CI 1.40 to 2.09 for the primary composite; 1.62, 1.28 to 2.06 for dose escalation) across 208 United States hospitals. The proximal signal did not translate into a measurable association with mortality in either database. The secondary AKI signal — null in the MIMIC-IV primary

propensity-matched analysis but positive in both the MIMIC-IV exact-plus-PSM sensitivity analysis (OR 1.25) and the eICU-CRD validation cohort (OR 2.01) — is compatible with a real association attenuated by residual confounding in the primary analysis, though differing AKI ascertainment across the two databases precludes quantitative comparison. No patient subgroup was predicted to derive net hemodynamic benefit on the primary composite outcome: causal-forest estimates were positive across all 6,759 patients, with current norepinephrine dose at index time as the dominant driver of heterogeneity, and the depth-2 policy tree identified only relative-cost ordering. Among the 57.7 percent of patients with a classifiable norepinephrine trajectory, the timing of diuretic administration relative to the trajectory of vasopressor requirements emerges as a candidate target for hypothesis-generating bedside heuristics, with specific operational thresholds requiring external validation.

Abbreviations

AKI
Acute kidney injury
APACHE IV
Acute Physiology and Chronic Health Evaluation IV
ATC
Average treatment effect on the controls
ATE
Average treatment effect (population-average)
ATT
Average treatment effect on the treated
CATE
Conditional average treatment effect
CI
Confidence interval
CVP
Central venous pressure
eICU-CRD
eICU Collaborative Research Database
ESKD
End-stage kidney disease
ICD
International Classification of Diseases
ICU
Intensive care unit
IPTW
Inverse probability of treatment weighting
KDIGO

Kidney Disease:Improving Global Outcomes
MAP
Mean arterial pressure
MIMIC-IV
Medical Information Mart for Intensive Care IV
NNH
Number needed to harm
OR
Odds ratio
PSM
Propensity score matching
RECORD
REporting of studies Conducted using Observational Routinely-collected health Data
RR
Risk ratio
RRT
Renal replacement therapy
SD
Standard deviation
SMD
Standardized mean difference
SOFA
Sequential Organ Failure Assessment
STROBE
Strengthening the Reporting of Observational Studies in Epidemiology
VExUS
Venous Excess Ultrasound Grading System

Declarations

Ethics approval and consent to participate

This study is a secondary analysis of two publicly available, fully de-identified critical care databases (MIMIC-IV and the eICU Collaborative Research Database) and was conducted in accordance with the Declaration of Helsinki. The establishment of MIMIC-IV was approved by the Institutional Review Boards of the Beth Israel Deaconess Medical Center (Boston, MA, USA) and the Massachusetts Institute of Technology (Cambridge, MA, USA), which granted a waiver of the requirement for individual informed consent; the eICU-CRD is an Institutional Review Board–exempt de-identified database. The authors accessed both databases as credentialed users after completing the mandated human-subjects research training and signing the PhysioNet data use agreement. Because all records were de-identified,

no additional institutional review board approval or individual informed consent was required for this analysis.

Consent for publication

Not applicable. This study does not contain data from any individual person in any identifiable form.

Clinical trial number

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Funding

The authors received no specific funding for this work.

Author Contribution

LS and YL conceived and designed the study. LS and CX performed the data extraction and curation from MIMIC-IV and the eICU-CRD. LS and HT carried out the statistical analysis, including propensity-score matching, causal-forest estimation, and the policy-tree and sensitivity analyses. LS, CX, HT, and YL interpreted the results. LS drafted the manuscript. CX, HT, and YL critically revised the manuscript for important intellectual content. YL supervised the study and is the corresponding author. All authors read and approved the final manuscript.

Acknowledgements

Not applicable.

Data Availability

The datasets analysed during the current study are derived from two third-party, publicly available critical care databases. MIMIC-IV (v3.1) and the eICU Collaborative Research Database (v2.0) are maintained by the Laboratory for Computational Physiology at the Massachusetts Institute of Technology and are accessible through the PhysioNet platform to credentialed researchers who have completed the required human-subjects research training and signed the applicable data use agreement, at

https://physionet.org/content/mimiciv/3.1/ and https://physionet.org/content/eicu-crd/2.0/, respectively [30, 31]. The complete data extraction and analysis pipeline, together with a data dictionary supporting the findings of this study, is openly available in the Zenodo repository at https://doi.org/10.5281/zenodo.20394738.

