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.

# Fluid Resuscitation Volume After Early Echocardiography and Mortality in Sepsis: Target Trial Emulations With Strictly Pre-Exposure Confounder Measurement in MIMIC-IV

**Hui Ye**
Seventh People's Hospital of Shanghai University of Traditional Chinese Medicine
**Qian Zhang**
Shanghai Pudong New Area People's Hospital, Shanghai Jiao Tong University School of Medicine
**Song Chen**
Shanghai Pudong New Area People's Hospital, Shanghai Jiao Tong University School of Medicine
**Peiwu Li**
Shanghai Pudong New Area People's Hospital, Shanghai Jiao Tong University School of Medicine
**Jian Wan**
Shanghai Pudong New Area People's Hospital, Shanghai Jiao Tong University School of Medicine
**Yilong Hao**
haoyilong@126.com

Shanghai Pudong New Area People's Hospital, Shanghai Jiao Tong University School of Medicine

**Research Article**

**Keywords:** sepsis, echocardiography, fluid resuscitation, treatment effect heterogeneity, target trial emulation, overlap weighting, MIMIC-IV

**Posted Date:** November 1st, 2026

**DOI:** https://doi.org/10.21203/rs.3.rs-11247095/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

Whether echocardiographic phenotypes modify the association between early fluid resuscitation volume and mortality in sepsis is unknown. We emulated three target trials within a single cohort of Sepsis-3 patients with early transthoracic echocardiography (TTE), measuring every confounder strictly before the start of the exposure window.

# Methods

Using MIMIC-IV v3.1 (2008–2022), we identified Sepsis-3 stays with TTE within ± 24 h of ICU admission (first stay per patient; n = 7,819). Emulation 1 (primary) compared post-TTE 24-h fluid volume ≥ 30 vs < 30 mL/kg with a landmark at TTE + 24 h (n = 7,398). Emulation 2 (secondary) compared dobutamine new-users versus non-users within the LVEF < 40% stratum (n = 1,201; 100 new-users). Emulation 3 (secondary) compared vasopressin new-users versus non-users (n = 6,474; 360 new-users). All adjustment covariates—including SOFA components, vasopressor exposure with norepinephrine dose, Sepsis-3 septic shock status, fluid balance and urine output—were measured strictly before the echocardiogram. Overlap weighting was used throughout; propensity scores were re-estimated within phenotype strata. The outcome was 28-day all-cause mortality from each landmark. Effect modification was assessed with formal interaction tests on multiplicative and additive scales. Bias analyses included E-values with hazard-to-risk-ratio conversion, tipping-point grids, and negative control outcomes restricted to diagnoses from prior admissions.

# Results

Post-echocardiography fluid volume ≥ 30 mL/kg was associated with higher 28-day mortality (HR 1.74, 95% CI 1.54–1.97; weighted risk 38.7% vs 24.8%; risk difference + 13.9%, 95% CI 10.8–17.0). The association was similar across LVEF strata (< 40%: HR 1.63, 1.26–2.10, n = 1,361; 40–54%: 2.02, 1.47–2.79, n = 1,180; 55–69%: 1.88, 1.51–2.34, n = 2,752; ≥70% hyperdynamic: 1.39, 1.02–1.89, n = 896) and across RV strata (normal: 1.67, 1.42–1.97, n = 4,814; dysfunction: 1.77, 1.43–2.18, n = 1,979). No formal interaction test supported effect modification (LVEF: multiplicative p = 0.44, additive p = 0.56; RV: multiplicative p = 0.65, additive p = 0.16; LVEF modelled continuously: p = 0.82). A previously observed protective association in hyperdynamic hearts without cardiac disease was not confirmed when cardiac disease was ascertained from pre-admission diagnoses (HR 1.33, 0.90–1.97). The association persisted within pre-TTE SOFA tertiles and septic-shock strata, at 20 and 40 mL/kg thresholds, per 10 mL/kg as a continuous exposure (HR 1.14, 1.11–1.17), with a 6-h exposure window (HR 1.43, 1.27–1.60), and after excluding pre-ICU echocardiograms (HR 1.64, 1.44–1.87). Dobutamine initiation (Emulation 2) was imprecisely estimated (HR 1.57, 0.95–2.60); vasopressin initiation (Emulation 3) was strongly associated with mortality (HR 2.55, 2.13–3.06) without phenotype interaction (multiplicative p = 0.23).

The E-value was 2.60 (2.29 for the confidence bound); prior-admission negative control outcomes indicated residual bias acting, if anything, toward the null.

# Conclusions

After strict temporal alignment of confounder measurement, exposure and follow-up, higher fluid volume in the 24 h after early echocardiography was consistently associated with higher 28-day mortality in sepsis, but we found no statistically supported evidence that LVEF or RV phenotype modifies this association. These results do not support phenotype-guided fluid restriction derived from observational subgroup patterns; adequately powered randomised trials with protocolised echocardiography remain necessary.

## 1. Introduction

Sepsis remains a leading cause of ICU mortality, and its haemodynamic management is still guided largely by population-average evidence [1,2]. Randomised trials of restrictive versus liberal fluid strategies (CLASSIC, CLOVERS) and of perfusion-targeted resuscitation (ANDROMEDA-SHOCK) were collectively neutral [3–5], motivating the hypothesis that treatment effects vary across physiological subtypes that these trials did not measure [6]. Echocardiography is the most accessible bedside tool for capturing cardiac subtypes [7]: septic cardiomyopathy [8], hyperdynamic states [9], right ventricular (RV) dysfunction [10], and diastolic impairment [11] are each common and prognostically relevant. Whether echocardiographic phenotypes modify the effect of fluid resuscitation itself—treatment effect heterogeneity (HTE)—has not been systematically quantified.

Answering HTE questions observationally is methodologically hazardous. Therapies are assigned to sicker patients (confounding by indication), illness severity evolves while exposure accumulates, and covariates summarised over calendar windows can be measured after treatment has begun, so that adjustment risks conditioning on consequences of treatment. Target trial emulation provides a framework for aligning eligibility, treatment assignment (time zero), and follow-up [12]; overlap weighting can balance measured covariates where treatment groups overlap [13]; and new-user designs mitigate prevalent-user bias [14].

Here we emulated three target trials within a single cohort of Sepsis-3 patients with early TTE in MIMIC-IV [15]: Emulation 1 (primary analysis) examined post-echocardiography 24-h fluid volume across LVEF and RV phenotypes; Emulation 2 (secondary) examined dobutamine initiation within the LVEF < 40% phenotype; and Emulation 3 (secondary) examined vasopressin initiation with a phenotype interaction test. In response to methodological review, the present version differs from a conventional analysis in four ways: every severity covariate is measured strictly before the echocardiogram (time zero); the outcome is 28-day all-cause mortality counted from each landmark, without treating discharge as a competing event; comorbidities are ascertained from admissions before the index hospitalisation; and

all effect-modification claims are evaluated with formal interaction tests on both relative and absolute scales.

## 2. Methods

# 2.1 Data source and ethics

We used MIMIC-IV v3.1 (Beth Israel Deaconess Medical Center, 2008–2022), a de-identified single-centre critical care database available under a data use agreement after credentialed training [15]. The study protocol was approved by the Ethics Committee of Shanghai Pudong New Area People's Hospital (approval No. K82); the requirement for individual informed consent was waived because the data are de-identified. Reporting follows the STROBE statement, its RECORD extension for routinely collected health data, and the TARGET guideline for observational studies emulating a target trial [16–18].

# 2.2 Cohort

Of 94,458 adult ICU stays, we identified Sepsis-3 stays as suspected infection (body-fluid culture, excluding surveillance screens, followed by systemic antibiotics within 72 h, or antibiotics first with culture within 24 h) with the time of suspected infection within [− 48 h, + 24 h] of ICU admission and an acute SOFA increase ≥ 2 within the first 24 h, yielding 32,165 stays [1]. We retained stays with a transthoracic echocardiogram within ± 24 h of ICU admission (n = 8,410), then the first ICU stay per patient (n = 7,819; Fig. 1). Because echocardiography is ordered selectively, we compared the 8,410 included stays with the 23,755 Sepsis-3 stays without a qualifying echocardiogram (Table S1). Analyses were conducted in PostgreSQL 18 and Python 3.12 (pandas, scikit-learn, statsmodels, lifelines).

