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.

Early benzodiazepine exposure and short-term risk of hepatic encephalopathy treatment initiation in critically ill patients with cirrhosis: a target trial emulation using the MIMIC-IV database

Chudi Huang
Zhejiang Rongjun Hospital
Qihuan Xia
Zhejiang Rongjun Hospital
Bin Du
Zhejiang Rongjun Hospital
Yue Lu
Zhejiang Rongjun Hospital
Miao Ding
Zhejiang Rongjun Hospital
Hong Shao
Zhejiang Rongjun Hospital
Sheng Xu
593382863@qq.com
Zhejiang Rongjun Hospital

Research Article

Keywords: hepatic encephalopathy, benzodiazepines, cirrhosis, target trial emulation, clone-censor-weight, MIMIC-IV, negative control

Posted Date: November 12th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11513932/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: Sedation is frequently required in the intensive care unit (ICU), yet benzodiazepines are traditionally avoided in cirrhosis because of their GABAergic potential to precipitate or worsen hepatic encephalopathy (HE). Whether early ICU benzodiazepine exposure actually increases short-term HE risk remains unquantified: to our knowledge, no previous study has estimated this contrast at the treatment-strategy level.

Objective
To evaluate whether benzodiazepine initiation within 24 h of ICU admission is associated with short-term HE worsening, operationalized as new initiation of lactulose or rifaximin between 24 h and 7 days, among critically ill patients with cirrhosis.

Methods: Using the Medical Information Mart for Intensive Care IV (MIMIC-IV, v3.1) database, we emulated a target trial and applied a clone-censor-weight (CCW) approach. Exposure was benzodiazepine administration (midazolam, lorazepam, or diazepam, any route) within a grace period after ICU admission (primary: 24 h; sensitivity: 12 and 36 h). The primary outcome was HE treatment initiation, defined as new lactulose or rifaximin dispensing between 24 h and 7 days. Censoring was addressed with per-arm inverse probability of censoring weights (IPCW). Two negative-control outcomes (new proton pump inhibitor or statin initiation) quantified residual confounding.

Introduction

Cirrhosis is a leading cause of critical illness and healthcare utilization worldwide [1], and hepatic encephalopathy (HE) is its defining complication, associated with prolonged hospitalization, readmission, and excess mortality [2, 3]. A substantial proportion of patients with decompensated cirrhosis require intensive care for complications such as variceal bleeding, spontaneous bacterial peritonitis, sepsis, and alcohol-associated hepatitis, and short-term mortality in this population remains high [4–6]. In the ICU, agitation, dyspnea, patient-ventilator asynchrony, and alcohol withdrawal frequently prompt sedation, and benzodiazepines—midazolam, lorazepam, and diazepam—remain among the most titratable and familiar agents [7, 8].

The pharmacologic rationale for caution in cirrhosis is longstanding. Benzodiazepine agonism of the GABA-A receptor is hypothesized to interact with the increased GABAergic tone and ammonia-related neurotoxicity of liver failure, a mechanism advanced by Basile and colleagues to explain hepatic encephalopathy itself [9–13], supported by the demonstration of elevated endogenous benzodiazepine-like compounds in brain [9] and by the transient—but not survival-improving—effect of the receptor antagonist flumazenil [14]. Clinical guidance has accordingly recommended avoiding benzodiazepines in patients with cirrhosis whenever possible [15], and impaired hepatic clearance further prolongs their effect in these patients [16, 17]. Yet agitation in critically ill patients with cirrhosis is common and difficult to manage, alternatives such as propofol and dexmedetomidine have their own limitations in hepatic

failure, antipsychotics did not improve outcomes in general ICU delirium [18], and alcohol withdrawal—highly prevalent in alcohol-associated cirrhosis—is still most often treated with benzodiazepines, although phenobarbital and other alternatives are increasingly used [19–21]. The net balance of risks at the bedside is therefore uncertain: it remains unclear whether early ICU benzodiazepine exposure worsens short-term hepatic encephalopathy outcomes, because no adequately controlled, strategy-level estimate of this contrast exists.

Observational evaluations of early ICU medications face well-described time-related biases. Because sedation is initiated after admission in response to evolving agitation or respiratory failure, naïve comparisons of "ever" versus "never" users are vulnerable to immortal time bias and confounding by indication [22, 23]. Target trial emulation (TTE) provides a principled framework—explicit eligibility, treatment strategies, time zero, follow-up, and outcomes—and clone-censor-weight (CCW) methods can approximate per-protocol contrasts when treatment is initiated within a grace period [22, 24–27]. Equally important, observational estimates of medication-initiation outcomes in ICU populations can be contaminated by a generic "sicker patients receive more medications" pathway; negative-control outcomes, which share the care-intensity pathway but not the putative biologic one, quantify this contamination [28, 29], complemented by sensitivity measures such as the E-value [30].

We therefore asked whether, among critically ill adults with cirrhosis, initiation of a benzodiazepine within 24 h of ICU admission, compared with no initiation, increases the short-term risk of HE treatment initiation (new lactulose or rifaximin between 24 h and 7 days). To answer this question we emulated a target trial in MIMIC-IV using a clone-censor-weight approach, and we pre-specified negative-control outcomes to distinguish a benzodiazepine-specific association from residual confounding. No directional hypothesis was prespecified; the primary contrast was tested two-sided.

Methods

Study design and data source

This study was designed as a target trial emulation using observational ICU data [22]. Reporting follows the STROBE guidelines [31] and the TARGET statement for target trial emulations [32, 33]. Data were obtained from MIMIC-IV v3.1, a large de-identified critical care dataset from Beth Israel Deaconess Medical Center covering 2008–2022 [34], following the database's reporting recommendations [35]; calendar-era analyses used the structured anchor_year_group field, because MIMIC dates are shifted per patient.

Target trial specification

The emulated trial specified: (eligibility) adults ≥ 18 years with cirrhosis (ICD-10 K70.2, K70.3, K70.4, K71.7, K74.x) whose ICU stay exceeded 24 h, with no benzodiazepine exposure before ICU admission

(new-user design, ascertained across intravenous administration records, pharmacy dispensing, and electronic medication administration records) and no lactulose or rifaximin within the first 24 h (outcome new-user); (strategies) A, initiate a benzodiazepine (midazolam, lorazepam, or diazepam, any route) within 24 h of ICU admission; B, do not; (time zero) ICU admission, with eligibility, assignment, and follow-up aligned. One eligibility condition—remaining in the ICU beyond 24 h—is realised only after time zero; the emulated trial is therefore explicit that its estimands apply to patients still in the ICU at 24 h, which coincides with the earliest point at which the outcome is observable, and patients dying or discharged within 24 h are outside the emulation by construction. Time zero was set at ICU admission rather than at the 24-hour landmark: cloning every eligible stay at admission and applying a grace period retains patients who receive benzodiazepines, die, or initiate the outcome within the first 24 h, whereas a 24-hour time zero would condition on surviving the window and reintroduce immortal-time structure; (outcome) HE treatment initiation between 24 h and 7 days; (estimand) per-protocol effect, because treatment intention is not recorded in routine data; (analysis plan) CCW with per-arm inverse probability of censoring weights.

Benzodiazepine exposure was identified from administration records (inputevents item IDs 221668, 221385, and 221623 for midazolam, lorazepam, and diazepam, respectively), supplemented by pharmacy and eMAR records of the same agents to capture oral administration. Cirrhosis was identified from explicit cirrhosis ICD-10 codes (ICD-10 K70.3, K71.7, K74.60, K74.69), which covered 94.0% of the cohort; the remainder qualified through fibrosis-family codes (65 stays) or alcoholic hepatic failure (7 stays) and were removed in a pre-specified sensitivity analysis. Hepatic encephalopathy treatment initiation was identified from pharmacy dispensing records of lactulose or rifaximin, which were corroborated by eMAR in 97.4% of outcome-window initiations; the complete medication-code list, including entries checked but containing no records, is provided in Supplementary Table S1. Alcohol-withdrawal protocol administrations retained under the drug-exposure definition were removed in a pre-specified sensitivity analysis.

Outcomes

The primary outcome was HE treatment initiation, defined as new lactulose or rifaximin dispensing between 24 h and 7 days after ICU admission. Secondary outcomes were 28-day and 7-day all-cause mortality (from date of ICU admission) and ICU length of stay. Negative-control outcomes were new proton pump inhibitor and new statin dispensing in the same 24 h-to-7-day window: these share the “sicker patients receive more medications” pathway but have no biologic plausibility for a benzodiazepine effect on HE.

Covariates

Censoring models included age, sex, and first-24-hour values of the components of the MELD score (total bilirubin, creatinine, INR [5, 36]), lactate, white cell count, hemoglobin, platelets, sodium, potassium,

albumin, heart rate, systolic blood pressure, respiratory rate, oxygen saturation, first Richmond Agitation-Sedation Scale (RASS) score, vasopressor use, and mechanical ventilation. Severity scores (SOFA, APSIII, GCS, CAM-ICU) were not included because they are measured over the first 24 h—partly after exposure—and are largely redundant with the direct physiologic variables. Because first-24-hour values can straddle exposure—among exposed stays, the first measurement preceded the first benzodiazepine dose for only 46–73% of variables (lowest for vasopressor initiation, 39%)—two pre-specified sensitivity analyses re-estimated the entire chain: one removed the three post-exposure-capable measures (first RASS, vasopressor use, mechanical ventilation) from the censoring models, and the stricter one restricted every covariate to its last measurement strictly before the first dose (for unexposed clones, before 24 h), with vasopressor and ventilation (ascertained from chartevents ventilator-mode entries) required to have been initiated before the dose; pre-dose measurement availability ranged from 49% (albumin) to 92% (heart rate), and unavailable values were handled by the same median-plus-indicator imputation. Continuous covariates were standardized; missing values were imputed with cohort medians plus missingness indicators. Laboratory coverage in the final cohort was high (creatinine 99.2%, INR 94.0%, bilirubin 84.7%).