References

1. Singer M, Deutschman CS, Seymour CW, Shankar-Hari M, Annane D, Bauer M, et al. The third international consensus definitions for sepsis and septic shock (Sepsis-3). JAMA. 2016;315:801–810. https://doi.org/10.1001/jama.2016.0287
2. Vincent JL, Sakr Y, Sprung CL, Ranieri VM, Reinhart K, Gerlach H, et al. Sepsis in European intensive care units: results of the SOAP study. Crit Care Med. 2006;34:344–353. https://doi.org/10.1097/01.CCM.0000194725.48928.3A
3. Rhodes A, Evans LE, Alhazzani W, Levy MM, Antonelli M, Ferrer R, et al. Surviving Sepsis Campaign: International Guidelines for Management of Sepsis and Septic Shock: 2016. Intensive Care Med. 2017;43:304–377. https://doi.org/10.1007/s00134-017-4683-6
4. Evans L, Rhodes A, Alhazzani W, Antonelli M, Coopersmith CM, French C, et al. Surviving Sepsis Campaign: International Guidelines for Management of Sepsis and Septic Shock 2021. Intensive Care Med. 2021;47:1181–1247. https://doi.org/10.1007/s00134-021-06506-y
5. Boyd JH, Forbes J, Nakada TA, Walley KR, Russell JA. Fluid resuscitation in septic shock: a positive fluid balance and elevated central venous pressure are associated with increased mortality. Crit Care Med. 2011;39:259–265. https://doi.org/10.1097/CCM.0b013e3181feeb15
6. Sakr Y, Rubatto Birri PN, Kotfis K, Nanchal R, Shah B, Kluge S, et al. Higher fluid balance increases the risk of death from sepsis: results from a large international audit. Crit Care Med. 2017;45:386–394. https://doi.org/10.1097/CCM.0000000000002189
7. Vaara ST, Korhonen AM, Kaukonen KM, Nisula S, Inkinen O, Hoppu S, et al. Fluid overload is associated with an increased risk for 90-day mortality in critically ill patients with renal replacement therapy: data from the prospective FINNAKI study. Crit Care. 2012;16:R197. https://doi.org/10.1186/cc11682
8. Hoste EAJ, Bagshaw SM, Bellomo R, Cely CM, Colman R, Cruz DN, et al. Epidemiology of acute kidney injury in critically ill patients: the multinational AKI-EPI study. Intensive Care Med. 2015;41:1411–1423. https://doi.org/10.1007/s00134-015-3934-7
9. Kidney Disease: Improving Global Outcomes (KDIGO) Acute Kidney Injury Work Group. KDIGO clinical practice guideline for acute kidney injury. Kidney Int Suppl. 2012;2:1–138. https://doi.org/10.1038/kisup.2012.1
10. National Heart, Lung, and Blood Institute Acute Respiratory Distress Syndrome (ARDS) Clinical Trials Network; Wiedemann HP, Wheeler AP, Bernard GR, Thompson BT, Hayden D, deBoisblanc B, et al. Comparison of two fluid-management strategies in acute lung injury. N Engl J Med. 2006;354:2564–2575. https://doi.org/10.1056/NEJMoa062200

11. Silversides JA, Major E, Ferguson AJ, Mann EE, McAuley DF, Marshall JC, et al. Conservative fluid management or deresuscitation for patients with sepsis or acute respiratory distress syndrome following the resuscitation phase of critical illness: a systematic review and meta-analysis. Intensive Care Med. 2017;43:155–170. https://doi.org/10.1007/s00134-016-4573-3
12. Hjortrup PB, Haase N, Bundgaard H, Thomsen SL, Winding R, Pettilä V, et al. Restricting volumes of resuscitation fluid in adults with septic shock after initial management: the CLASSIC randomised, parallel-group, multicentre feasibility trial. Intensive Care Med. 2016;42:1695–1705. https://doi.org/10.1007/s00134-016-4500-7
13. McCoy IE, Chertow GM, Chang TI. Patterns of diuretic use in the intensive care unit. PLoS One. 2019;14:e0217911. https://doi.org/10.1371/journal.pone.0217911
14. Brater DC. Diuretic therapy. N Engl J Med. 1998;339:387–395. https://doi.org/10.1056/NEJM199808063390607
15. Shankar SS, Brater DC. Loop diuretics: from the Na-K-2Cl transporter to clinical use. Am J Physiol Renal Physiol. 2003;284:F11-F21. https://doi.org/10.1152/ajprenal.00119.2002
16. Mehta RL, Pascual MT, Soroko S, Chertow GM; PICARD Study Group. Diuretics, mortality, and nonrecovery of renal function in acute renal failure. JAMA. 2002;288:2547–2553. https://doi.org/10.1001/jama.288.20.2547
17. Uchino S, Doig GS, Bellomo R, Morimatsu H, Morgera S, Schetz M, et al. Diuretics and mortality in acute renal failure. Crit Care Med. 2004;32:1669–1677. https://doi.org/10.1097/01.CCM.0000132892.51063.2F
18. Grams ME, Estrella MM, Coresh J, Brower RG, Liu KD; National Heart, Lung, and Blood Institute Acute Respiratory Distress Syndrome Network. Fluid balance, diuretic use, and mortality in acute kidney injury. Clin J Am Soc Nephrol. 2011;6:966–973. https://doi.org/10.2215/CJN.08781010
19. Cantarovich F, Rangoonwala B, Lorenz H, Verho M, Esnault VLM; High-Dose Furosemide in Acute Renal Failure Study Group. High-dose furosemide for established ARF: a prospective, randomized, double-blind, placebo-controlled, multicenter trial. Am J Kidney Dis. 2004;44:402–409. https://doi.org/10.1053/j.ajkd.2004.05.021
20. Bagshaw SM, Gibney RTN, Kruger P, Hassan I, McAlister FA, Bellomo R. The effect of low-dose furosemide in critically ill patients with early acute kidney injury: a pilot randomized blinded controlled trial (the SPARK study). J Crit Care. 2017;42:138–146. https://doi.org/10.1016/j.jcrc.2017.07.030
21. Bagshaw SM, Delaney A, Haase M, Ghali WA, Bellomo R. Loop diuretics in the management of acute renal failure: a systematic review and meta-analysis. Crit Care Resusc. 2007;9:60–68.
22. Hollenberg SM. Vasopressor support in septic shock. Chest. 2007;132:1678–1687. https://doi.org/10.1378/chest.07-0291
23. De Backer D, Biston P, Devriendt J, Madl C, Chochrad D, Aldecoa C, et al. Comparison of dopamine and norepinephrine in the treatment of shock. N Engl J Med. 2010;362:779–789. https://doi.org/10.1056/NEJMoa0907118