# 2.3 Echocardiographic phenotyping

Echocardiographic measurements were taken from the MIMIC-IV-ECHO v1.0.1 structured measurement table (PhysioNet), in which report findings are already standardised to controlled categories. LVEF was categorised as < 40%, 40–54%, 55–69%, and ≥ 70% (hyperdynamic). The ≥ 70% hyperdynamic threshold follows Paonessa et al. [9]; the remaining boundaries were chosen to separate guideline-defined reduced ejection fraction (< 40% [19]) from preserved ejection fraction, splitting the preserved range at the approximate cohort median; robustness to the ESC categorisation (< 40%, 40–49%, ≥ 50% [19]) and to LVEF modelled continuously was assessed (Section 2.10). RV function was mapped from the standardised rv_function field: "normal" (Nl RV function, low normal function, hyperdynamic) versus "dysfunction" (mild, moderate or severe global hypokinesis; RV function depressed; apical free-wall hypokinesis; basal RV hypokinesis with McConnell's sign); "RV not well seen", "cannot assess RV function" and blank entries were treated as missing (Table S2 lists the mapping of all 12 observed categories). The mapping is deterministic and covers 100% of observed categories; convergent validity was assessed against tricuspid annular plane systolic excursion (TAPSE) and tricuspid regurgitation gradient. Because basal RV hypokinesis (McConnell's sign) can indicate acute pulmonary embolism rather than sepsis-related RV dysfunction, affected stays were excluded in a sensitivity analysis. When a

patient had more than one TTE study within the eligibility window (464 of 7,819 stays, 5.9%), the first study was used. Studies performed up to 24 h before ICU admission (n = 1,345, 17.2%) were flagged, and post-admission studies were analysed separately in a sensitivity analysis. Inferior vena cava (IVC) diameter was available for only 20.2% of stays and is reported descriptively.

## 2.4 Study design: three target trial emulations within a single cohort

Three separate target trial emulations were performed in the same source cohort; they are not nested or sequential trials, and they share patients, so no cross-emulation comparisons are made. Table 2 summarises each emulation as a target trial protocol (eligibility, treatment strategies, assignment, time zero, grace/exposure window, start of follow-up, outcome, causal contrast, analysis plan), following the TARGET reporting guideline [18]. Figure S1 shows the temporal structure common to all emulations: ICU admission; echocardiogram (time zero); covariate ascertainment window (strictly before time zero); exposure window; landmark; and follow-up.

Table 2
Target trial protocol summary for the three emulations (TARGET guideline [18]).

| Component | Emulation 1 (primary): fluid volume | Emulation 2: dobutamine | Emulation 3: vasopressin |
|---|---|---|---|
| Eligibility | Sepsis-3, first ICU stay, TTE within ± 24 h of ICU admission, alive & hospitalised at landmark, body weight available (n = 7,398) | Emulation 1 eligibility + LVEF < 40% (n = 1,201) | Sepsis-3, first ICU stay, TTE within ± 24 h, alive & hospitalised at landmark (n = 6,474) |
| Treatment strategies | ≥ 30 vs < 30 mL/kg crystalloid+colloid in 24 h after TTE | Dobutamine initiation vs no dobutamine within (TTE, TTE + 48 h] | Vasopressin initiation vs no vasopressin within (TTE, TTE + 48 h] |
| Assignment procedure | Emulated randomisation via overlap weighting on pre-TTE propensity score | Same; + fluid-volume covariates; parsimonious penalised model as sensitivity | Same |
| Time zero | First qualifying echocardiogram | First qualifying echocardiogram | First qualifying echocardiogram |
| Landmark / start of follow-up | TTE + 24 h | TTE + 48 h | TTE + 48 h |
| Follow-up | 28 days from landmark, all-cause mortality | 28 days from landmark | 28 days from landmark |
| Causal contrast | Observational analogue of ITT (assignment by achieved exposure at landmark) | New-user ITT analogue | New-user ITT analogue |
| Effect modification | LVEF (4 strata), RV (2), SOFA tertiles, septic shock; formal interactions on multiplicative & additive scales | None (precision) | Hyperdynamic LVEF × treatment interaction (measured LVEF only) |
| Analysis plan | Weighted cause-specific Cox; weighted Kaplan–Meier risks; RD bootstrap (1,000 reps, PS re-estimated per replicate) | Same | Same + product-term and bootstrap RD-difference interaction tests |

**Emulation 1 (primary).** Time zero was the first qualifying echocardiogram. Exposure was the total crystalloid plus colloid volume administered in the ICU during the 24 h after time zero, indexed to body weight, with infusion events prorated by window overlap; volumes include maintenance fluids and drug carriers, because inputevents records total administered volume (item codes in Table S3). The contrast was ≥ 30 vs < 30 mL/kg. Because 30 mL/kg over 24 h (≈ 1.25 mL/kg/h) spans maintenance as well as resuscitation volumes and is applied after, not during, initial resuscitation, we treat the threshold as a pragmatic upper-quintile split of cumulative early volume rather than a direct operationalisation of the

Surviving Sepsis Campaign 3-h bolus recommendation [2]; the exposure was therefore also modelled continuously and at 20 and 40 mL/kg thresholds (Section 2.10), and the full exposure distribution is reported. To avoid immortal time bias, a landmark was set at TTE + 24 h; stays that died (hospital death) or were discharged alive from hospital before the landmark (n = 421; 270 deaths, 29 live discharges, 127 missing body weight; categories overlap) were excluded, leaving 7,398 stays. Estimates therefore apply to patients surviving and remaining hospitalised to the landmark.

**Emulation 2 (secondary).** Within the LVEF < 40% stratum, dobutamine new-users (first administration in (TTE, TTE + 48 h]; pre-TTE users and late initiators excluded) were compared with non-users, landmark at TTE + 48 h (n = 1,201; 100 new-users; Figure S2). A class-level sensitivity pooled dobutamine and milrinone initiators (n = 132). Because concurrent fluid administration modifies and confounds the response to dobutamine, the propensity model for this emulation additionally included fluid volume administered before the echocardiogram, the Emulation 1 exposure (post-TTE 24-h volume per kg), and fluid volume administered between the echocardiogram and inotrope initiation.

**Emulation 3 (secondary).** Vasopressin new-users (TTE, TTE + 48 h]; n = 360 after excluding prevalent and late users) were compared with non-users (n = 6,114), landmark at TTE + 48 h, with multiplicative and additive interaction by the hyperdynamic phenotype (Figure S3). A stratum-restricted design was not feasible (56 treated stays in the hyperdynamic stratum). The interaction analysis was restricted to stays with measured LVEF (n = 5,413) to avoid classifying missing LVEF as non-hyperdynamic.

# 2.5 Outcome

The primary outcome for all emulations was 28-day all-cause mortality counted from each landmark, with no competing event. MIMIC-IV provides out-of-hospital dates of death from hospital and state death records, so deaths after discharge are counted; a patient discharged alive on day 10 who died on day 20 is classified as an event on day 20. Follow-up is defined from each landmark (not from ICU admission), giving a uniform 28-day horizon in all emulations. The previously reported analysis—26-day in-hospital mortality with discharge alive as a competing event—is retained as a sensitivity analysis for comparability (Table S4); cause-specific models for that analysis follow recommended competing-risks reporting [20].

# 2.6 Confounding control

For each emulation we fit a logistic propensity model. Every covariate was measured strictly before time zero (the echocardiogram): laboratory values from hospital admission to time zero, and charted ICU data, organ support and medications from ICU admission to time zero. The covariate set comprised demographics (age, sex, race, anchor year group); nine comorbidities ascertained from ICD codes of admissions before the index hospitalisation (heart failure, arrhythmia, ischaemic heart disease, chronic kidney disease, chronic liver disease, diabetes, COPD, hypertension, malignancy); pre-TTE SOFA components (all six, scored from pre-TTE physiology); peak pre-TTE lactate; minimum pre-TTE mean arterial pressure; mean pre-TTE heart rate; pre-TTE mechanical ventilation, renal replacement therapy and systemic corticosteroids; any pre-TTE vasopressor and the maximum pre-TTE norepinephrine-

equivalent rate; Sepsis-3 septic shock status (pre-TTE vasopressor with lactate > 2 mmol/L); fluid volume administered before the echocardiogram; pre-TTE urine output rate; pre-TTE diuretic use; infection source (blood, respiratory, urinary, other); surgical versus medical admission; and pre-ICU location (emergency department, ward, operating room/PACU, other ICU, transfer-in, other). Continuous variables were median-imputed with a missingness indicator for each covariate with missing values; multiple imputation by chained equations (5 imputations, pooled by Rubin's rules) was performed as a sensitivity analysis. Per-covariate missingness is reported in Table S5.