Statistical analysis

Each ICU stay was cloned into both strategies. Clones were censored when observed treatment deviated from the assigned strategy: strategy-B clones at actual benzodiazepine initiation, strategy-A clones at the grace-period boundary if not initiated. Death within the outcome window without the outcome was treated as a censoring event addressed by weighting (death was never used to censor survival analyses). This choice defines the primary fixed-window estimand as a hypothetical-survival contrast—HE treatment initiation if clones were, counterfactually, kept under observation through day 7—and was made for consistency with the fixed-window risk difference; the Aalen-Johansen analysis (Fig. 5a) reports the complementary competing-risk quantity, in which death before the outcome removes the clone from the risk set rather than being imputed away. The two estimands answer different questions and are reported side by side. Censoring models were fitted separately per arm with L2-regularized logistic regression (perfect separation occurred with unpenalized fits); stabilized weights were the ratio of the marginal to the fitted probability of remaining uncensored, truncated at the 99th percentile of uncensored weights within each arm. The marginal odds ratio was derived from the weighted risks. Confidence intervals and p-values came from patient-level nonparametric cluster bootstrap (1,000 replicates, resampling patients together with all of their ICU stays), re-estimating the entire clone-censor-weight chain in each replicate; replicates in which the regularized censoring models failed to converge were discarded, and at least 93% of replicates converged in every analysis.

Diagnostics and sensitivity analyses were pre-specified: weight-truncation percentile curves (none, 0.5/99.5, 1/99, 2.5/97.5, 5/95); effective sample sizes $(\Sigma w)^2/\Sigma w^2$ by arm; censoring-model discrimination; covariate-omission tests; grace periods of 12 and 36 h; IV-only exposure; alternative outcome definitions (lactulose only, rifaximin only, both); first ICU stay per patient; and the two negative-control outcomes. An E-value was computed for the weighted risk ratio (1.22). Analyses used Python

3.11 (polars, statsmodels, scikit-learn, matplotlib). Two-sided $p < 0.05$ was considered statistically significant for the single primary hypothesis; all secondary, sensitivity, and subgroup analyses are exploratory.

Results

Cohort

Of 94,458 ICU stays, 3,755 met the cirrhosis definition; after requiring length of stay > 24 h (n = 3,044), excluding prior benzodiazepine use across all medication channels (n = 750 excluded), and excluding prevalent HE treatment within 24 h (n = 1,097 excluded), 1,197 stays (340 exposed [28.4%], 857 unexposed) formed the analytical cohort (Fig. 1). Table 1 shows the full baseline distribution; standardized mean differences (SMDs) are reported in preference to p-values.

Table 1
Baseline characteristics of the analytical cohort within the first 24 hours of ICU admission.

Variable	Total	Benzodiazepine	Control	p-value	SMD (pre-weighting)	SMD (post-weighting)
N	1,197	340	857			
Demographics						
Age, years	61.0 (52–68)	56.0 (49–64)	62.0 (54–69)	0.000	-0.39	-0.03
Male	797 (66.6%)	223 (65.6%)	574 (67.0%)	0.695	-0.03	-0.04
White	803 (67.1%)	226 (66.5%)	577 (67.3%)	0.829	-0.02	0.03
Black	99 (8.3%)	26 (7.6%)	73 (8.5%)	0.706	-0.03	-0.01
Liver disease severity						
Bilirubin, mg/dL	1.6 (0.8–3.5)	1.7 (0.9–3.9)	1.6 (0.8–3.3)	0.037	0.15	0.00
Creatinine, mg/dL	1.1 (0.8–1.8)	1.0 (0.7–1.6)	1.1 (0.8–1.8)	0.003	-0.01	-0.05
INR	1.6 (1.3–1.9)	1.6 (1.3–1.9)	1.6 (1.3–1.9)	0.814	0.01	0.08
Albumin, g/dL	2.9 (2.5–3.3)	2.8 (2.4–3.3)	2.9 (2.5–3.3)	0.639	-0.05	-0.11
Platelets, K/uL	121.0 (78–186)	113.0 (67.8–182.2)	125.0 (81–187)	0.005	-0.11	-0.00
Sodium, mEq/L	138.0 (134–140)	137.0 (134–140)	138.0 (134–140)	0.044	-0.14	-0.01
Potassium, mEq/L	4.2 (3.8–4.8)	4.2 (3.8–4.7)	4.3 (3.8–4.8)	0.129	-0.05	-0.01

Data are presented as median (interquartile range) for continuous variables and n (%) for categorical variables. SMD, standardized mean difference. SMD < 0.1 was defined as acceptable covariate balance. Pre-weighting SMD denotes balance before inverse probability of censoring weighting; post-weighting SMD denotes balance after weighting. P-values are from unweighted Mann-Whitney U tests (continuous) or chi-square tests (categorical). BZD, benzodiazepine; ICU, intensive care unit; RASS, Richmond Agitation-Sedation Scale.

Variable	Total	Benzodiazepine	Control	p-value	SMD (pre-weighting)	SMD (post-weighting)
Vital signs (first day)						
Heart rate, bpm	90.0 (78–105)	96.0 (80.8–110)	89.0 (76–102)	0.000	0.26	0.03
Systolic BP, mmHg	116.0 (102–133)	118.0 (101–133)	115.0 (102–133)	0.496	0.01	0.04
Mean arterial pressure, mmHg	81.0 (71–93)	82.0 (71–93)	80.5 (71–93)	0.493	0.04	0.02
Respiratory rate, /min	19.0 (16–23)	19.0 (16–23)	18.0 (16–23)	0.049	0.09	0.03
Temperature, C	36.8 (36.5–37.1)	36.8 (36.6–37.1)	36.8 (36.5–37.1)	0.731	-0.02	-0.05
SpO2, %	98.0 (95–100)	98.0 (95–100)	98.0 (95–100)	0.457	0.08	0.04
RASS (first)	0.0 (-3.8–0)	0.0 (-3–0)	0.0 (-4–0)	0.033	0.12	0.03
Laboratory (first day)						
WBC, K/uL	10.0 (6.5–14.9)	8.8 (5.9–13.9)	10.3 (6.7–15.2)	0.003	-0.03	0.08
Hemoglobin, g/dL	9.5 (8–11.2)	9.4 (7.8–11)	9.5 (8.1–11.2)	0.248	-0.05	0.00
Glucose, mg/dL	134.0 (105–181.5)	127.0 (100–168.2)	137.0 (108–186)	0.003	-0.16	-0.12
Lactate, mmol/L	2.3 (1.6–3.7)	2.7 (1.7–4.6)	2.2 (1.5–3.4)	0.000	0.34	0.10

Data are presented as median (interquartile range) for continuous variables and n (%) for categorical variables. SMD, standardized mean difference. SMD < 0.1 was defined as acceptable covariate balance. Pre-weighting SMD denotes balance before inverse probability of censoring weighting; post-weighting SMD denotes balance after weighting. P-values are from unweighted Mann-Whitney U tests (continuous) or chi-square tests (categorical). BZD, benzodiazepine; ICU, intensive care unit; RASS, Richmond Agitation-Sedation Scale.

Variable	Total	Benzodiazepine	Control	p-value	SMD (pre-weighting)	SMD (post-weighting)
Urine output 0-24h, mL	500.0 (250–900)	200.0 (100–750)	500.0 (300–935)	0.008	-0.48	-0.45
Comorbidities and cirrhosis phenotype						
Myocardial infarction	107 (8.9%)	33 (9.7%)	74 (8.6%)	0.636	0.04	0.03
Congestive heart failure	347 (29.0%)	68 (20.0%)	279 (32.6%)	0.000	-0.29	-0.18
Dementia	23 (1.9%)	6 (1.8%)	17 (2.0%)	0.988	-0.02	0.10
Diabetes	448 (37.4%)	107 (31.5%)	341 (39.8%)	0.009	-0.17	-0.13
Renal disease	283 (23.6%)	64 (18.8%)	219 (25.6%)	0.017	-0.16	-0.09
Malignancy	254 (21.2%)	49 (14.4%)	205 (23.9%)	0.000	-0.24	-0.20
Alcohol-related cirrhosis	535 (44.7%)	198 (58.2%)	337 (39.3%)	0.000	0.39	0.25
Variceal bleeding history	335 (28.0%)	128 (37.6%)	207 (24.2%)	0.000	0.30	0.22
Hepatic encephalopathy history	120 (10.0%)	37 (10.9%)	83 (9.7%)	0.606	0.04	0.08
Spontaneous bacterial peritonitis	52 (4.3%)	17 (5.0%)	35 (4.1%)	0.587	0.04	-0.01
ICU management (first day)						
Mechanical ventilation	630 (52.6%)	205 (60.3%)	425 (49.6%)	0.001	0.22	0.04
Antibiotic use	681 (56.9%)	193 (56.8%)	488 (56.9%)	1.000	-0.00	-0.02

Data are presented as median (interquartile range) for continuous variables and n (%) for categorical variables. SMD, standardized mean difference. SMD < 0.1 was defined as acceptable covariate balance. Pre-weighting SMD denotes balance before inverse probability of censoring weighting; post-weighting SMD denotes balance after weighting. P-values are from unweighted Mann-Whitney U tests (continuous) or chi-square tests (categorical). BZD, benzodiazepine; ICU, intensive care unit; RASS, Richmond Agitation-Sedation Scale.