24. Bandak G, Sakhuja A, Andrijasevic NM, Gunderson TM, Gajic O, Kashani K. Use of diuretics in shock: temporal trends and clinical impacts in a propensity-matched cohort study. PLoS One. 2020;15:e0228274. https://doi.org/10.1371/journal.pone.0228274
25. Andrei S, Bahr PA, Nguyen M, Bouhemad B, Guinot PG. Prevalence of systemic venous congestion assessed by Venous Excess Ultrasound Grading System (VExUS) and association with acute kidney injury in a general ICU cohort: a prospective multicentric study. Crit Care. 2023;27:224. https://doi.org/10.1186/s13054-023-04524-4
26. Hernán MA, Robins JM. Using big data to emulate a target trial when a randomized trial is not available. Am J Epidemiol. 2016;183:758–764. https://doi.org/10.1093/aje/kwv254
27. Wager S, Athey S. Estimation and inference of heterogeneous treatment effects using random forests. J Am Stat Assoc. 2018;113:1228–1242. https://doi.org/10.1080/01621459.2017.1319839
28. Athey S, Tibshirani J, Wager S. Generalized random forests. Ann Stat. 2019;47:1148–1178. https://doi.org/10.1214/18-AOS1709
29. Athey S, Wager S. Policy learning with observational data. Econometrica. 2021;89:133–161. https://doi.org/10.3982/ECTA15732
30. 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. https://doi.org/10.1038/s41597-022-01899-x
31. Pollard TJ, Johnson AEW, Raffa JD, Celi LA, Mark RG, Badawi O. The eICU Collaborative Research Database, a freely available multi-center database for critical care research. Sci Data. 2018;5:180178. https://doi.org/10.1038/sdata.2018.178
32. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP; STROBE Initiative. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. Lancet. 2007;370:1453–1457. https://doi.org/10.1016/S0140-6736(07)61602-X
33. Benchimol EI, Smeeth L, Guttmann A, Harron K, Moher D, Petersen I, et al. The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) statement. PLoS Med. 2015;12:e1001885. https://doi.org/10.1371/journal.pmed.1001885
34. Khwaja A. KDIGO clinical practice guidelines for acute kidney injury. Nephron Clin Pract. 2012;120:c179-c184. https://doi.org/10.1159/000339789
35. Vincent JL, Moreno R, Takala J, Willatts S, De Mendonça A, Bruining H, et al. The SOFA (Sepsis-related Organ Failure Assessment) score to describe organ dysfunction/failure. Intensive Care Med. 1996;22:707–710. https://doi.org/10.1007/BF01709751
36. Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. J Chronic Dis. 1987;40:373–383. https://doi.org/10.1016/0021-9681(87)90171-8
37. Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. Biometrika. 1983;70:41–55. https://doi.org/10.1093/biomet/70.1.41

38. Austin PC. An introduction to propensity score methods for reducing the effects of confounding in observational studies. Multivariate Behav Res. 2011;46:399–424. https://doi.org/10.1080/00273171.2011.568786
39. Stuart EA. Matching methods for causal inference: a review and a look forward. Stat Sci. 2010;25:1–21. https://doi.org/10.1214/09-STS313
40. VanderWeele TJ, Ding P. Sensitivity analysis in observational research: introducing the E-value. Ann Intern Med. 2017;167:268–274. https://doi.org/10.7326/M16-2607
41. Robins JM, Rotnitzky A, Zhao LP. Estimation of regression coefficients when some regressors are not always observed. J Am Stat Assoc. 1994;89:846–866. https://doi.org/10.1080/01621459.1994.10476818
42. Ho DE, Imai K, King G, Stuart EA. MatchIt: nonparametric preprocessing for parametric causal inference. J Stat Softw. 2011;42:1–28. https://doi.org/10.18637/jss.v042.i08