Overlap weights (treated × (1 − ps); controls × ps) target the average treatment effect in the overlap population [13]. Balance was assessed by absolute standardised mean differences (SMDs) and variance ratios for all model covariates and for severity summaries not in the model (24-h SOFA total, tricuspid regurgitation gradient); weighted covariate means—not only SMDs—are reported to describe the overlap population clinically (Table 1, Table S6). Propensity score distributions by treatment group and post-weighting effective sample sizes (ESS) are reported for every emulation (Figure S4). No trimming was applied to overlap weights; inverse-probability-of-treatment weights (ATE estimand, propensity scores clipped to [0.01, 0.99]) were used as a sensitivity analysis. Propensity scores and weights were re-estimated within each LVEF and RV stratum for stratum-specific estimates. For Emulation 2 (100 treated), a parsimonious L1-penalised propensity model was fitted in addition to the full model. Weighted Cox models used robust sandwich standard errors; proportional-hazards diagnostics (Schoenfeld residuals) are reported, and because the treatment hazard was non-proportional, period-specific hazard ratios (days 0–7 and 8–28) are reported alongside absolute risk differences, which do not require proportionality. Risk differences used subject-level non-parametric bootstrap percentile intervals with 1,008 replicates, re-estimating the propensity model and weights within each replicate.

Table 1

Baseline characteristics of the Emulation 1 landmark cohort before and after overlap weighting (n = 7,398). All severity variables measured strictly before the echocardiogram (time zero); comorbidities from pre-admission diagnoses. Weighted values describe the overlap population. (Full version in Table S6.)

| Characteristic | ≥ 30 mL/kg (n = 1,357) | < 30 mL/kg (n = 6,041) | SMD unadj. | SMD after OW |
|---|---|---|---|---|
| Age, yr, median (IQR) | 65.0 (52.0–77.0) | 69.0 (59.0–79.0) | −0.278 | −0.002 |
| Male, % | 53.6 | 56.5 | −0.057 | −0.003 |
| Weight, kg, median (IQR) | 73.5 (61.2–88.0) | 80.8 (67.7–97.2) | −0.350 | 0.007 |
| Pre-TTE SOFA total, median (IQR) | 6.0 (3.0–9.0) | 4.0 (1.0–6.0) | 0.587 | 0.000 |
| Lactate max pre-TTE, mmol/L | 3.2 (1.9–5.7) | 2.0 (1.4–3.2) | 0.485 | 0.001 |
| MAP min pre-TTE, mmHg | 57.0 (49.0–65.0) | 61.0 (54.0–70.0) | −0.302 | 0.001 |
| Heart rate mean pre-TTE, /min | 94.8 (81.2–109.0) | 87.5 (74.2–100.3) | 0.366 | 0.001 |
| Norepinephrine max pre-TTE, µg/kg/min | 0.3 (0.2–0.5) | 0.2 (0.1–0.3) | 0.183 | 0.006 |
| Mechanical ventilation pre-TTE, % | 55.6 | 24.8 | 0.661 | 0.001 |
| RRT pre-TTE, % | 3.7 | 1.1 | 0.172 | 0.001 |
| Systemic corticosteroids pre-TTE, % | 18.1 | 16.4 | 0.046 | 0.001 |
| Vasopressor pre-TTE, % | 56.2 | 24.7 | 0.677 | −0.003 |
| Septic shock pre-TTE (Sepsis-3), % | 40.2 | 12.2 | 0.671 | −0.001 |
| Diuretic pre-TTE, % | 16.0 | 29.8 | −0.334 | 0.004 |
| Surgical admission, % | 22.3 | 13.3 | 0.235 | 0.002 |
| Fluid before TTE, mL, median (IQR) | 1324.2 (308.4–2722.9) | 212.6 (0.0–873.7) | 0.763 | 0.001 |

IQR, interquartile range; SMD, standardised mean difference; OW, overlap weighting; SOFA, Sequential Organ Failure Assessment; MAP, mean arterial pressure; RRT, renal replacement therapy; CKD, chronic kidney disease; COPD, chronic obstructive pulmonary disease; TTE, transthoracic echocardiography. Medians are unweighted; weighted summaries of the overlap population are given in Table S6. Missingness per covariate is reported in Table S5.

| Characteristic | ≥ 30 mL/kg (n = 1,357) | < 30 mL/kg (n = 6,041) | SMD unadj. | SMD after OW |
|---|---|---|---|---|
| Heart failure (prior admission), % | 14.2 | 20.0 | −0.154 | −0.002 |
| Arrhythmia (prior admission), % | 15.2 | 19.5 | −0.114 | −0.003 |
| Ischaemic heart disease (prior admission), % | 14.0 | 20.1 | −0.164 | −0.007 |
| CKD (prior admission), % | 12.2 | 17.2 | −0.142 | −0.010 |
| Chronic liver disease (prior admission), % | 5.4 | 5.3 | 0.001 | 0.001 |
| Diabetes (prior admission), % | 16.4 | 20.3 | −0.102 | 0.001 |
| COPD (prior admission), % | 6.7 | 7.8 | −0.043 | −0.002 |
| Hypertension (prior admission), % | 22.7 | 27.4 | −0.109 | −0.006 |
| Malignancy (prior admission), % | 12.2 | 12.8 | −0.017 | −0.001 |
| 24-h SOFA total (not in PS model), median (IQR) | 11.0 (8.0–14.0) | 8.0 (6.0–11.0) | 0.635 | 0.154 |
| Post-TTE 24-h fluid, mL/kg, median (IQR) | 45.1 (36.1–62.2) | 7.1 (1.1–15.0) | — | — |

IQR, interquartile range; SMD, standardised mean difference; OW, overlap weighting; SOFA, Sequential Organ Failure Assessment; MAP, mean arterial pressure; RRT, renal replacement therapy; CKD, chronic kidney disease; COPD, chronic obstructive pulmonary disease; TTE, transthoracic echocardiography. Medians are unweighted; weighted summaries of the overlap population are given in Table S6. Missingness per covariate is reported in Table S5.

## 2.7 Effect modification

Effect modification by LVEF (four strata), RV function (two strata), pre-TTE SOFA tertiles and septic shock status was assessed with formal interaction tests on two scales: multiplicatively, by Wald tests of treatment × stratum product terms in weighted Cox models (global test across the four LVEF strata), and additively, by heterogeneity tests of stratum-specific risk differences using their bootstrap covariance. LVEF was additionally modelled as a continuous covariate with a treatment × LVEF product term. To control multiplicity across the six phenotype-specific hazard ratios, Benjamini–Hochberg adjusted p-values are reported. The credibility of any subgroup pattern was assessed against ICEMAN criteria [21] (Table S7). Differences between subgroup estimates, or between their statistical significances, are not interpreted as effect modification.

## 2.8 Bias quantification

We computed E-values for the primary contrast [22]. Because 28-day mortality is common (≈ 24%), hazard ratios were first converted to approximate risk ratios using the weighted control risk [23]; directly

computed risk-ratio E-values are also reported. Tipping-point grids show the joint strength (prevalence imbalance × outcome association) an unmeasured binary confounder would need to fully explain the overall association. Negative control outcomes (NCOs) were pre-existing benign conditions that early fluid cannot plausibly cause (benign prostatic hyperplasia, cataract, osteoarthritis, non-toxic goitre, benign skin neoplasms); to avoid conflating coding intensity during the index admission with pre-existing disease, NCOs were ascertained only from diagnoses recorded in admissions before the index hospitalisation, and their weighted associations with treatment were estimated with the same propensity model.

## 2.9 Sensitivity analyses

Sensitivity analyses comprised: (i) alternative fluid thresholds (20 and 40 mL/kg, analysed with the same overlap-weighted approach as the primary analysis); (ii) continuous exposure using generalised-propensity-score (GPS) stabilised weights, per 10 mL/kg and with a four-knot restricted cubic spline; (iii) a 6-h exposure window (≥ 7.5 mL/kg in 6 h, landmark TTE + 6 h) to shrink within-window evolution of illness severity; (iv) ESC LVEF categorisation and continuous LVEF interaction; (v) exclusion of McConnell's sign; (vi) exclusion of echocardiograms performed before ICU admission; (vii) exclusion of stays with pre-admission heart failure, ischaemic heart disease or arrhythmia, with propensity scores and weights re-estimated in the restricted cohort; (viii) multiple imputation; and (ix) the legacy 26-day in-hospital competing-risks analysis.