Variable	Total	Benzodiazepine	Control	p-value	SMD (pre-weighting)	SMD (post-weighting)
Any vasopressor	453 (37.8%)	143 (42.1%)	310 (36.2%)	0.068	0.12	0.02
Norepinephrine	321 (26.8%)	118 (34.7%)	203 (23.7%)	0.000	0.24	0.13
Vasopressin	124 (10.4%)	48 (14.1%)	76 (8.9%)	0.010	0.17	0.09
Phenylephrine	212 (17.7%)	60 (17.6%)	152 (17.7%)	1.000	-0.00	-0.03

Data are presented as median (interquartile range) for continuous variables and n (%) for categorical variables. SMD, standardized mean difference. SMD < 0.1 was defined as acceptable covariate balance. Pre-weighting SMD denotes balance before inverse probability of censoring weighting; post-weighting SMD denotes balance after weighting. P-values are from unweighted Mann-Whitney U tests (continuous) or chi-square tests (categorical). BZD, benzodiazepine; ICU, intensive care unit; RASS, Richmond Agitation-Sedation Scale.

Exposed patients were younger (median 56.0 (49.0–64.0) vs 62.0 (54.0–69.0) years; p < 0.001); sex was similar (male 223 (65.6%) vs 574 (67.0%); p = 0.695). Liver disease severity at baseline was comparable between groups: MELD components were nearly identical (total bilirubin 1.7 (0.9–3.9) vs 1.6 (0.8–3.3) mg/dL, p = 0.037; creatinine 1.0 (0.7–1.6) vs 1.1 (0.8–1.8) mg/dL, p = 0.003; INR 1.6 (1.3–1.9) in both, p = 0.814), as were albumin (2.8 vs 2.9 g/dL, p = 0.639), platelets (113.0 vs 125.0 K/μL, p = 0.005) and sodium (137.0 vs 138.0 mEq/L, p = 0.044). In contrast, the clinical phenotype differed markedly: alcohol-related cirrhosis (58.2% vs 39.3%; p < 0.001) and a history of variceal bleeding (37.6% vs 24.2%; p < 0.001) were both more prevalent among exposed patients, whereas prior hepatic encephalopathy (10.9% vs 9.7%; p = 0.606) and spontaneous bacterial peritonitis (5.0% vs 4.1%; p = 0.587) were balanced.

First-day physiology indicated greater acute stress among exposed patients: lactate was higher (2.7 (1.7–4.6) vs 2.2 (1.5–3.4) mmol/L; p < 0.001) and heart rate faster (96.0 (80.8–110.0) vs 89.0 (76.0–102.0) bpm; p < 0.001), while respiratory rate (19.0 vs 18.0/min, p = 0.049), SpO2 (98.0% in both, p = 0.457), white cell count (8.8 vs 10.3 K/μL, p = 0.003) and first RASS score (0.0 (− 3.0 to 0.0) vs 0.0 (− 4.0 to 0.0); p = 0.033) were similar.

ICU management in the first 24 h was more intensive among exposed patients: mechanical ventilation (60.3% vs 49.6%; p = 0.001), norepinephrine (34.7% vs 23.7%; p < 0.001) and vasopressin (14.1% vs 8.9%; p = 0.010) were all more common, while phenylephrine use (17.6% vs 17.7%; p = 1.000) and crystalloid volume (1,248.8 (495.4–3,100.2) vs 1,267.9 (482.3–2,934.7) mL; p = 0.741) were similar. Before weighting, HE treatment initiation was more frequent in the exposed group (36.2% [123/340] vs 23.7% [203/857]) and 28-day mortality higher (30.6% [104/340] vs 21.0% [180/857]), highlighting the need for weighting to account for baseline imbalance in phenotype and treatment intensity rather than liver disease severity.

After weighting, covariate balance was restored: all post-weighting standardized mean differences were below 0.11 (age − 0.390 to − 0.033; lactate 0.337 to 0.099; mechanical ventilation 0.216 to 0.043). Benzodiazepine exposure within 24 h was fairly stable across calendar eras, from 30.7% of eligible stays in 2008–2010 to 25.1% in 2020–2022.

Primary outcome

Before weighting, HE treatment initiation occurred in 36.2% (123/340) of exposed versus 23.7% (203/857) of unexposed stays (crude difference + 12.5 percentage points). After CCW with IPCW (the clone-and-censor design is illustrated in Fig. 2), the weighted incidence was 34.6% versus 28.3% (risk difference + 6.3 percentage points, 95% CI 0.7 to 13.2; weighted odds ratio 1.34; p = 0.030, from a patient-level cluster bootstrap with 1,000 replicates resampling all ICU stays of each patient) (Table 2, Fig. 3). Effective sample sizes were 216 (arm A) and 750 (arm B); censoring-model c-statistics were 0.73 in both arms; 1,197 stays contributed 304 and 804 uncensored clones. Among these uncensored clones, the unweighted risk of HE treatment initiation was 123/304 (40.5%) in arm A and 203/804 (25.2%) in arm B; IPCW re-weighting shifted these to 34.6% and 28.3%, respectively, because clones censored for treatment deviation or death before the outcome were differentially weighted by the censoring models. Exploratory subgroup analyses are presented in Fig. 4b, and competing-risk and survival analyses in Fig. 5.

Table 2

Weighted primary, negative-control, and sensitivity analyses for hepatic encephalopathy treatment initiation (analytical cohort, n = 1,197 ICU stays; 304 and 804 uncensored clones contributed weighted risk estimates, with 123 and 203 events among uncensored clones in arms A and B). All estimates re-estimate the full clone-censor-weight chain; risk differences and 95% confidence intervals are from 1,000-replicate patient-level cluster bootstraps (resampling all ICU stays of each patient). Negative-control outcomes share the care-intensity pathway but not the putative biologic pathway; their confidence intervals should include the null in the absence of residual confounding. PPI, proton pump inhibitor; HE, hepatic encephalopathy; IV, intravenous.

Analysis	**Weighted risk, Strategy A (%)**	**Weighted risk, Strategy B (%)**	**Risk difference, pp (95% CI)**	**p-value**
Primary: HE treatment initiation (24 h grace)	34.6	28.3	+ 6.3 (0.7 to 13.2)	0.030
Negative control: new PPI initiation	48.2	44.3	+ 3.9 (-3.4 to 11.1)	0.371
Negative control: new statin initiation	9.0	9.4	−0.3 (-4.6 to 4.7)	0.903
Grace period 12 h	36.2	28.9	+ 7.3 (0.8 to 14.5)	0.019
Grace period 36 h	35.3	28.1	+ 7.2 (1.6 to 13.9)	0.004
IV-channel exposure only	40.3	28.2	+ 12.2 (5.0 to 21.0)	0.004
Outcome: lactulose only	34.2	27.6	+ 6.6 (0.6 to 13.3)	0.038
Outcome: rifaximin only	16.6	13.3	+ 3.3 (-1.9 to 7.9)	0.211
Outcome: both drugs required	16.1	12.6	+ 3.6 (-1.2 to 8.5)	0.143
First ICU stay per patient	-	-	+ 5.5 (-0.9 to 13.3)	0.096
Censoring models without post-exposure-capable covariates (RASS, vasopressors, ventilation)	-	-	+ 7.7 (1.4 to 13.5)	0.008
All covariates restricted to pre-dose measurements	-	-	+ 10.3 (1.9 to 18.7)	0.012
Cohort restricted to explicit cirrhosis ICD-10 codes (n = 1,125)	-	-	+ 7.2 (0.9 to 14.0)	0.027

Analysis	Weighted risk, Strategy A (%)	Weighted risk, Strategy B (%)	Risk difference, pp (95% CI)	p-value
List of additional files				
• **Additional file 1** (format: PDF): **Figure S1.** Distribution of stabilised inverse-probability-of-censoring weights by arm among uncensored clones (dashed lines, arm-specific 99th-percentile truncation).				
• **Additional file 2** (format: PDF): **Figure S2.** Covariate balance before and after weighting (Love plot): standardized mean differences for all Table 1 variables, before (grey) and after (blue) weighting; dashed lines, ± 0.1.				
• **Additional file 3** (format: PDF): **S1.** Frozen statistical analysis plan (English translation).				
• **Additional file 4** (format: PDF): **S2.** Deviation register (English translation).				
• **Additional file 5** (format: CSV): tableS1_medication_codes.csv				

Figure 2: Schematic of target trial emulation with clone-censor and inverse-probability-of-censoring weighting (CCW-IPCW). (a) Analytical workflow. Time zero was defined as ICU admission. Each eligible ICU stay was cloned into two hypothetical strategies: benzodiazepine (BZD) within 24 h of admission (Strategy A) and no benzodiazepine within 24 h (Strategy B). Clones deviating from their assigned strategy were artificially censored, and uncensored clones were weighted by the inverse probability of censoring. The primary outcome was HE treatment initiation occurring between 24 h and 7 days; the secondary outcome was 28-day mortality. (b) Swimmer plot illustrating cloning and censoring rules using hypothetical patient examples (not real study data). The grace period was 0–24 h; HE outcomes were assessed only in the window from 24 h to 7 days. Strategy A clones were censored if no BZD dose was administered by 24 h; Strategy B clones were censored immediately if BZD was given within 24 h. Symbols: blue circle, BZD administration; orange square, HE treatment initiation; black star, death (competing event); cross, artificial censoring.

Figure 3: Primary, sensitivity, and negative-control estimates (weighted risk differences, percentage points). Squares: primary and sensitivity analyses; circles: negative-control outcomes (new proton pump inhibitor, new statin).

Negative-control outcomes

New PPI initiation in the outcome window had a weighted risk difference of + 3.9 percentage points (95% CI − 3.4 to 11.1; weighted risks 48.2% vs 44.3%; p = 0.37), and new statin initiation − 0.3 percentage points (95% CI − 4.6 to 4.7; 9.0% vs 9.4%; p = 0.90); both are based on 1,000-replicate patient-level cluster bootstraps. By the pre-specified negative-control criterion (control 95% CI should include the null), both controls were compatible with the null, providing no evidence of a residual care-intensity confounding pathway after weighting; residual confounding by unmeasured variables nonetheless cannot be excluded.