Figures

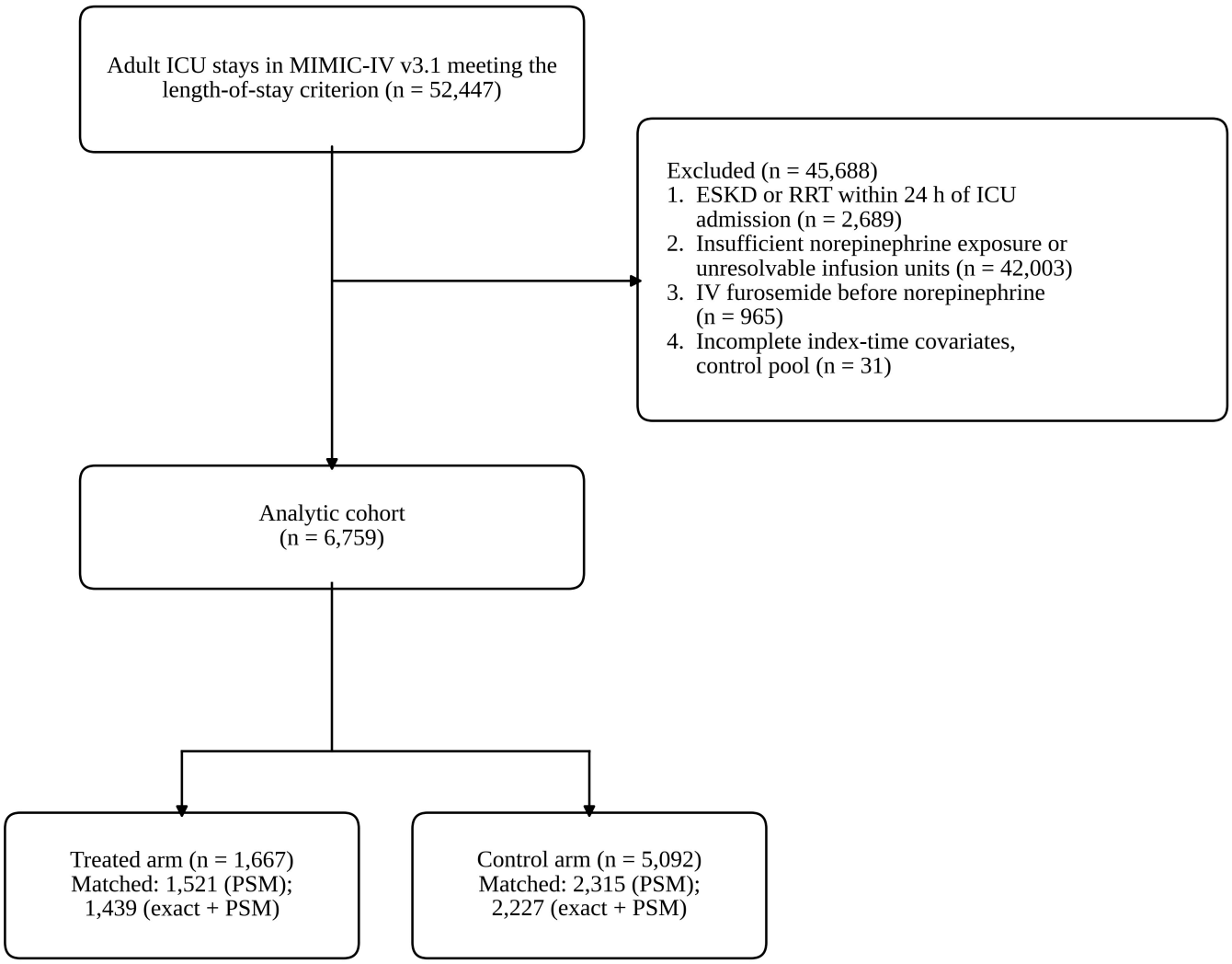


Figure 1

Cohort assembly flow diagram. Patient flow from the MIMIC-IV v3.1 source population to the analytic cohort and the two treatment-strategy arms. From all adult ICU stays meeting the length-of-stay criterion (n = 52,447), patients were excluded for end-stage kidney disease (ESKD) or renal replacement therapy (RRT) within the first 24 hours of ICU admission (n = 2,689), insufficient norepinephrine (NE) exposure or unresolvable infusion units (n = 42,003), intravenous furosemide before norepinephrine initiation (n = 965), and incomplete index-time covariates in the control pool (n = 31), yielding an analytic cohort of 6,759. Treated patients (n = 1,667) received intravenous furosemide during active norepinephrine infusion; control patients (n = 5,092), for whom one index time per patient was sampled at random, did not. Matched cohort sizes are shown for the primary propensity-score matching (PSM) and for the exact-plus-propensity sensitivity analysis (exact + PSM).