## 2.10 Protocol, registration and code

The analyses were conducted under an internal statistical analysis plan that was not prospectively registered; no claims of prospective registration are made. The Emulation 3 interaction design replaced an infeasible stratum-restricted design after a feasibility check and before phenotype-specific results were examined. All extraction and analysis code is publicly archived at https://doi.org/10.5281/zenodo.23113388 (Zenodo); MIMIC-IV permits code sharing under its data use agreement.

# 3. Results

## 3.1 Cohort and selection into early echocardiography

The Emulation 1 landmark cohort comprised 7,398 stays (Fig. 1). Median age was 69 years, 56.0% were male, the qualifying echocardiogram was performed a median of 10.6 h (IQR 3.1–16.1) after ICU admission, and median pre-TTE SOFA was 4 (IQR 1–7); 28-day mortality from the landmark was 23.9% (1,766/7,398). Stays with a qualifying echocardiogram were more severely ill than the 23,755 Sepsis-3 stays without one (median 24-h SOFA 9 vs 8; mechanical ventilation 44.9% vs 40.9%; RRT 4.4% vs 2.5%; hospital mortality 22.2% vs 12.8%; cardiac care unit admission 22.4% vs 6.2%; Table S1), confirming that early echocardiography is ordered selectively.

High-volume resuscitation (≥ 30 mL/kg; n = 1,357, 18.3%) was strongly patterned by pre-exposure illness severity—pre-TTE mechanical ventilation 55.6% vs 24.8%, pre-TTE vasopressor 56.2% vs 24.7%, pre-TTE septic shock 40.2% vs 12.2% (Table 1). Overlap weighting reduced the maximum absolute SMD from 0.763 to ≤ 0.010 across all model covariates; weighted variance ratios lay within 0.5–2.0 for 54 of 55 model covariates (the exception, pre-TTE norepinephrine dose at 0.26, reflects down-weighting of the highest-dose treated stays), and the 24-h SOFA total, a severity summary deliberately excluded from the model, retained an SMD of 0.154 after weighting (Table S6). Post-weighting effective sample sizes were 1,130 (treated) and 2,858 (control); propensity score distributions overlapped across the full range (Figure S4).

Baseline severity also differed across echocardiographic phenotypes: LVEF < 40% and RV dysfunction strata had higher pre-TTE SOFA, more vasopressor use and higher 28-day mortality than other strata (Table S8). Phenotype-by-severity confounding is therefore possible a priori, and stratum-specific estimates must be read against the severity gradient within each phenotype.

## 3.2 Emulation 1: overall association

Post-echocardiography fluid volume ≥ 30 mL/kg was associated with higher 28-day mortality (HR 1.74, 95% CI 1.54–1.97; weighted risk 38.7% vs 24.8%; RD + 13.9%, 95% CI 10.8–17.0; Table 3, Fig. 2). The treatment hazard was non-proportional (Schoenfeld p < 0.001): period-specific HRs were 1.91 (1.63–2.24) for days 0–7 and 1.53 (1.27–1.85) for days 8–28; absolute risk differences are unaffected by this violation. For comparison, the legacy covariate set (first-24-h variables measured partly after exposure onset) yielded HR 1.45 (1.28–1.64) for the same outcome, indicating that temporal misalignment of covariates had biased the earlier estimate toward the null. The ATE sensitivity with inverse-probability weights gave HR 2.05 (1.75–2.39; RD + 17.8%).

Table 3

Overlap-weighted associations between post-echocardiography fluid volume (≥ 30 vs < 30 mL/kg) and 28-day all-cause mortality from landmark, overall and by stratum (Emulation 1, n = 7,398). Propensity scores re-estimated within phenotype strata.

| Stratum | n (treated) | Deaths | HR (95% CI) | Risk ≥ 30, % | Risk < 30, % | RD, % (95% CI) |
|---|---|---|---|---|---|---|
| Overall | 7,398 (1,357) | 1,766 | 1.74 (1.54–1.97) | 38.7 | 24.8 | + 13.9 (10.8 to 17.0) |
| LVEF < 40% | 1,361 (279) | 386 | 1.63 (1.26–2.10) | 45.0 | 31.4 | + 13.6 (5.4 to 21.1) |
| LVEF 40–54% | 1,180 (231) | 268 | 2.02 (1.47–2.79) | 38.2 | 21.1 | + 17.2 (9.6 to 24.7) |
| LVEF 55–69% | 2,752 (490) | 532 | 1.88 (1.51–2.34) | 32.8 | 19.2 | + 13.6 (9.0 to 17.8) |
| LVEF ≥ 70% (hyperdynamic) | 896 (207) | 267 | 1.39 (1.02–1.89) | 41.2 | 32.5 | + 8.7 (0.8 to 18.0) |
| RV normal | 4,814 (837) | 998 | 1.67 (1.42–1.97) | 34.0 | 21.8 | + 11.8 (8.1 to 15.6) |
| RV dysfunction | 1,979 (409) | 639 | 1.77 (1.43–2.18) | 49.6 | 32.1 | + 16.9 (10.5 to 22.8) |
| Pre-TTE SOFA tertile 1 (0–2) | 2,660 (306) | 379 | 2.45 (1.91–3.13) | 32.5 | 14.7 | + 17.9 (11.8 to 22.7) |
| Pre-TTE SOFA tertile 2 (3–6) | 2,758 (416) | 612 | 1.87 (1.50–2.33) | 34.6 | 20.1 | + 14.5 (9.1 to 19.8) |
| Pre-TTE SOFA tertile 3 (7–20) | 1,980 (635) | 775 | 1.45 (1.22–1.73) | 47.9 | 36.6 | + 11.3 (5.5 to 16.7) |
| No septic shock pre-TTE | 6,113 (811) | 1,269 | 1.89 (1.63–2.19) | 35.4 | 20.5 | + 14.9 (10.9 to 18.2) |

Interaction tests: LVEF multiplicative $\chi^2(3) = 2.68$, p = 0.44; LVEF additive Q = 2.05, p = 0.56; RV multiplicative p = 0.65; RV additive p = 0.16; continuous LVEF × treatment p = 0.82; SOFA additive p = 0.32; septic shock additive p = 0.58. HR, hazard ratio from weighted Cox models with robust standard errors; risk, weighted Kaplan–Meier mortality at 28 days; RD, risk difference (bootstrap percentile CI, 1,008 replicates, propensity scores re-estimated within each replicate).

| Stratum | n (treated) | Deaths | HR (95% CI) | Risk ≥ 30, % | Risk < 30, % | RD, % (95% CI) |
|---|---|---|---|---|---|---|
| Septic shock pre-TTE | 1,285 (546) | 497 | 1.53 (1.24–1.88) | 47.8 | 34.4 | + 13.4 (5.8 to 19.1) |

Interaction tests: LVEF multiplicative $\chi^2(3) = 2.68$, p = 0.44; LVEF additive Q = 2.05, p = 0.56; RV multiplicative p = 0.65; RV additive p = 0.16; continuous LVEF × treatment p = 0.82; SOFA additive p = 0.32; septic shock additive p = 0.58. HR, hazard ratio from weighted Cox models with robust standard errors; risk, weighted Kaplan–Meier mortality at 28 days; RD, risk difference (bootstrap percentile CI, 1,008 replicates, propensity scores re-estimated within each replicate).

# 3.3 Phenotype-stratified estimates and formal interaction tests

Stratum-specific associations were positive in every phenotype stratum (Table 3, Fig. 3): LVEF < 40% HR 1.63 (1.26–2.10), RD + 13.6% (5.4–21.1); LVEF 40–54% HR 2.02 (1.47–2.79), RD + 17.2% (9.6–24.7); LVEF 55–69% HR 1.88 (1.51–2.34), RD + 13.6% (9.0–17.8); LVEF ≥ 70% (hyperdynamic) HR 1.39 (1.02–1.89), RD + 8.7% (0.8–18.0); RV normal HR 1.67 (1.42–1.97), RD + 11.8% (8.1–15.6); RV dysfunction HR 1.77 (1.43–2.18), RD + 16.9% (10.5–22.8). Propensity scores were re-estimated within each stratum; within-stratum balance is reported in Table S6. All six stratum-specific hazard ratios remained nominally significant after Benjamini–Hochberg adjustment (hyperdynamic adjusted p = 0.037; all others < 0.001).