Identifiability diagnostics and sensitivity analyses

All sensitivity analyses re-estimated the entire clone-censor-weight chain, with 95% confidence intervals from 1,000-replicate patient-level cluster bootstraps. With a 12-hour grace period, 293 stays (24.5%) were classified as exposed; the weighted risk difference was + 7.3 percentage points (95% CI 0.8 to 14.5; p = 0.019). With a 36-hour grace period, 381 stays (31.8%) were exposed; the weighted risk difference was + 7.2 percentage points (95% CI 1.6 to 13.9; p = 0.004). Restricting the exposure definition to the intravenous channel (187 exposed stays) gave a larger estimate of + 12.2 percentage points (95% CI 5.0 to 21.0; p = 0.004), consistent with intravenous administration marking more severe agitation. Alternative outcome definitions gave directionally consistent results: lactulose only (+ 6.6, 95% CI 0.6 to 13.3; p = 0.038), rifaximin only (+ 3.3, 95% CI − 1.9 to 7.9; p = 0.211), and the joint definition requiring both drugs (+ 3.6, 95% CI − 1.2 to 8.5; p = 0.143); the wider intervals for rifaximin-based definitions reflect their lower incidence (about 16% vs 34% weighted risk), not a change in direction. Restricting the analysis to each patient's first ICU stay (n = 1,026; 306 exposed) gave + 5.5 percentage points (95% CI − 0.9 to 13.3; p = 0.096). Removing the three post-exposure-capable covariates from the censoring models gave + 7.7 percentage points (95% CI 1.4 to 13.5; p = 0.008), and a stricter analysis restricted every covariate to its last measurement before the first benzodiazepine dose (pre-exposure availability 49–92% across variables; vasopressor and ventilation required initiation before the dose) gave + 10.3 percentage points (95% CI 1.9 to 18.7; p = 0.012) — larger than the primary estimate, indicating that adjustment on post-exposure measurements, if anything, attenuates rather than manufactures the association. Restricting the cohort to explicit cirrhosis codes (n = 1,125; 327 exposed) gave + 7.2 percentage points (95% CI 0.9 to 14.0; p = 0.027). Across weight-truncation choices (none through 5/95 percentiles) the weighted risk difference remained positive throughout, ranging from + 3.9 to + 8.2 percentage points (Fig. 4a); all primary and sensitivity estimates are summarised in Fig. 3. An E-value of 1.74 was computed for the weighted risk ratio (1.22): unmeasured confounding would need to be associated with both the strategy and the outcome by a risk ratio of at least 1.74-fold, above and beyond the measured covariates, to explain away the observed association.

Exploratory subgroup analyses

In exploratory analyses using the primary weights (Fig. 4b), the weighted risk difference was larger among more severely ill patients—mechanical ventilation in the first 24 h (13.1, 95% CI 2.9 to 23.4 vs − 1.9, − 11.2 to 7.3 without), first RASS ≤ − 2 (19.1, 6.2 to 31.9 vs − 1.8, − 10.0 to 6.4 above), and INR at or above the cohort median (16.7, 5.4 to 28.0 vs − 1.0, − 9.9 to 8.0 below)—and broadly similar by cirrhosis etiology (6.2, − 3.9 to 16.3 alcohol-related vs 1.7, − 7.9 to 11.4 non-alcohol-related). These analyses were not pre-specified, are based on small strata, involve multiple comparisons, and inherit the confounding demonstrated by the negative controls; they are hypothesis-generating only. Notably, the pattern of larger estimates in sicker patients is equally consistent with residual confounding by severity and with a true effect concentrated in patients with the greatest susceptibility.

Time-to-event analyses

Weighted Kaplan–Meier analysis (Fig. 5) estimated 28-day survival of 74.8% in the benzodiazepine strategy versus 83.6% in the no-benzodiazepine strategy (risk difference − 8.8 percentage points, 95% CI − 15.9 to − 2.9; p = 0.008), with a weighted Cox hazard ratio for 28-day mortality of 1.65 (point estimate only; model-based confidence intervals for weighted Cox are understated and omitted); 7-day weighted survival was 93.5% versus 97.1% (risk difference − 3.6, 95% CI − 7.1 to − 0.3; p = 0.017), and, in a post hoc extension beyond the pre-specified 28-day window, 90-day weighted survival was 70.3% versus 73.8% (risk difference − 3.4, 95% CI − 10.8 to 2.8; p = 0.25). All curve-endpoint risk differences and confidence intervals are from 1,000-replicate patient-level cluster bootstraps re-estimating the full clone-censor-weight chain. The restricted mean survival time at 7 days was 6.91 versus 6.94 days. Weighted median ICU length of stay was 3.6 versus 3.1 days, and the 28-day-truncated weighted mean was 5.75 versus 5.12 days; because longer stays mechanically increase the time during which new medications can be observed, these differences should not be read as longer “at-risk” time implying benefit or harm. In competing-risk analysis treating death before the outcome as a competing event (Aalen–Johansen; Fig. 5a), the 7-day cumulative incidence of HE treatment initiation was 55.3% versus 51.2% (risk difference + 4.2, 95% CI − 6.4 to 15.8; p = 0.45), exceeding the fixed-window weighted proportions (34.6% versus 28.3%) because HE ascertainment was censored at ICU discharge and the estimator imputes hazard beyond discharge.

Discussion

In this target trial emulation among critically ill patients with cirrhosis, early ICU benzodiazepine exposure was associated with a modestly higher rate of short-term HE treatment initiation (+ 6.3 percentage points) that survived clone-censor-weight adjustment, was directionally stable across pre-specified design choices, and was accompanied by restored covariate balance and null negative-control outcomes. Unlike the unweighted contrast (+ 12.5 percentage points), the adjusted estimate is approximately half as large, consistent with confounding by indication contributing substantially to the crude association (a descriptive contrast, not a formal decomposition).

To our knowledge, no previous study has estimated this specific strategy contrast—early ICU benzodiazepine initiation versus avoidance against short-term HE treatment—so our result can only be positioned against indirect evidence, and its direction is consistent with all of it. The closest emulated contrast published to date comes from a different organ system: a recent target trial emulation of early midazolam in acute exacerbations of COPD applied the same clone-censor-weight framework with inverse probability of censoring weights and likewise found an adjusted risk difference whose confidence interval included the null [37]—suggesting that early-sedation emulations in MIMIC-IV share a common profile of modest point estimates and limited precision. Antagonist trials occupy the next position: in a Cochrane review, the benzodiazepine-receptor antagonist flumazenil produced short-lived clinical improvement in a minority of patients with cirrhosis and hepatic encephalopathy, with substantial heterogeneity and low certainty [14]—compatible with a real but modest and transient benzodiazepine contribution, the same profile our adjusted estimate shows. Pharmacokinetic studies occupy the next

position: cirrhosis prolongs benzodiazepine exposure several-fold [8, 16], and elevated endogenous benzodiazepine-like ligands have been demonstrated in overt encephalopathy [9, 12], both supporting susceptibility at the receptor level. What our study adds to this chain is the strategy-level, population-level link that neither antagonism experiments nor pharmacology could supply; equally important, the halving of the crude estimate on weighting shows how much of the apparent literature-level association between sedation and encephalopathy may reflect confounding by indication.

Our results extend a methodological point to the specific setting of medication-proxy outcomes in ICU populations. Drug-initiation endpoints such as “new lactulose or rifaximin”—both established treatments for hepatic encephalopathy [3, 38, 39]—are attractive because they are objective and time-stamped, but they are also directly coupled to the same clinical deterioration that prompts the exposure decision [40, 41]. Both negative-control outcomes were compatible with the null after identical weighting, providing no evidence of a residual care-intensity pathway; we nonetheless suggest that target trial emulations with medication-proxy outcomes in ICU data routinely include such controls, because in this cohort they materially strengthened the interpretation.

The E-value for the weighted risk ratio (1.22) was 1.74, illustrating the fragility from the complementary direction: an unmeasured confounder associated with both exposure and outcome by a risk ratio of 3.6 would suffice to explain the point estimate. Candidate confounders of this strength plausibly exist here: hepatic encephalopathy itself at presentation (grading is incompletely coded), alcohol withdrawal severity, and agitation—all indications for sedation and predictors of HE treatment. First-day RASS, our best-measured marker of agitation, was available in 99.8% of stays; removing or adding it moved the estimate trivially, but RASS captures only part of this pathway.

Two findings require explicit interpretive caution. First, restricting the exposure definition to the intravenous channel produced the largest estimate (+ 12.2 percentage points). Intravenous benzodiazepine administration generally marks more severe agitation—alcohol withdrawal, ventilator dyssynchrony, or procedural urgency—so this gradient is consistent both with a true susceptibility that is greatest when exposure is most intense and with residual confounding by agitation severity that a single baseline RASS score cannot fully capture. Second, the same ambiguity applies to the exploratory subgroup pattern, in which estimates were larger among mechanically ventilated patients, those with a first RASS of − 2 or lower, and those with INR at or above the cohort median: a genuine effect concentrated in the most susceptible patients and residual confounding by illness severity are equally consistent with this pattern. The null negative-control outcomes, although reassuring on average, cannot exclude residual confounding that co-occurs with severity in specific subgroups; all secondary, sensitivity, and subgroup contrasts therefore remain descriptive. Because a single primary hypothesis was pre-specified and all other contrasts are exploratory, no multiplicity adjustment was applied; the subgroup findings in particular should be read as hypothesis-generating given the number of comparisons. Finally, although the Aalen-Johansen curves separate visually (Fig. 5a), their pointwise confidence bands overlap substantially and the endpoint risk-difference interval includes the null; the

apparent separation should not be read as evidence of a difference, and the same caution applies to the survival curves, whose between-arm contrasts (7-day − 3.6, 28-day − 8.8 percentage points) are more precise but remain vulnerable to the residual confounding discussed above.