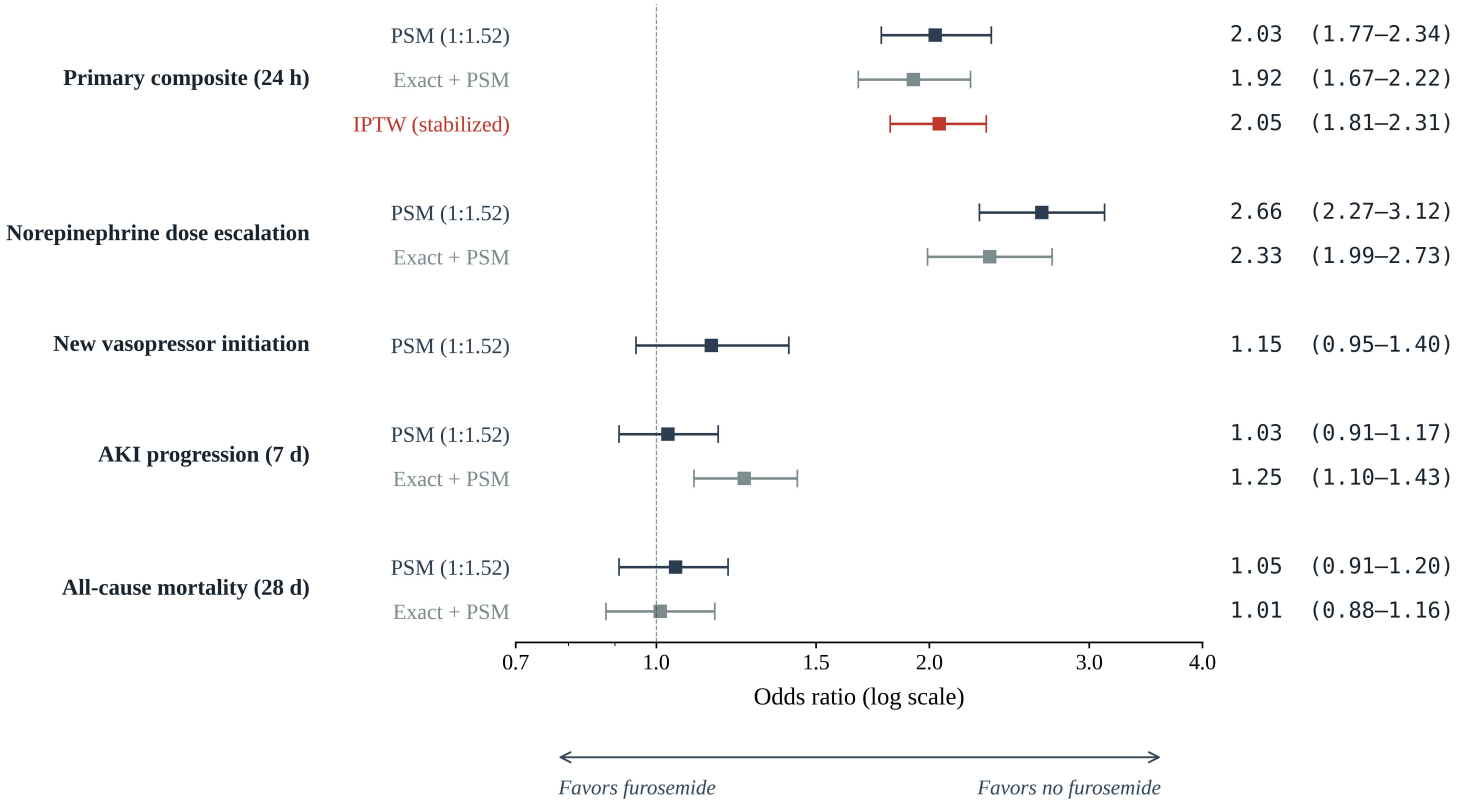


Figure 2

Forest plot summarizing the robustness of the treatment effect across three matching strategies. Each point represents the odds ratio (OR) for the indicated outcome with 95% confidence interval, on a log10 scale, comparing furosemide-exposed with matched control patients. Rows are grouped by outcome (primary composite, norepinephrine dose escalation, new vasopressor initiation, acute kidney injury progression at 7 days, all-cause mortality). The dashed vertical line at OR = 1 indicates the null. Sample sizes refer to the matched or weighted cohort for the corresponding analysis.