No formal interaction test supported effect modification (Table 3, Fig. 3): LVEF multiplicative global test $\chi^2(3) = 2.68$, p = 0.44; LVEF additive heterogeneity Q = 2.05, p = 0.56; RV multiplicative p = 0.65; RV additive difference + 5.1%, p = 0.16; continuous LVEF × treatment p = 0.82; ESC categorisation interaction p = 0.85. The ICEMAN assessment rated the credibility of effect modification by echocardiographic phenotype as low (Table S7): interaction tests are null on both scales, there is no a priori direction from randomised evidence, and subgroup contrasts were not the basis of a pre-registered hypothesis.

# 3.4 Severity stratification

The association persisted within every baseline-severity stratum (Table 3): pre-TTE SOFA tertiles T1 HR 2.45 (1.91–3.13), T2 HR 1.87 (1.50–2.33), T3 HR 1.45 (1.22–1.73) (additive heterogeneity p = 0.32); septic shock absent HR 1.89 (1.63–2.19), present HR 1.53 (1.24–1.88) (additive p = 0.58). The attenuation of the relative association with increasing baseline severity—while absolute risk differences remained + 11 to + 17% across strata—is the expected pattern when part of the crude association reflects confounding by indication; within-stratum persistence indicates the association is not generated by severity differences between the exposure groups alone.

# 3.5 Sensitivity analyses

All sensitivity analyses were directionally consistent (Table 4). Weighted analyses at 20 and 40 mL/kg thresholds gave HR 1.67 (1.49–1.87; RD + 11.5%, 8.9–14.0; n = 2,198 treated) and HR 1.65 (1.44–1.90;

RD + 12.8%, 9.1–16.5; n = 855 treated). As a continuous exposure with GPS-stabilised weights, each additional 10 mL/kg carried HR 1.14 (1.11–1.17); the restricted-cubic-spline curve rose monotonically across the observed distribution (median 10.5 mL/kg; IQR 2.4–23.7; 95th percentile 60.2; Figure S5). With a 6-h exposure window (≥ 7.5 mL/kg; n = 7,610; 1,621 treated), the association was HR 1.43 (1.27–1.60; RD + 8.7%). Excluding echocardiograms performed before ICU admission (1,299 of 7,398 landmark stays) gave HR 1.64 (1.44–1.87). Excluding McConnell's sign (51 stays) left the RV-dysfunction estimate unchanged (HR 1.72, 1.39–2.12). Multiple imputation gave a pooled HR of 1.91 (1.69–2.15).

Table 4
Emulations 2 and 3 (overlap-weighted; 28-day all-cause mortality from landmark at TTE + 48 h).

| Analysis | n (treated) | Deaths | HR (95% CI) | RD, % (95% CI) |
|---|---|---|---|---|
| Emulation 2: dobutamine new-user, LVEF < 40% | 1,201 (100) | 291 | 1.57 (0.95–2.60) | + 10.8 (− 1.7 to 23.2) |
| Emulation 2: penalised (L1) propensity model | 1,201 (100) | 291 | 1.55 (0.98–2.47) | + 10.3 (—) |
| Emulation 2: any inotrope new-user (sensitivity) | 1,201 (132) | 291 | 1.12 (0.70–1.79) | + 2.4 (—) |
| Emulation 3: vasopressin new-user, overall | 6,474 (360) | 1,287 | 2.55 (2.13–3.06) | + 23.9 (18.6 to 28.7) |
| Emulation 3: hyperdynamic (measured LVEF ≥ 70%) | 777 (56) | 192 | 3.30 (2.16–5.04) | + 36.5 (—) |
| Emulation 3: non-hyperdynamic (measured LVEF < 70%) | 4,636 (231) | 861 | 2.42 (1.91–3.06) | + 21.6 (—) |
| Emulation 3: treatment × hyperdynamic interaction (measured LVEF, n = 5,413) | — | — | 1.34 (0.83–2.17), p = 0.23 | + 14.9, p = 0.065 |
| Supplementary figures: Figure S1 temporal structure; Figures S2–S3 flow diagrams for Emulations 2 and 3; Figure S4 propensity score distributions; Figure S5 continuous-exposure (GPS-weighted) restricted cubic spline; Figure S6 weighted cumulative mortality for Emulations 2 and 3; Figure S7 cumulative mortality by RV phenotype. | | | | |
| Supplementary tables: Table S1 echo-selected vs non-selected Sepsis-3 stays; Table S2 RV category mapping; Table S3 fluid item codes; Table S4 legacy 26-day in-hospital competing-risks analysis; Table S5 per-covariate missingness; Table S6 full balance table incl. variance ratios, weighted means and within-stratum balance; Table S7 ICEMAN credibility assessment; Table S8 baseline severity by echocardiographic phenotype. | | | | |

Excluding stays with pre-admission heart failure, ischaemic heart disease or arrhythmia (2,247 excluded, 30.4%) left the overall association intact (HR 1.61, 1.39–1.87; RD + 11.5%, 8.3–15.3), and the hyperdynamic-stratum estimate was HR 1.33 (0.90–1.97; RD + 7.2%, − 2.6 to 17.9; n = 670, 157 treated). The protective reversal reported in the previous version of this analysis (HR 0.54) is therefore attributable to the earlier definition of cardiac disease from index-admission codes—which removed three quarters of

the cohort, enriched for post-exposure cardiac diagnoses—and to reuse of full-cohort weights; it does not replicate with pre-admission ascertainment and re-estimated weights. We no longer report a protective subgroup effect.

## 3.6 Emulation 2: dobutamine new-user within LVEF < 40%

Among 1,201 landmark stays with LVEF < 40%, dobutamine initiation (n = 100) was not significantly associated with 28-day mortality after adjustment for concurrent fluid administration (HR 1.57, 0.95–2.60; RD + 10.8%, − 1.7 to 23.2; weighted control risk 22.8%; Figure S6). The parsimonious penalised propensity model gave a similar estimate (HR 1.55, 0.98–2.47), as did pooling dobutamine and milrinone initiators (n = 132; HR 1.12, 0.70–1.79). Post-weighting effective treated sample size was 53–64, so these intervals are wide and compatible with clinically important harm or benefit; no post-hoc power threshold is quoted.

## 3.7 Emulation 3: vasopressin × hyperdynamic interaction

Vasopressin initiation (n = 360) was strongly associated with 28-day mortality (HR 2.55, 2.13–3.06; RD + 23.9%, 18.6–28.7; Figure S6). The median propensity score among non-users was 0.03, documenting that vasopressin initiation is nearly synonymous with refractory shock; we interpret the main association as dominated by rescue-therapy indication. Among stays with measured LVEF, the association did not differ significantly by hyperdynamic phenotype: multiplicative interaction HR 1.34 (0.83–2.17; p = 0.23); additive difference + 14.9% (p = 0.065) (hyperdynamic n = 777, 56 treated: HR 3.30, 2.16–5.04; non-hyperdynamic n = 4,636: HR 2.42, 1.91–3.06).

## 3.8 Bias analyses

With conversion of the hazard ratio to an approximate risk ratio at the weighted control risk (24.8%), the E-value was 2.60 for the point estimate and 2.29 for the confidence bound; the directly computed 28-day risk-ratio E-value was 2.57 (risk ratio 1.59). Tipping-point analysis showed that an unmeasured confounder with an outcome risk ratio of 2.0 would need a prevalence imbalance of roughly 60% versus 20% between exposure groups to fully explain the association. Such an imbalance is not implausible—imbalances of similar magnitude were observed for measured covariates before weighting (e.g., pre-TTE vasopressor use 56.2% vs 24.7%)—so the association cannot be considered robust to unmeasured confounding. Prior-admission negative control outcomes remained modestly under-recorded in the high-volume group after weighting (any NCO: OR 0.82; osteoarthritis OR 0.73; non-toxic goitre OR 0.78; benign prostatic hyperplasia OR 0.91; cataract OR 0.92). Because these conditions are documented before exposure, an inverse association indicates residual bias that makes the exposure appear protective in the chronic-disease dimension; for the harmful primary estimate this implies bias toward the null (underestimation of harm), although the magnitude and direction of residual bias for the mortality outcome itself cannot be inferred directly. Together with the 24-h SOFA imbalance after weighting (SMD 0.154), the NCO results demonstrate residual systematic bias despite excellent measured-covariate balance.