Clinically, our findings do not overturn the traditional caution—they reinforce it, while making clear that the existing observational database evidence cannot settle the question. PADIS guidelines recommend minimizing benzodiazepine exposure in critically ill adults generally [7]; in cirrhosis, where GABAergic neurotoxicity has a specific mechanistic rationale [10, 13] and impaired clearance prolongs drug effect [17], alternatives—propofol, dexmedetomidine, and where withdrawal dominates, carefully titrated protocols or phenobarbital [19]—deserve earlier consideration, consistent with contemporary guidance for critically ill patients with cirrhosis [6].

Limitations

This study has several limitations. (1) The primary outcome is a medication-based proxy: lactulose and rifaximin may be initiated for incompletely recorded indications (lactulose is also prescribed for constipation, and rifaximin has non-HE uses), dispensing timestamps approximate administration, and benzodiazepine sedation itself can mask the signs that trigger HE recognition—so ascertainment bias and outcome misclassification may operate in either direction. (2) Residual and unmeasured confounding—especially baseline encephalopathy grade, withdrawal severity, and sedation indication—remains the central threat; the negative controls were null, but the E-value (1.74) is within reach of plausible unmeasured confounders. (3) The exposed group mixes continuous sedation, single boluses, and alcohol-withdrawal protocol administrations; the CIWA-protocol sensitivity analysis removed only 33 stays and could not separate these strategies. (4) Subgroup analyses were not pre-specified, involve multiple comparisons, and are hypothesis-generating only. (5) Eligibility required ICU length of stay beyond 24 h—a post-time-zero event—so patients who died or were discharged within the first 24 h are excluded; this selection is shared by the outcome definition (HE treatment is only observable after 24 h) but means the emulated strategies apply to 24-hour survivors of ICU admission. (6) Single-center data without external validation limit generalizability. (7) Curve-endpoint risk differences carry cluster-bootstrap confidence intervals, but hazard-ratio, restricted-mean, and subgroup estimates are descriptive; the 90-day survival panel is a post hoc extension, and 28-day mortality has ± 1-day boundary imprecision from date-only records. (8) Analyses not performed: external validation, time-varying confounder adjustment (censoring models use baseline covariates only), and HE severity grading, which is not codable from MIMIC-IV.

Conclusions

Early ICU benzodiazepine exposure in cirrhosis was associated with more frequent short-term initiation of hepatic encephalopathy treatment, and this association was stable across all pre-specified design choices. Both negative-control outcomes were compatible with the null after identical weighting, providing no evidence of a residual care-intensity confounding pathway; residual confounding by

unmeasured severity, particularly agitation and withdrawal, nonetheless remains a live alternative explanation, especially for the larger estimates seen with intravenous exposure and in sicker subgroups. The cautious use of benzodiazepines in cirrhosis remains justified on existing grounds; adequately controlled studies—ideally with protocolized sedation comparisons—are needed before any causal claim, in either direction, is made.

Tables

Table S1: Complete medication, laboratory, and ICD-10 code audit: every item ID, pharmacy/eMAR name, and ICD code considered, with row counts in the source tables, including entries checked but containing no records (e.g., rifaximin has no inputevents entries; statins have no inputevents entries).

Declarations

Ethics approval and consent to participate

The institutional review board (IRB) that approved the establishment and sharing of MIMIC-IV is the Institutional Review Board of the Beth Israel Deaconess Medical Center (Boston, MA, USA). The IRB of the authors' institution is the Ethics Committee of Zhejiang Rongjun Hospital (Jiaxing, Zhejiang, China). MIMIC-IV is a de-identified database of routinely collected clinical data, released through PhysioNet to credentialed users under a data use agreement; it was accessed under PhysioNet credentialed access (approved user: Chudi Huang, huangchudi2026), and all patient records were fully de-identified before release. Because this retrospective analysis used only fully de-identified records obtained from that existing approved repository, it did not require separate ethics approval, and no approval or reference number applies. The study was conducted in accordance with the Declaration of Helsinki (2024 revision), all methods were performed in accordance with the relevant guidelines and regulations, and individual informed consent was not required and was not obtained.

Consent for publication

Not applicable.

Abbreviations

key

C.H. Chudi Huang
Q.X. Qihuan Xia
B.D. Bin Du
Y.L. Yue Lu
M.D. Miao Ding
H.S. Hong Shao
S.X. Sheng Xu.

Declarations

Results

Among 1,197 eligible ICU stays, 340 (28.4%) received a benzodiazepine within 24 h of ICU admission. After IPCW weighting, HE treatment initiation occurred in 34.6% of the benzodiazepine group versus 28.3% of controls (risk difference + 6.3 percentage points, 95% CI 0.7 to 13.2; weighted odds ratio 1.34), and 28-day survival was 74.8% versus 83.6% (risk difference − 8.8 percentage points, 95% CI − 15.9 to − 2.9; weighted Cox hazard ratio 1.65). The estimate was directionally stable across grace periods, outcome definitions, weight-truncation percentiles, and covariate-omission analyses, and adjustment approximately halved the crude association (+ 12.5 to + 6.3 percentage points). Pre-specified negative-control outcomes—new proton-pump inhibitor (risk difference + 3.9 percentage points, 95% CI − 3.4 to 11.1) and new statin initiation (− 0.3 percentage points, 95% CI − 4.6 to 4.7)—both yielded confidence intervals compatible with the null, and covariate balance was restored after weighting (all standardized mean differences < 0.11).

Conclusion

In this first strategy-level emulation of the question, early ICU benzodiazepine exposure was associated with a modestly higher rate of short-term hepatic encephalopathy treatment initiation that persisted after clone-censor-weight adjustment, with negative-control outcomes compatible with the null and restored covariate balance. The findings remain observational, do not establish causation with certainty, and support continued caution with benzodiazepines in cirrhosis.

Competing interests

The authors declare that they have no competing interests.

Funding

This study received no specific funding.

Author Contribution

C.H. conceptualised the study, performed the formal analysis, developed the methodology and software, validated the results, and wrote the original draft. Q.X. curated the data, conducted the investigation, and contributed to the methodology. B.D. contributed to the investigation and validation. Y.L. curated the data and prepared the figures. M.D. provided resources and contributed to the investigation. H.S. supervised

the study and administered the project. S.X. conceptualised the study, supervised the work, and reviewed and edited the manuscript. All authors read and approved the final manuscript.

Acknowledgement

We thank the MIT Laboratory for Computational Physiology and the PhysioNet team for maintaining the MIMIC-IV database.

Data Availability

MIMIC-IV v3.1 is a de-identified, publicly available database distributed through PhysioNet (https://physionet.org/content/mimiciv/3.1/) to credentialed users under a data use agreement; it is not redistributed here. The analysis code, cohort definition files, and the aggregate result artefacts that support every statistic reported in this article are provided as Supplementary Information.

References

1. Wang Y, Wang M, Liu C, Chen Y, Pan Y. Global burden of liver cirrhosis 1990–2019 and 20 years forecast: results from the global burden of disease study 2019. Ann Med. 2024;56:2328521. 10.1080/07853890.2024.2328521.
2. Poordad FF. Review article: the burden of hepatic encephalopathy. Aliment Pharmacol Ther. 2007;25(Suppl 1):3–9. 10.1111/j.1746-6342.2006.03215.x.
3. Moon AM, Kim HP, Jiang Y, Schladenhauffen R, Tapper EB, Lok ASF. Systematic review and meta-analysis on the effects of lactulose and rifaximin on patient-reported outcomes in hepatic encephalopathy. Am J Gastroenterol. 2023;118:284–93. 10.14309/ajg.0000000000002008.
4. Wehler M, Kokoska J, Reulbach U, Hahn EG, Denzer C. Short-term prognosis in critically ill patients with cirrhosis assessed by prognostic scoring systems. Hepatology. 2001;34:255–61. 10.1053/jhep.2001.26522.
5. Kamath PS, Wiesner RH, Malinchoc M, Kremers W, Therneau TM, Kosberg CL, et al. A model to predict survival in patients with end-stage liver disease. Hepatology. 2001;33:464–70. 10.1053/jhep.2001.22172.
6. Karvellas CJ, Bajaj JS, Kamath. P S AASLD practice guidance on acute-on-chronic liver failure and the management of critically ill patients with cirrhosis. Hepatology. 2024;79:1463–502. 10.1097/HEP.0000000000000671.
7. Devlin JW, Skrobik Y, Gélinas C, Needham DM, Slooter AJC, Pandharipande PP, et al. Clinical practice guidelines for the prevention and management of pain, agitation/sedation, delirium, immobility, and sleep disruption in adult patients in the ICU. Crit. Care Med. 2018;46:e825–73. 10.1097/CCM.0000000000003299.