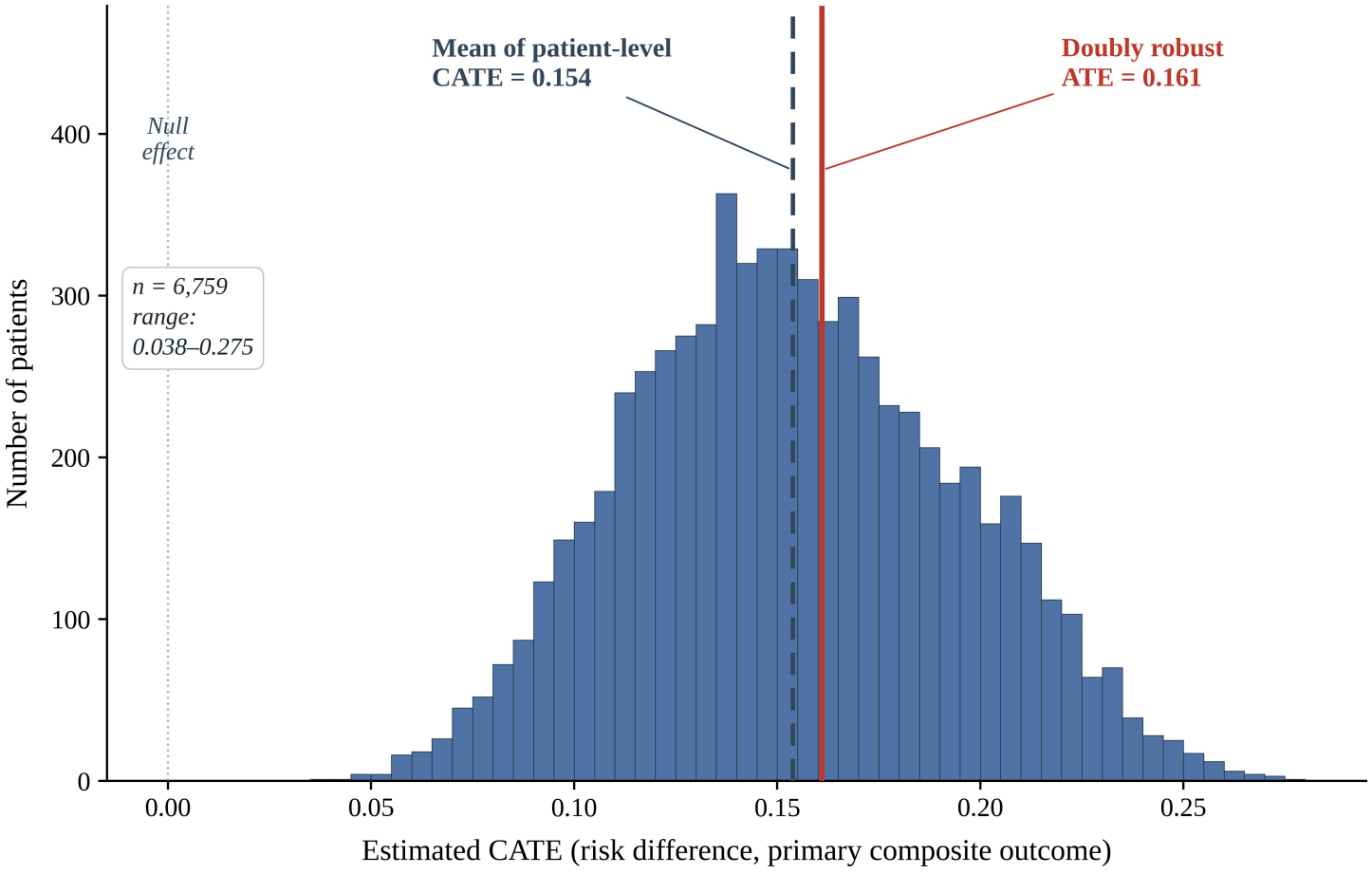


Figure 3

Distribution of conditional average treatment effect (CATE) predictions from the causal forest. Each bar is the count of patients with predicted CATE within a 0.005 bin, where CATE is the risk difference for the primary composite outcome (norepinephrine escalation or new vasopressor initiation within 24 hours). The solid line is the doubly robust population-average ATE (0.161); the dashed line is the arithmetic mean of patient-level CATE predictions (0.154); the dotted line at zero is the null. All 6,759 patient-level point estimates are positive (pointwise CIs not displayed).