# 4. Discussion

## 4.1 Principal findings

In 7,398 Sepsis-3 ICU stays with early echocardiography, and with every confounder measured strictly before the exposure window, higher fluid volume in the 24 h after echocardiography was consistently associated with higher 28-day mortality. The association was remarkably uniform: it persisted in every LVEF and RV phenotype stratum, within every baseline-severity stratum, at alternative thresholds, as a continuous exposure, with a shortened 6-h exposure window, and after excluding pre-ICU echocardiograms. Formal interaction tests on both relative and absolute scales found no evidence that echocardiographic phenotype modifies the association, and the previously reported protective reversal in hyperdynamic hearts without cardiac disease did not survive correction of comorbidity ascertainment. The dobutamine emulation was too small for firm conclusions, and the vasopressin emulation illustrates the reach of rescue-therapy confounding even with overlap weighting.

## 4.2 Relation to prior work

The association we observe is directionally consistent with the observational literature coupling positive fluid balance and higher resuscitation volumes with mortality [24–26], and with restrictive strategies being at least as safe as liberal ones in randomised trials (CLASSIC, CLOVERS) [3,4]. Our estimate cannot explain the neutrality of those trials: the trials randomised strategies early, achieved smaller volume separation than the contrast studied here, and enrolled patients before the selective ordering of echocardiography that defines our cohort. Our results are best read as consistent with—but not causal proof of—the harm signalled by conservative fluid management. For dobutamine, the absence of randomised mortality benefit and the neutral milrinone-versus-dobutamine comparison in cardiogenic shock align with our imprecise estimate [27]; for vasopressin, randomised comparisons against norepinephrine were neutral on mortality (VASST, VANISH), while our strongly positive observational association illustrates why rescue therapies resist observational emulation [28,29]. ANDROMEDA-SHOCK showed that targeting a physiological phenotype is feasible at trial scale [5]; our data, however, do not provide observational support for choosing cardiac ultrasound phenotypes as the stratification axis of such trials.

## 4.3 Interpretation

Three readings of the overall association remain open. First, it may be causal: excess early volume after the initial resuscitation phase could plausibly add harm through oedema and venous congestion without correcting vasoplegia or hypovolaemia. Second, it may reflect residual confounding by indication: severity evolves during even a 24-h window, patients who deteriorate receive more fluid, and our negative control and non-model-covariate analyses demonstrate residual systematic bias after weighting. Third, both may operate together; the attenuation of relative risks with increasing baseline severity is compatible with either. What the data do not support is effect modification by echocardiographic phenotype: stratum-specific estimates are all positive with overlapping intervals, and every formal

interaction test is null. We therefore no longer claim that the fluid–mortality association is "phenotype-dependent", "absent" in hyperdynamic hearts, or "strongest" with RV dysfunction, and we present the uniform association—not subgroup contrasts—as the primary finding.

# 4.4 Strengths

The study assembles a large echocardiography-linked sepsis cohort (n = 7,819; LVEF strata n = 6,189; RV status classifiable in 91.8%) and applies a deliberately conservative design: a target trial framework with an explicit protocol table per emulation (TARGET guideline); all covariates measured strictly before time zero, including vasopressor dose, septic shock status, fluid balance and urine output; comorbidities from pre-admission diagnoses; a uniform 28-day all-cause outcome counted from each landmark; formal interaction tests on two scales with multiplicity control and ICEMAN credibility assessment; within-stratum re-estimation of propensity scores; and bias analyses (E-values with hazard-to-risk conversion, tipping points, prior-admission negative controls) reported beside the primary estimates.

# 4.5 Limitations

First, confounding by indication remains the central threat. The association is not robust to a plausible unmeasured confounder (E-value 2.60; tipping-point imbalance 60% vs 20% at RR 2.0), the negative control analysis demonstrates residual bias, and a non-model severity summary (24-h SOFA) remained imbalanced after weighting (SMD 0.154). We report associations, not effects, and do not use them to recommend phenotype-guided management. Second, although covariates were measured strictly before time zero, severity continues to evolve within the 24-h exposure window; the 6-h window sensitivity mitigates but cannot eliminate this time-varying confounding, and we did not implement a clone–censor–weight analysis because adherence to a cumulative-dose strategy is only knowable at the end of the window. Third, echocardiography is ordered selectively (included stays had nearly twice the hospital mortality of excluded stays), clinicians act on echocardiographic findings not captured in structured fields (IVC diameter was available for only 20.2%, and dynamic fluid-responsiveness assessments were unavailable [30]), and 17.2% of echocardiograms preceded ICU admission with incomplete capture of pre-ICU fluids; the cohort is therefore not representative of all patients with sepsis, and phenotype-specific confounding by indication cannot be excluded. Fourth, landmark exclusion of 421 early deaths and discharges removes the sickest and the least severely ill, so estimates apply to patients remaining hospitalised at 24 h. Fifth, LVEF was missing in 16.3% and RV status in 8.2%; analyses were complete-case per phenotype, and median imputation with missingness indicators was used for covariates (multiple imputation gave consistent results). Sixth, the treatment hazard was non-proportional; we emphasise absolute risk differences and report period-specific hazard ratios. Seventh, Emulations 2 and 3 are small (100 and 360 treated; effective treated samples of 53–64 and 347), share patients with Emulation 1, and—in the case of vasopressin—suffer near-structural positivity violations (median control propensity score 0.03). Eighth, single-centre practice patterns, ICD-based comorbidity ascertainment, and the qualitative grading of RV function limit generalisability.

# 4.6 Conclusions

After strict temporal alignment of confounder measurement, exposure and follow-up, higher fluid volume in the 24 h after early echocardiography was consistently associated with higher 28-day mortality in sepsis. We found no statistically supported evidence that LVEF or RV phenotype modifies this association, and an earlier subgroup signal of protection in hyperdynamic hearts was an artefact of comorbidity ascertainment. These associations are vulnerable to unmeasured confounding and should not guide bedside fluid decisions; they do, however, define the evidentiary baseline against which future randomised, phenotype-stratified resuscitation trials should be designed.

## Declarations

**Ethics approval and consent to participate:**

Approved by the Ethics Committee of Shanghai Pudong New Area People's Hospital (approval No. K82). MIMIC-IV is de-identified; the institutional requirement for individual informed consent was waived.

**Consent for publication:**

Not applicable.

# Competing interests:

The authors declare that they have no competing interests.

# Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this manuscript, the authors used generative AI tools to assist with drafting, language editing, and formatting. All authors have read and approved the final manuscript and take full responsibility for its scientific content, data interpretation, and conclusions.

# Funding:

In-hospital Research Project of the People's Hospital of Pudong New Area, Shanghai (E23-03, Yilong Hao); Clinical Outstanding Discipline Construction Project of Pudong New Area, Shanghai (PWYgy2021-09, Jian Wan); Discipline Leadership Talent Training Program of the Pudong New Area Health System, Shanghai (PWRd2024-12, Song Chen); Key Medical Discipline Team Construction Project of Pudong New Area, Shanghai (PWZxq2022-13, Jian Wan); Clinical New-Quality Specialty Project of the Peak and Plateau Discipline Construction in Pudong New Area, Shanghai (2026-PWXZ-05, Peiwu Li).

# Author Contribution

H.Y. and Y.H. conceived and designed the study. H.Y. performed data extraction and statistical analysis. H.Y. and Y.H. drafted the main manuscript text. Q.Z., S.C., P.L. and J.W. contributed to study design, clinical interpretation and critical revision of the manuscript. All authors read and approved the final manuscript.

# Acknowledgement

The authors thank the MIT Laboratory for Computational Physiology for maintaining MIMIC-IV, and the peer reviewers whose methodological critique prompted the redesign of this analysis.

# Data Availability

All data used in this study are from the MIMIC-IV and MIMIC-IV-ECHO databases, which are available to credentialed researchers via PhysioNet (https://physionet.org/) upon completion of the required training and data use agreement. No new patient-level data were generated or collected. All extraction and analysis code is publicly archived at https://doi.org/10.5281/zenodo.23113388 (Zenodo). MIMIC-IV permits code sharing under its data use agreement.

## 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(8):801–810.

2. 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(11):1181–1247.