8. Reves JG, Fragen RJ, Vinik HR, Greenblatt DJ. Midazolam: pharmacology and uses. Anesthesiology. 1985;62:310–24. 10.1097/00000542-198503000-00017.
9. Basile AS, Hughes RD, Harrison PM, Murata Y, Pannell L, Jones EA, et al. Elevated brain concentrations of 1,4-benzodiazepines in fulminant hepatic failure. N Engl J Med. 1991;325:473–8. 10.1056/NEJM199108153250705.
10. Basile AS, Jones EA. The involvement of benzodiazepine receptor ligands in hepatic encephalopathy. Hepatology. 1994;20:541–3. 10.1016/0270-9139(94)90217-8.
11. Basile AS, Jones EA, Skolnick P. The pathogenesis and treatment of hepatic encephalopathy: evidence for the involvement of benzodiazepine receptor ligands. Pharmacol Rev. 1991b;43:27–71. 10.1016/S0031-6997(25)06659-1.
12. Yurdaydin C, Gu ZQ, Nowak G, Fromm C, Holt AG, Basile AS. Benzodiazepine receptor ligands are elevated in an animal model of hepatic encephalopathy: relationship between brain concentration and severity of encephalopathy. J Pharmacol Exp Ther. 1993;265:565–71. 10.1016/s0022-3565(25)38171-1.
13. Heidari R. Brain mitochondria as potential therapeutic targets for managing hepatic encephalopathy. Life Sci. 2019;218:65–80. do.
14. Goh ET, Andersen ML, Morgan MY, Gluud LL. Flumazenil versus placebo or no intervention for people with cirrhosis and hepatic encephalopathy. Cochrane Database Syst Rev. 2017;8:CD002798. 10.1002/14651858.CD002798.pub4.
15. Vilstrup H, Amodio P, Bajaj J, Cordoba J, Ferenci P, Mullen KD, et al. Hepatic encephalopathy in chronic liver disease: 2014 practice guideline by the American Association for the Study of Liver Diseases and the European Association for the Study of the Liver. Hepatology. 2014;60:715–35. 10.1002/hep.27210.
16. Gershkovich P, Wasan KM, Ribeyre C, Moreau X. Effect of variations in treatment regimen and liver cirrhosis on exposure to benzodiazepines during treatment of alcohol withdrawal syndrome. Drugs Context. 2015;4:212287. 10.7573/dic.212287.
17. Ozbey AC, Keemink J, Wagner B, Schaefer HG, Lehr T. Physiologically based pharmacokinetic modeling to predict the impact of liver cirrhosis on glucuronidation via UGT1A4 and UGT2B7/2B4—a case study with midazolam. Drug Metab Dispos. 2024;52:614–25. 10.1124/dmd.123.001635.
18. Girard TD, Exline MC, Carson SS, Hough CL, Rock P, Gong MN, et al. Haloperidol and ziprasidone for treatment of delirium in critical illness. N Engl J Med. 2018;379:2506–16. 10.1056/NEJMoa1808217.
19. Govalan R, Hao S, Kenes MT, Scott SD, Szunar A. Safety and effectiveness of phenobarbital vs. benzodiazepines for severe alcohol withdrawal in alcohol-associated liver disease in the ICU. JHEP Rep. 2026;8:101769. 10.1016/j.jhepr.2026.101769.
20. Ingebrigtson M, Schang S, Hyzy RC, McSparron JI. Phenobarbital versus benzodiazepines for alcohol withdrawal in the medical intensive care unit: a retrospective cohort study. J Pharm Pract. 2025;38:231–6. 10.1177/08971900241273144.

21. Salomoni M, Missanelli A, Crescioli G, Vannacci A, Lombardi N, Maggi L, et al. Real-world analysis on the use of gamma-hydroxybutyric acid for alcohol withdrawal syndrome in hospitalized patients with diagnosis of cirrhosis. Intern Emerg Med. 2025;20:119–29. 10.1007/s11739-024-03761-x.
22. Hernán MA, Robins JM. Using big data to emulate a target trial when a randomized trial is not available. Am J Epidemiol. 2016;183:758–64. 10.1093/aje/kwv254.
23. Manzanilla A, Brunetta C, Peyre-Pradat F, Filion KB, Suissa S. Use of clone-censor-weight to avoid immortal-time bias: a systematic methodological review. J Clin Epidemiol. 2026;199:112468. 10.1016/j.jclinepi.2026.112468.
24. Gaber CE, Ghazarian AA, Strassle PD, Murray EJ. De-mystifying the clone-censor-weight method for causal research using observational data: a primer for cancer researchers. Cancer Med. 2024;13:e70461. 10.1002/cam4.70461.
25. Dickerman BA, García-Albéniz X, Logan RW, Denaxas S, Hernán. M A Avoidable flaws in observational analyses: an application to statins and cancer. Nat Med. 2019;25:1601–6. 10.1038/s41591-019-0597-x.
26. Nicolaisen SK, Thomsen RW, Pedersen L. When to start, how to classify, and whom to include: a methodological assessment and illustrative example of the clone-censor-weight approach for evaluating intervention timing in observational studies. Clin Epidemiol. 2026;18:618946. 10.2147/CLEP.S618946.
27. Hernán MA. Methods of public health research — strengthening causal inference from observational data. N Engl J Med. 2021;385:1345–8. 10.1056/NEJMp2113319.
28. Lipsitch M, Tchetgen Tchetgen E, Cohen T. Negative controls: a tool for detecting confounding and bias in observational studies. Epidemiology. 2010;21:383–8. 10.1097/EDE.0b013e3181d61eeb.
29. Tchetgen Tchetgen EJ. The control outcome calibration approach for causal inference with unobserved confounding. Am J Epidemiol. 2014;179:633–40. 10.1093/aje/kwt303.
30. VanderWeele TJ, Ding P. Sensitivity analysis in observational research: introducing the E-value. Ann Intern Med. 2017;167:268–74. 10.7326/M16-2607.
31. 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;370:1453–7. 10.1016/s0140-6736(07)61602-x.
32. Cashin AG, Hansford HJ, Hernán MA, Lee H, Jones MD, Karelse IM, et al. Transparent reporting of observational studies emulating a target trial: the TARGET Statement. BMJ. 2025;390:e087179. 10.1136/bmj-2025-087179.
33. Cashin AG, Hansford HJ, Hernán MA, Lee H, Jones MD, Karelse IM, et al. Transparent reporting of observational studies emulating a target trial (TARGET): the TARGET guideline. PLoS Med. 2025b;22:e1004787. 10.1371/journal.pmed.1004787.
34. 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. 10.1038/s41597-022-01899-x.

35. MIT Laboratory for Computational. Physiology Guidelines for reviewers and authors of MIMIC data 2026.
36. Kim WR, Biggins SW, Kremers WK, Wiesner RH, Kamath PS, Benson JT, et al. Hyponatremia and mortality among patients on the liver-transplant waiting list. N Engl J Med. 2008;359:1018–26. 10.1056/NEJMoa0801209.
37. Qin G, Zou K. Early midazolam exposure and short-term risk of acute exacerbation of COPD in critically ill patients: a target trial emulation using the MIMIC-IV database. Front Artif Intell. 2026;9:1836581. 10.3389/frai.2026.1836581.
38. Conn HO, Leevy CM, Vlahcevic ZR, Rodgers JB, Maddrey WC, Seeff L, et al. Comparison of lactulose and neomycin in the treatment of chronic portal-systemic encephalopathy. A double blind controlled trial. Gastroenterology. 1977;72:573–83. PMID:14049.
39. Bass NM, Mullen KD, Sanyal A, Poordad F, Neff G, Leevy CB, et al. Rifaximin treatment in hepatic encephalopathy. N Engl J Med. 2010;362:1071–81. 10.1056/NEJMoa0907893.
40. Dickerman BA, García-Albéniz X, Logan RW, Hernán MA. Emulating a target trial in case-control designs: an application to statins and colorectal cancer. Int J Epidemiol. 2020;49:1637–46. 10.1093/ije/dyaa144.
41. Parienti JJ. Understanding the clone-censor-weight method in observational studies of antibiotic duration. Clin Infect Dis (online ahead print). 2026. 10.1093/cid/ciag388.

Figures

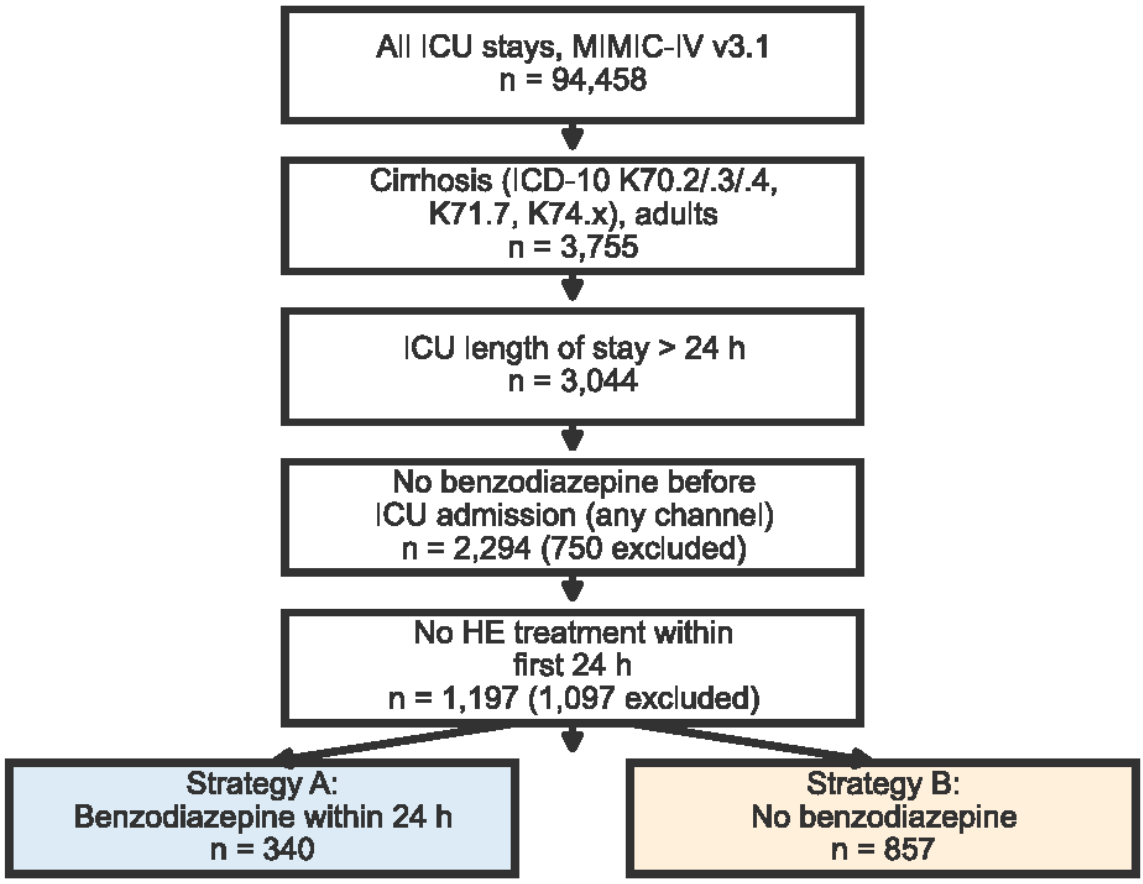


Figure 1

Cohort flow diagram. HE, hepatic encephalopathy; ICU, intensive care unit.