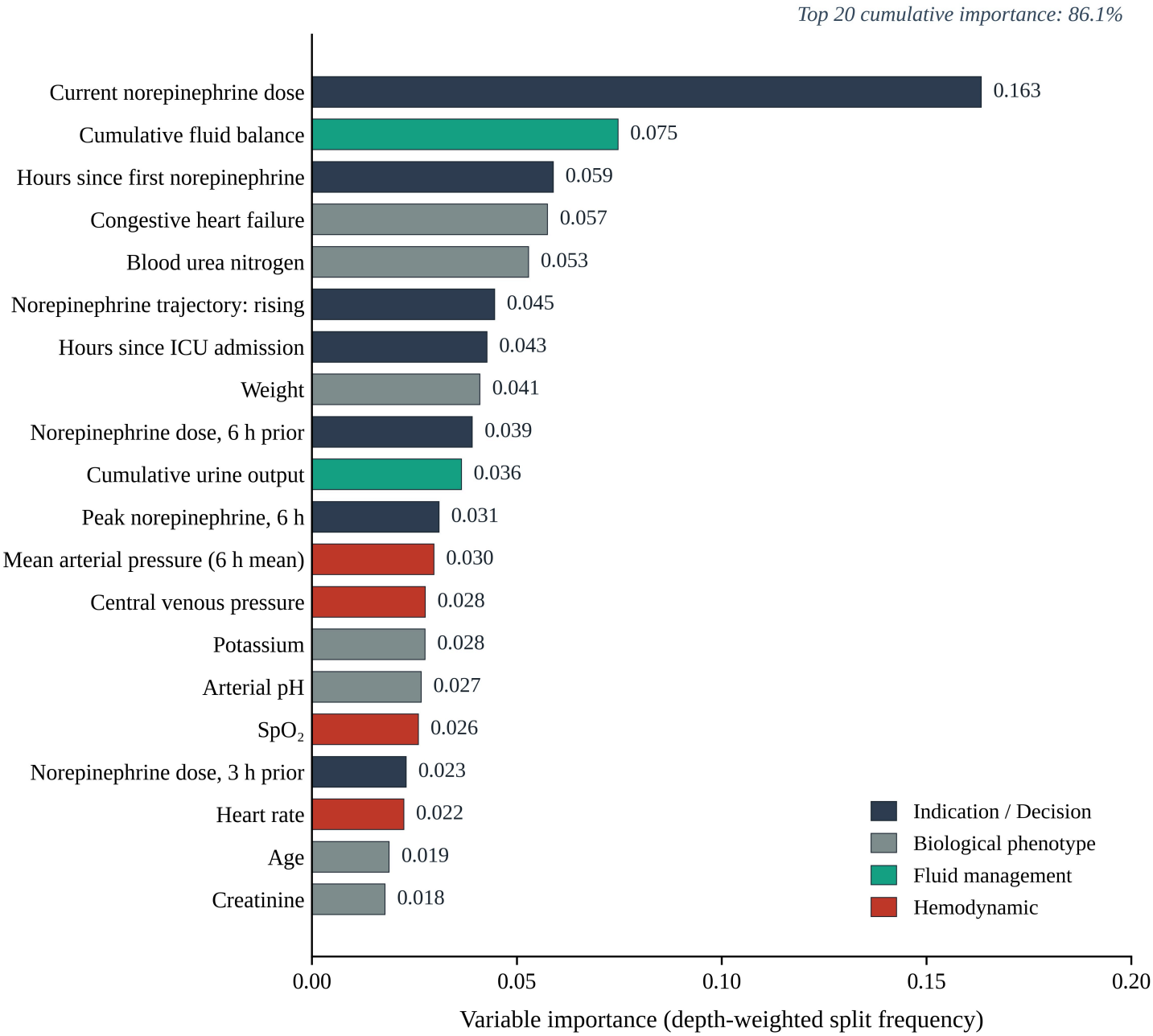


Figure 4

Top 20 covariates ranked by variable importance from the causal forest. Variable importance is the depth-weighted frequency with which each covariate was selected as a splitting variable across the 4,000 trees, scaled to sum to one. Current norepinephrine dose at index time is the dominant covariate (importance 0.163, more than twice the next), with the remaining variables spanning the indication/decision, biological-phenotype, fluid-management, and hemodynamic domains (color-coded by category). The displayed 20 variables account for approximately 86 percent of cumulative importance; the full top-25 ranking is provided in Supplementary Table S3.

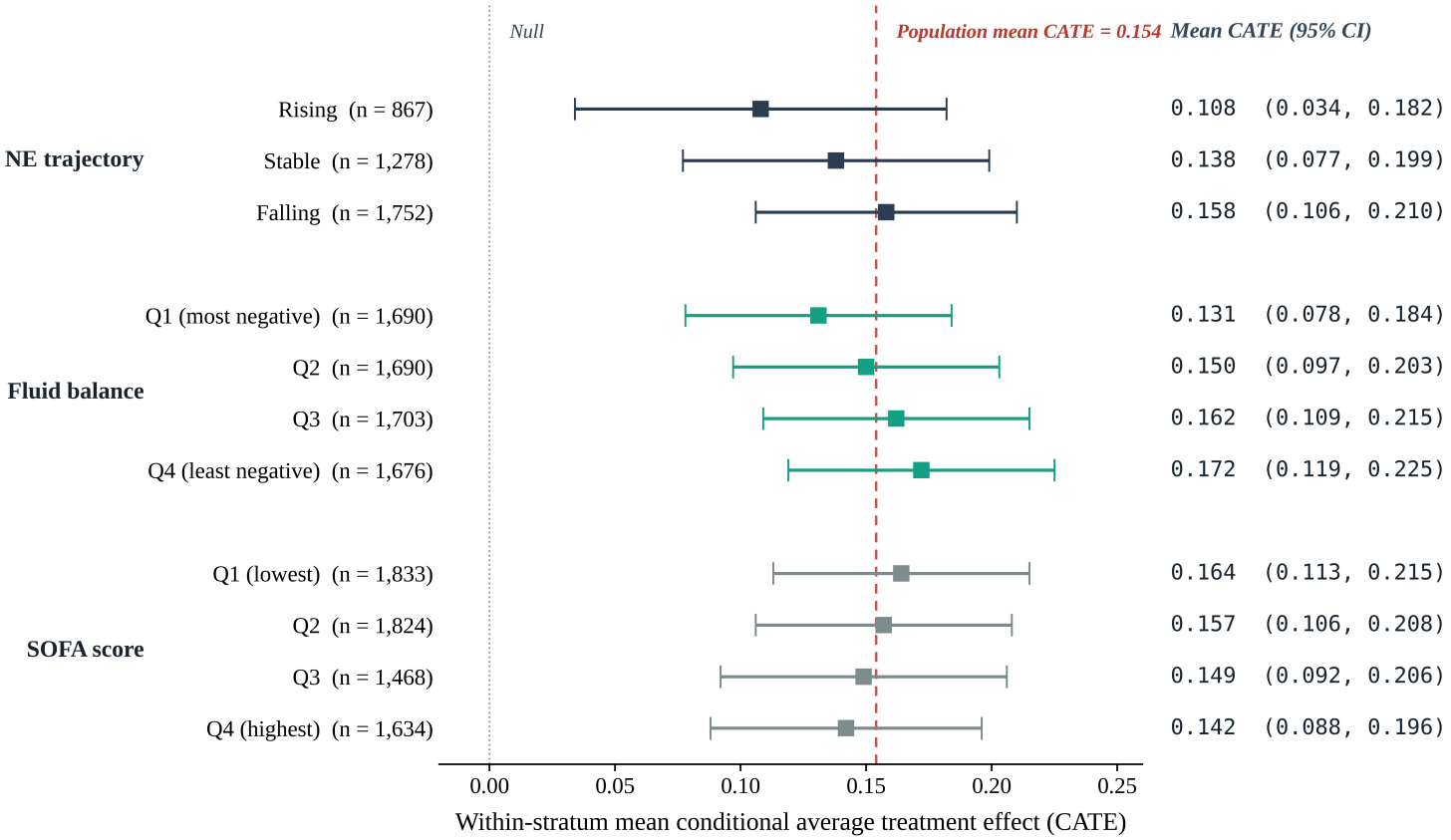


Figure 5

Within-stratum mean conditional average treatment effect (CATE) by prespecified subgroup, from the primary causal forest. Each point is the mean of patient-level CATE predictions within the stratum; horizontal lines are 95% CIs (doubly robust influence-function variance scaled to subgroup size; see Methods). Subgroups: norepinephrine trajectory class (rising, stable, falling), cumulative fluid balance quartile, and SOFA score quartile. The dotted line at zero is the null; the dashed line is the population-average mean CATE (0.154). All eleven CIs exclude zero.