3. Meyhoff TS, Hjortrup PB, Wetterslev J, Sivapalan P, Laake JH, Cronhjort M, et al. Restriction of intravenous fluid in ICU patients with septic shock. N Engl J Med. 2022;386(26):2459–2470.

4. National Heart, Lung, and Blood Institute PETAL Clinical Trials Network; Shapiro NI, Douglas IS, Brower RG, Brown SM, Exline MC, et al. Early restrictive or liberal fluid management for sepsis-induced hypotension. N Engl J Med. 2023;388(6):499–510.

5. Hernández G, Ospina-Tascón GA, Damiani LP, Estenssoro E, Dubin A, Hurtado J, et al. Effect of a resuscitation strategy targeting peripheral perfusion status vs serum lactate levels on 28-day mortality among patients with septic shock: the ANDROMEDA-SHOCK randomized clinical trial. JAMA. 2019;321(7):654–664.

6. Seymour CW, Kennedy JN, Wang S, Chang CCH, Elliott CF, Xu Z, et al. Derivation, validation, and potential treatment implications of novel clinical phenotypes for sepsis. JAMA. 2019;321(20):2003–2017.

7. Vieillard-Baron A, Millington SJ, Sanfilippo F, Chew M, Diaz-Gomez J, McLean A, et al. A decade of progress in critical care echocardiography: a narrative review. Intensive Care Med. 2019;45(6):770–788.

8. Beesley SJ, Weber G, Sarge T, Nikravan S, Grissom CK, Lanspa MJ, et al. Septic cardiomyopathy. Crit Care Med. 2018;46(4):625–634.

9. Paonessa JR, Brennan T, Pimentel M, Steinhaus D, Feng M, Celi LA. Hyperdynamic left ventricular ejection fraction in the intensive care unit. Crit Care. 2015;19(1):288.

10. Vallabhajosyula S, Kumar M, Pandompatam G, Sakhuja A, Kashyap R, Kashani K, et al. Prognostic impact of isolated right ventricular dysfunction in sepsis and septic shock: an 8-year historical cohort study. Ann Intensive Care. 2017;7(1):94.

11. Sanfilippo F, Corredor C, Fletcher N, Landesberg G, Benedetto U, Foex P, et al. Diastolic dysfunction and mortality in septic patients: a systematic review and meta-analysis. Intensive Care Med. 2015;41(6):1004–1013.

12. Hernán MA, Robins JM. Using big data to emulate a target trial when a randomized trial is not available. Am J Epidemiol. 2016;183(8):758–764.

13. Li F, Morgan KL, Zaslavsky AM. Balancing covariates via propensity score weighting. J Am Stat Assoc. 2018;113(521):390–400.

14. Ray WA. Evaluating medication effects outside of clinical trials: new-user designs. Am J Epidemiol. 2003;158(9):915–920.

15. Johnson AEW, Bulgarelli L, Shen L, Gayles A, Shammout A, Horng S, et al. MIMIC-IV, a freely accessible electronic health record dataset. Sci Data. 2023;10:1.

16. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. Lancet. 2007;370(9596):1453–1457.

17. 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(10):e1001885.

18. Cashin AG, Hansford HJ, McAuley JH, et al. Transparent reporting of observational studies emulating a target trial: the TARGET statement. JAMA. 2025;334(12):1084–1093.

19. McDonagh TA, Metra M, Adamo M, et al. 2021 ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure. Eur Heart J. 2021;42(36):3599–3726.

20. Latouche A, Allignol A, Beyersmann J, Labopin M, Fine JP. A competing risks analysis should report results on all cause-specific hazards and cumulative incidence functions. J Clin Epidemiol. 2013;66(6):648–653.

21. Schandelmaier S, Briel M, Varadhan R, et al. Development of the Instrument to assess the Credibility of Effect Modification Analyses (ICEMAN) in randomized controlled trials and meta-analyses. CMAJ. 2020;192(32):E901–E906.

22. VanderWeele TJ, Ding P. Sensitivity analysis in observational research: introducing the E-value. Ann Intern Med. 2017;167(4):268–274.

23. VanderWeele TJ. Optimal approximate conversions of odds ratios and hazard ratios to risk ratios. Biometrics. 2020;76(3):746–752.

24. 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(2):259–265.

25. Marik PE, Bellomo R. A rational approach to fluid therapy in sepsis. Br J Anaesth. 2016;116(3):339–349.

26. National Heart, Lung, and Blood Institute Acute Respiratory Distress Syndrome (ARDS) Clinical Trials Network; Wiedemann HP, Wheeler AP, Bernard GR, Thompson BT, Hayden S, et al. Comparison of two fluid-management strategies in acute lung injury. N Engl J Med. 2006;354(24):2564–2575.

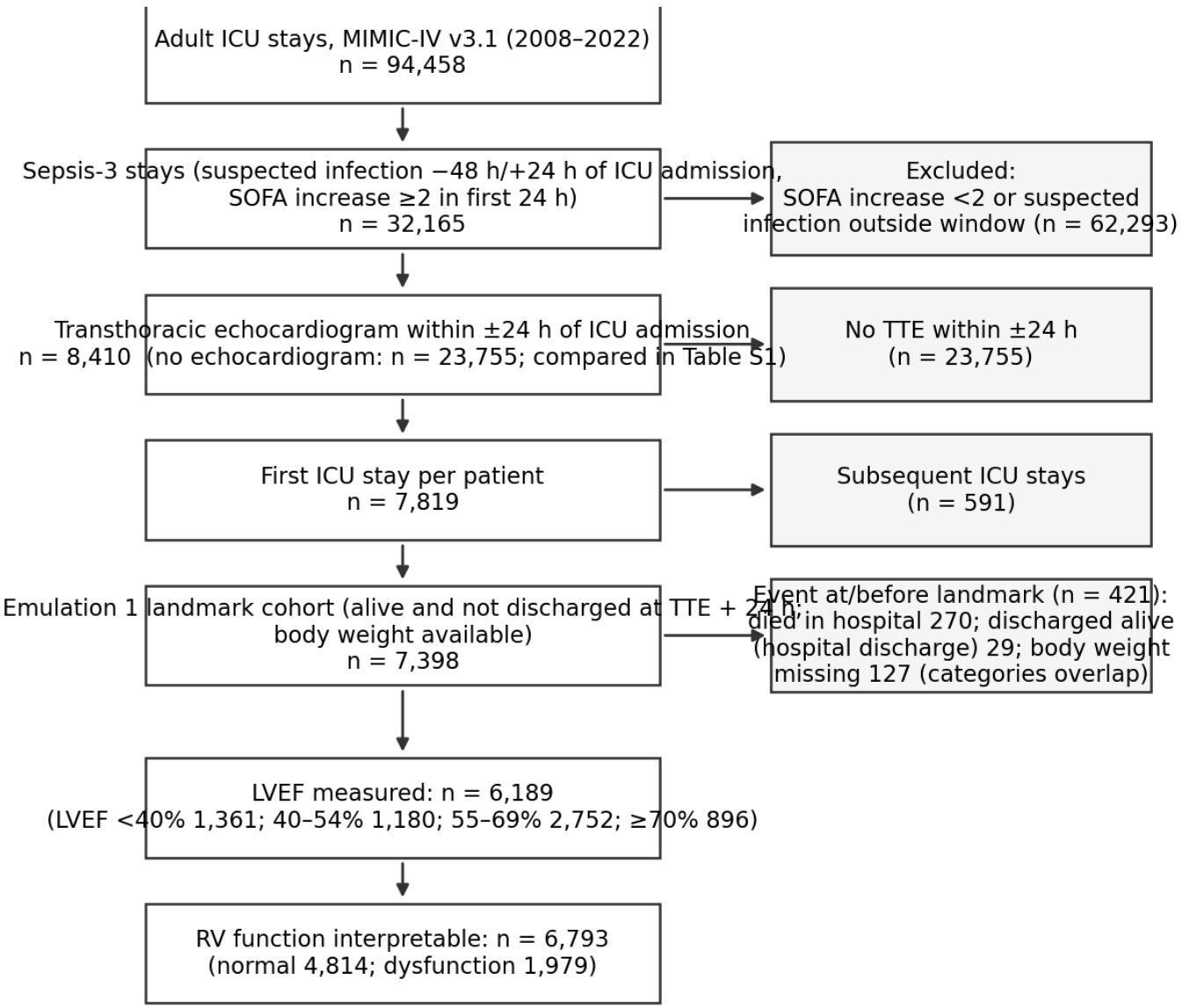
27. Mathew R, Di Santo P, Jung RG, Marbach JA, Hutson J, Simard T, et al. Milrinone as compared with dobutamine in the treatment of cardiogenic shock. N Engl J Med. 2021;385(6):516–525.