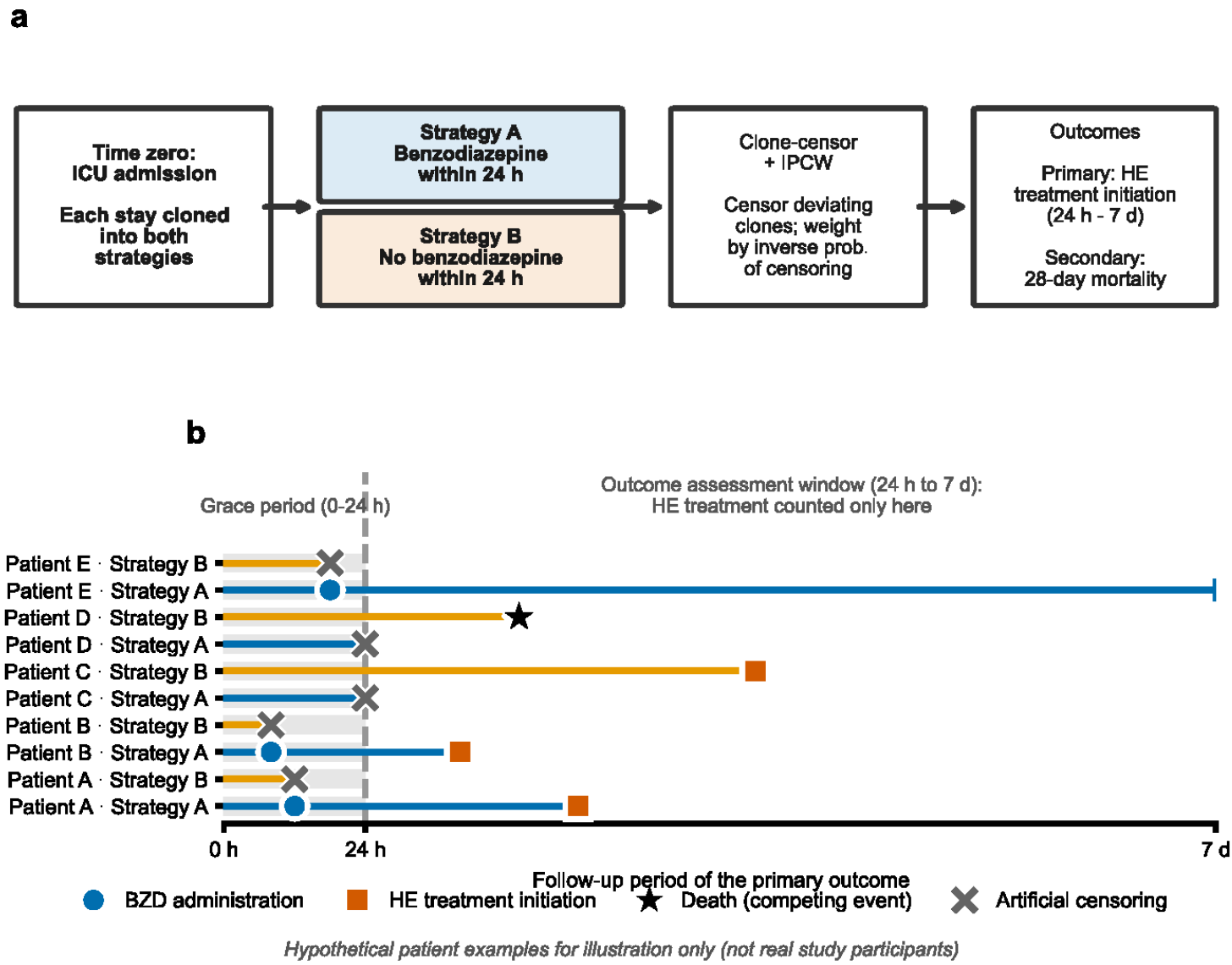


Figure 2

Schematic of target trial emulation with clone-censor and inverse-probability-of-censoring weighting (CCW-IPCW). (a) Analytical workflow. Time zero was defined as ICU admission. Each eligible ICU stay was cloned into two hypothetical strategies: benzodiazepine (BZD) within 24 h of admission (Strategy A) and no benzodiazepine within 24 h (Strategy B). Clones deviating from their assigned strategy were artificially censored, and uncensored clones were weighted by the inverse probability of censoring. The primary outcome was HE treatment initiation occurring between 24 h and 7 days; the secondary outcome was 28-day mortality. (b) Swimmer plot illustrating cloning and censoring rules using hypothetical patient examples (not real study data). The grace period was 0-24 h; HE outcomes were assessed only in the window from 24 h to 7 days. Strategy A clones were censored if no BZD dose was administered by 24 h; Strategy B clones were censored immediately if BZD was given within 24 h. Symbols: blue circle, BZD administration; orange square, HE treatment initiation; black star, death (competing event); cross, artificial censoring.

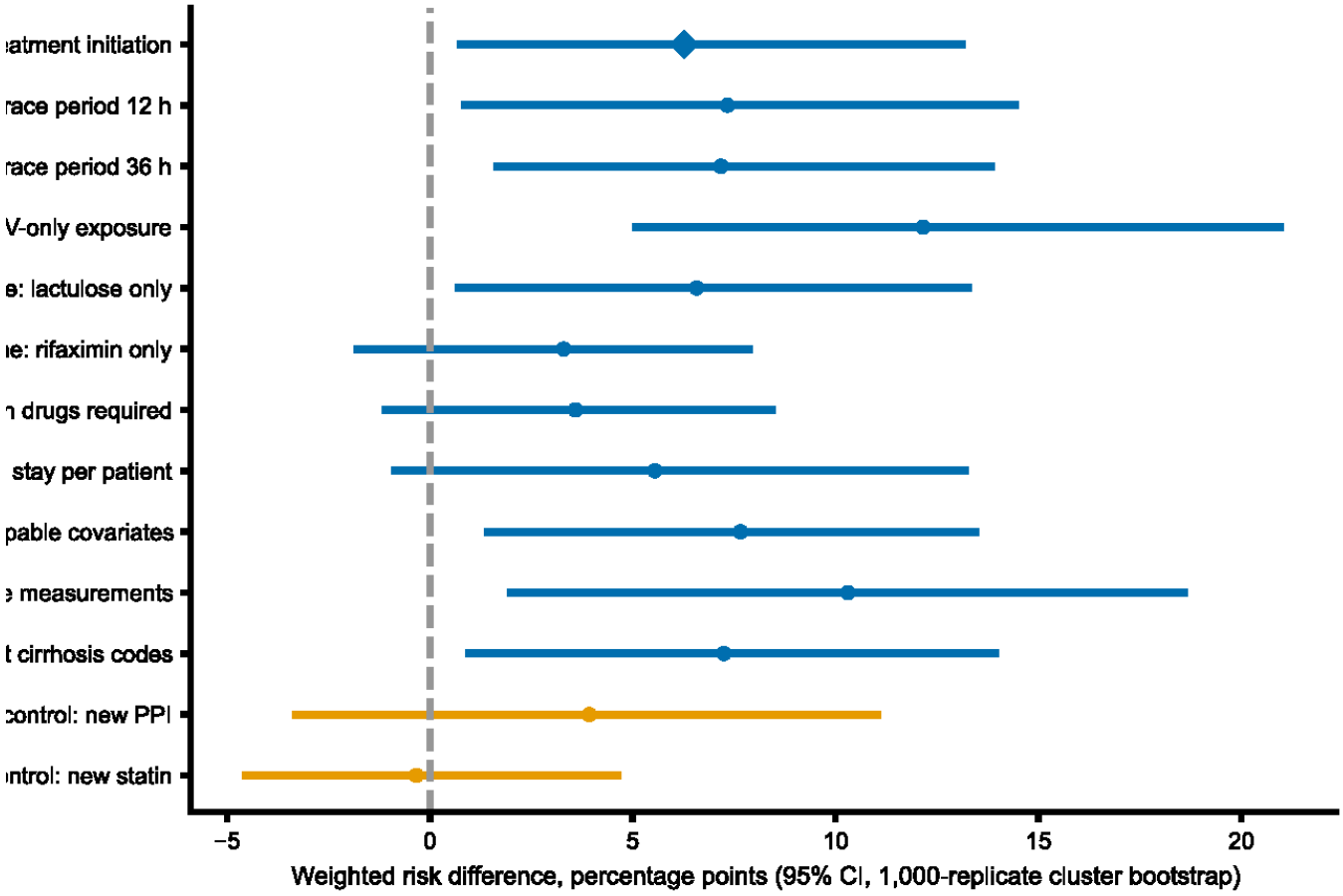


Figure 3

Primary, sensitivity, and negative-control estimates (weighted risk differences, percentage points). Squares: primary and sensitivity analyses; circles: negative-control outcomes (new proton pump inhibitor, new statin).