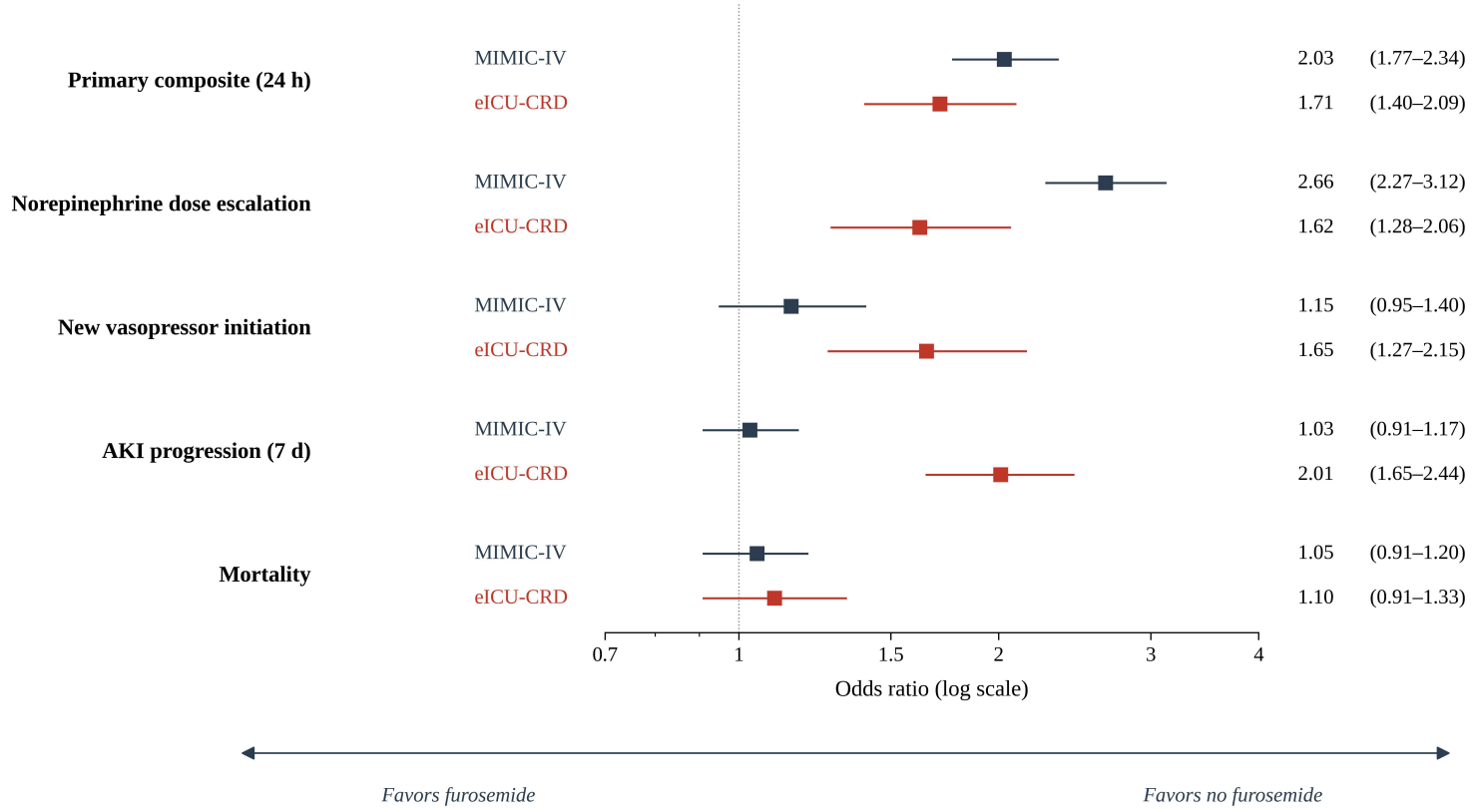


Figure 6

External validation forest plot comparing treatment effects in the MIMIC-IV derivation cohort (navy, n = 3,836 matched) and the eICU-CRD validation cohort (red, n = 2,307 matched) across the five prespecified outcomes; within each row the MIMIC-IV estimate is plotted above the eICU-CRD estimate. Horizontal lines are 95% CIs; the dotted line at OR = 1 is the null. Odds ratios (95% CI) are annotated at right. Direction of effect was concordant for all five outcomes, with attenuated magnitude in the multicenter validation cohort. AKI and mortality definitions differ between databases (see Table 4 footnotes).

Supplementary Files

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

- RECORDChecklist.docx
- Supplementaryfiles.docx
- FigS1.jpg
- FigS2.jpg
- FigS3.jpg