28. Russell JA, Walley KR, Singer J, Gordon AC, Hébert PC, Cooper DJ, et al. Vasopressin versus norepinephrine infusion in patients with septic shock. N Engl J Med. 2008;358(9):877–887.

29. Gordon AC, Mason AJ, Thirunavukkarasu N, Perkins GD, Cecconi M, Cepkova M, et al. Effect of early vasopressin vs norepinephrine on kidney failure in patients with septic shock: the VANISH randomized clinical trial. JAMA. 2016;316(5):509–518.

30. Monnet X, Marik P, Teboul JL. Passive leg raising for predicting fluid responsiveness: a systematic review and meta-analysis. Intensive Care Med. 2016;42(12):1935–1947.

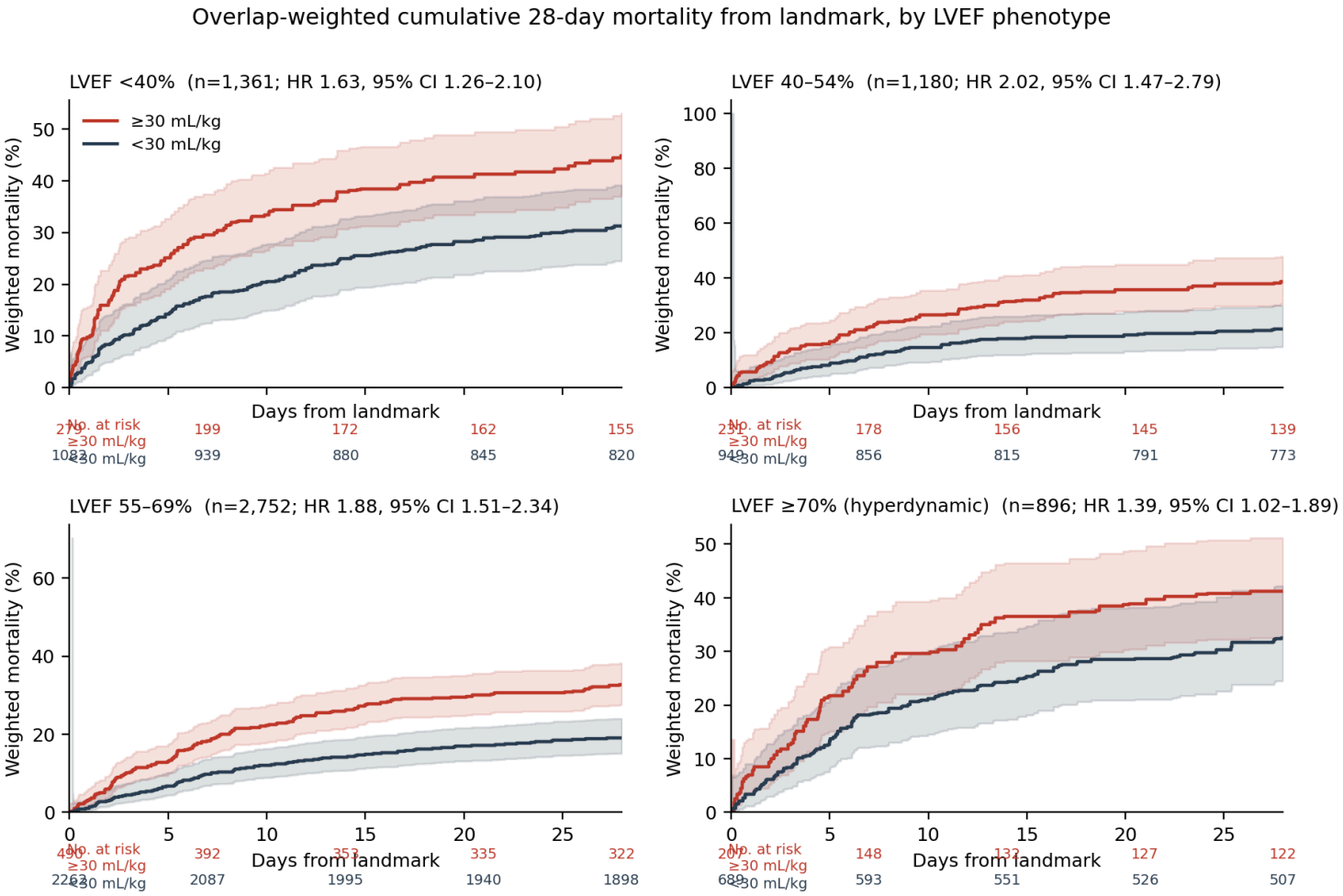
# Figures




TTE = transthoracic echocardiography. Landmark exclusion uses hospital discharge (not ICU discharge). LVEF and RV strata are not mutually exclusive; counts are complete cases per phenotype.

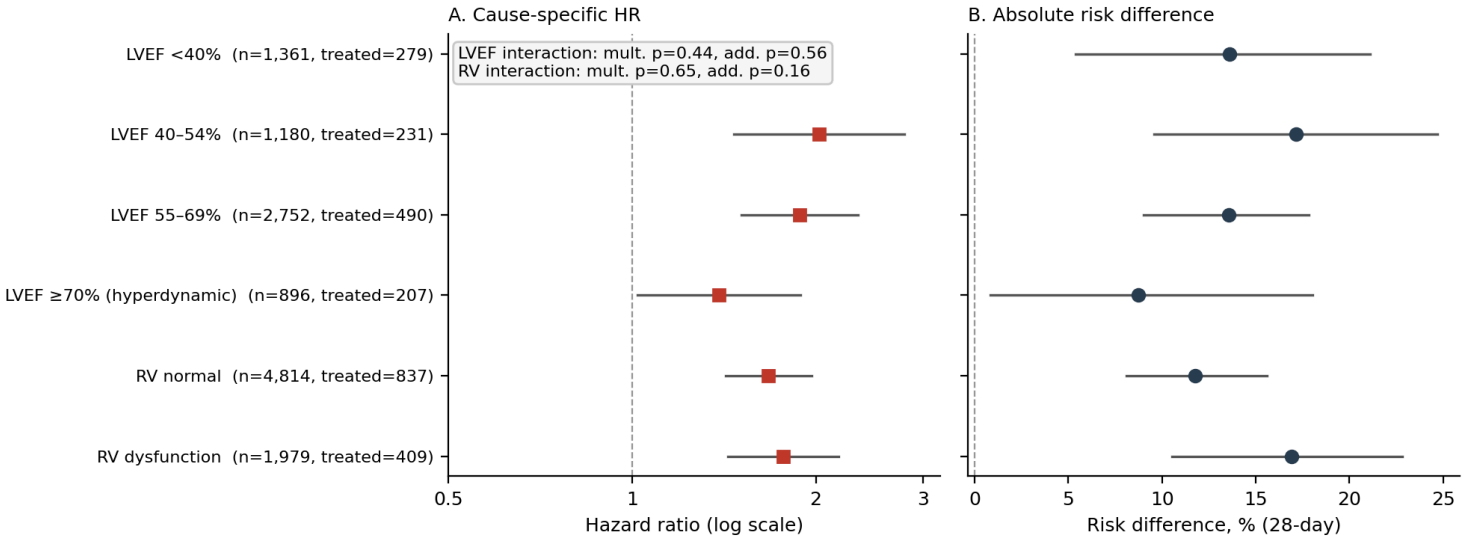
**Figure 1**

Cohort flow diagram. Landmark exclusions use hospital discharge; death, live discharge and missing body weight are reported separately. The comparison of Sepsis-3 stays with and without qualifying echocardiography is in Table S1.




## Figure 2

Overlap-weighted cumulative 28-day mortality from landmark by LVEF phenotype, with 95% confidence bands and numbers at risk. Propensity scores re-estimated within each stratum.




## Figure 3

Phenotype-stratified associations for Emulation 1 on the relative (A, hazard ratio) and absolute (B, 28-day risk difference) scales, with formal interaction p-values.

## Supplementary Files

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

- Table1.docx
- Table2.docx
- Table4.docx
- Table3.docx
- Supplementv4.docx
- v4figS1kmrv.png
- v4figS2ps.png
- v4figS3rcs.png
- v4figS4kme2e3.png