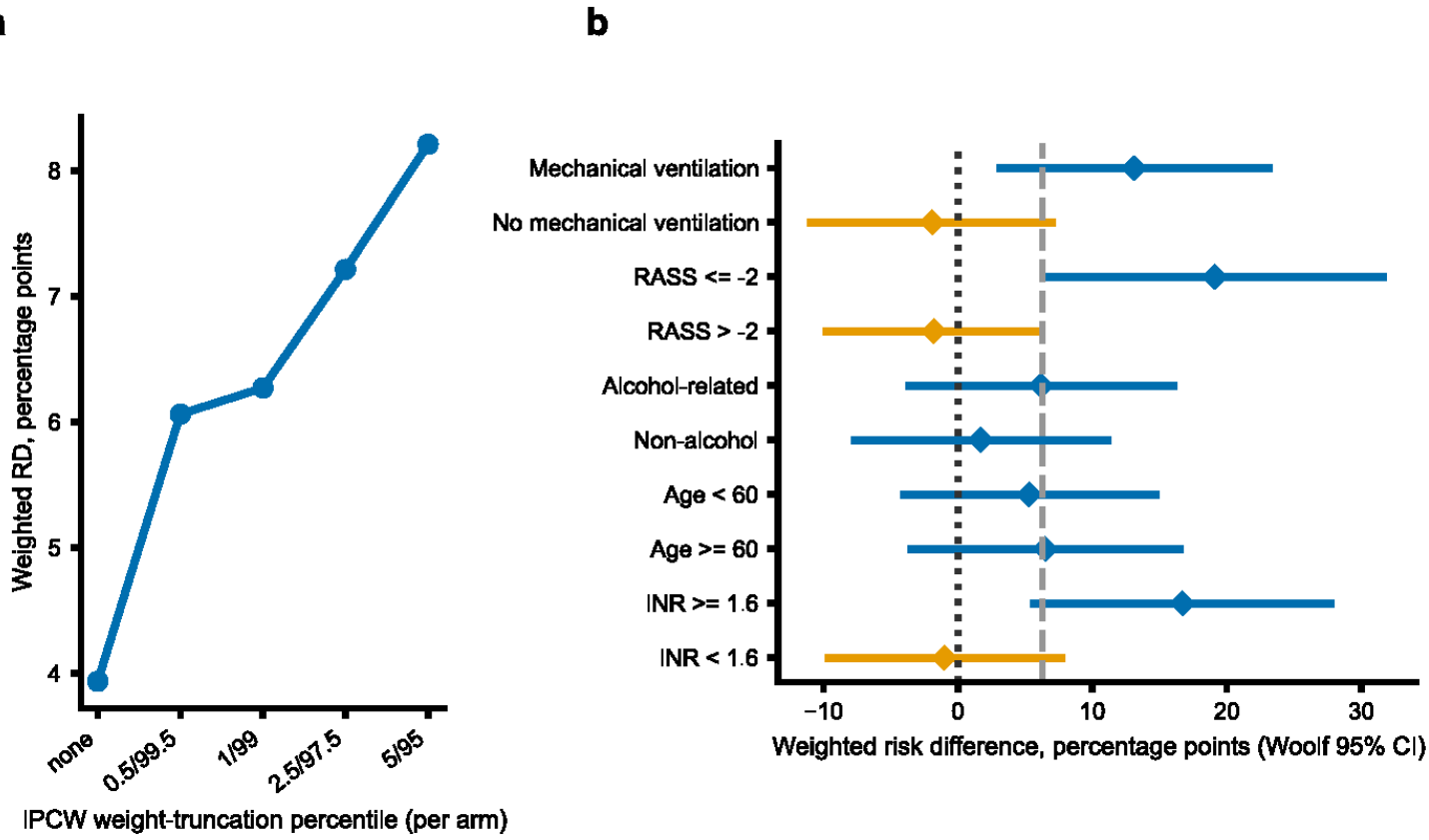


Figure 4

Weight diagnostics and exploratory subgroup analyses. (a) Weight-truncation sensitivity curve: the weighted risk difference remains positive across truncation percentiles from none to 5/95, indicating the estimate is not driven by extreme weights (point estimates). (b) Exploratory subgroup analyses: weighted risk differences for HE treatment initiation, stratified by pre-exposure variables (Woolf 95% CI from effective sample sizes; dashed line, overall estimate +6.3 percentage points; dotted line, null). Not pre-specified; hypothesis-generating only.

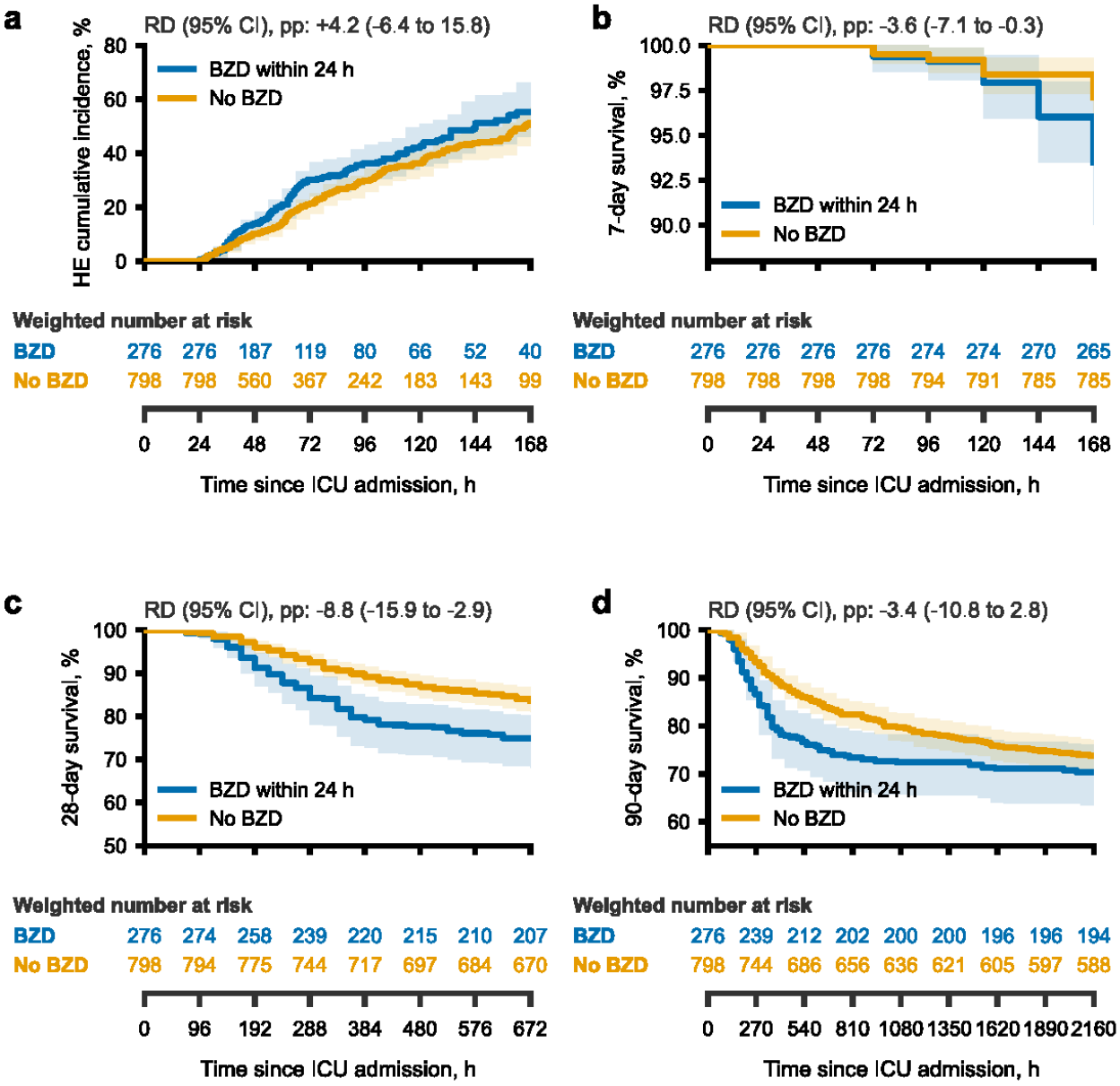


Figure 5

Time-to-event analyses under stabilised inverse-probability-of-censoring weights (IPCW). (a) Aalen-Johansen cumulative incidence of hepatic encephalopathy (HE) treatment initiation, with death before the outcome as a competing event and HE ascertainment censored at ICU discharge. (b) Weighted Kaplan-Meier 7-day survival. (c) Weighted Kaplan-Meier 28-day survival. (d) Weighted Kaplan-Meier 90-day survival (post hoc extension). Shaded bands are pointwise 95% confidence bands from 1,000-

replicate patient-level cluster bootstraps re-estimating the full clone-censor-weight chain; annotated risk differences (RD, BZD within 24 h minus no BZD) at the panel-specific horizon are from the same bootstrap. Weighted number-at-risk tables give the sum of stabilised IPCW weights of clones still at risk (not raw counts). The y-axes of panels b-d are truncated for visibility. The weighted Cox hazard ratio is reported as a point estimate only. BZD, benzodiazepine; h, hours; pp, percentage points.

Supplementary Files

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

- figS1weightdistribution.pdf
- figS2smdloveplot.pdf
- S1SAPEN.md
- S2deviationsEN.md
- tableS1medicationcodes.csv