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.

Deep versus light sedation and 28-day mortality in mechanically ventilated sepsis: a causal analysis of two intensive care databases

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

The Affiliated Taizhou People's Hospital of Nanjing Medical University

Article

Keywords: sepsis, deep sedation, mechanical ventilation, intensive care unit, mortality, causal inference

Posted Date: July 3rd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9738160/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

Almost all patients with sepsis who require invasive mechanical ventilation receive sedation, yet the consequences of the depth of that sedation in this population are poorly defined. We examined whether deep sedation during the first 48 hours of ventilation is associated with 28-day mortality in two large intensive care databases. Adults with sepsis were classified, from the median Richmond Agitation-Sedation Scale score, as receiving deep or light sedation, and follow-up began at the end of the 48-hour window. The effect on 28-day mortality was estimated using inverse probability of treatment weighting, with propensity score matching, multiple imputation, and an E-value as sensitivity analyses, and was tested for replication in an independent multicentre cohort. In the primary cohort of 8,477 patients, deep sedation was associated with higher 28-day mortality (odds ratio 2.37, 95% confidence interval 2.09 to 2.69; adjusted risk difference 15.4 percentage points), and with higher ICU and in-hospital mortality; the incidence of delirium, assessed in this cohort only, was also higher. The association was consistent across all sensitivity analyses and was reproduced in the validation cohort of 844 patients (odds ratio 2.01). Deep early sedation is associated with substantially worse survival in ventilated patients with sepsis.

Introduction

Sepsis, defined as life-threatening organ dysfunction caused by a dysregulated host response to infection [1], is among the most common reasons for admission to an intensive care unit (ICU) and a leading cause of death worldwide, with an estimated 11 million sepsis-related deaths each year [2]. A large proportion of patients with sepsis develop respiratory failure and require invasive mechanical ventilation, and almost all ventilated patients receive sedative medication to relieve discomfort, facilitate tolerance of the ventilator, and maintain safety. The depth of sedation that results is one of the few directly modifiable elements of early critical care, and the level maintained during the first hours and days of ventilation is determined as much by unit culture and clinician preference as by any fixed physiological requirement.

Over the past two decades, evidence has converged on the principle that sedation should be minimised. randomised trials of daily interruption of sedative infusions [3] and of a protocol of minimal or no sedation [4] have shown that lighter sedation shortens the duration of mechanical ventilation and ICU stay, and a randomised comparison of light with deep sedation found the lighter strategy to be feasible and without harm to subsequent psychological recovery [5]. Observational studies have extended these findings to harder outcomes: greater sedation intensity, and deep sedation in particular during the first 48 hours of ventilation, has been independently associated with delayed extubation, delirium, and increased short- and long-term mortality [6–8], a relationship summarized in a systematic review and meta-analysis [9]. On the basis of this evidence, current guidelines for the management of pain, agitation, and sedation in critically ill adults recommend that a light level of sedation be targeted whenever it is clinically tolerated, with depth monitored using a validated instrument such as the Richmond Agitation-Sedation Scale (RASS) [10, 11].

Whether these conclusions apply without qualification to patients with sepsis is, however, uncertain. Most evidence on sedation depth derives from general or mixed ICU populations, within which patients with sepsis form only a subgroup. Several features of sepsis may make the depth of sedation particularly consequential: these patients carry a high burden of acute brain dysfunction and delirium, which is itself associated with worse long-term outcomes [12]; they are haemodynamically fragile and therefore more exposed to the cardiovascular depressant effects of sedative agents; and they frequently require neuromuscular blockade and prolonged mechanical ventilation, both of which constrain how sedation can be delivered. Sepsis-associated encephalopathy is, moreover, influenced by potentially modifiable aspects of management [13]. Despite these considerations, sepsis-specific data on the consequences of sedation depth are sparse, and no randomised trial has been designed to test sedation depth in this population. The question is further complicated by strong confounding by indication: deeper sedation is preferentially administered to more severely ill patients, so that a direct comparison of outcomes between deeply and lightly sedated patients conflates the effect of sedation with that of the illness that prompted it. Isolating the effect of sedation depth itself therefore requires both rigorous adjustment for the determinants of treatment and replication in an independent population.

We addressed this question using two large, independent ICU databases and a causal-inference approach based on inverse probability of treatment weighting. Our aim was to estimate the effect of deep, as compared with light, sedation during the first 48 hours of invasive mechanical ventilation on 28-day mortality in adults with sepsis, and on the secondary outcomes of delirium and ICU and in-hospital mortality, and to determine whether any such effect was reproduced in an independent multicentre cohort.

Methods

Study design and data sources

We conducted a retrospective cohort study informed by the principles of the target trial emulation framework [14], in which an observational analysis is structured to approximate the protocol of a hypothetical randomised trial with respect to eligibility, the definition of the treatment strategies, time zero, and outcome assessment. The primary analysis used the Medical Information Mart for Intensive Care IV (MIMIC-IV, version 3.1) [15], a single-centre database containing the de-identified electronic health records of patients admitted to the Beth Israel Deaconess Medical Center, a tertiary academic hospital in Boston, Massachusetts, United States. External validation used the eICU Collaborative Research Database (eICU-CRD, version 2.0) [16], a multicentre database compiled from intensive care units (ICUs) at hospitals across the continental United States. Both databases were queried through Google BigQuery.

Both databases are publicly available and fully de-identified. Their establishment and the use of the data for research were approved by the institutional review boards of the Massachusetts Institute of Technology and, for MIMIC-IV, the Beth Israel Deaconess Medical Center, with a waiver of the

requirement for individual informed consent. Investigators completed the required human-subjects research training and signed the data use agreement before access. The present study used only de-identified data and therefore did not require additional ethical review. Reporting follows the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) recommendations [17].

Study population and cohort derivation

Eligible patients were adults (aged at least 18 years) who met the Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3) [1] and who received invasive mechanical ventilation initiated within 24 hours of ICU admission. To allow the exposure to be defined over a fixed early window, we additionally required an ICU length of stay of at least 48 hours and at least six Richmond Agitation-Sedation Scale (RASS) [11] assessments recorded during the first 48 hours of invasive ventilation. In MIMIC-IV, sepsis was identified using the database's pre-computed Sepsis-3 definition, which operationalizes suspected infection together with an acute increase of at least two points in the Sequential Organ Failure Assessment (SOFA) score [18]. Because eICU-CRD does not provide a pre-computed Sepsis-3 definition, sepsis in the validation cohort was identified from International Classification of Diseases (ICD) codes for sepsis and septic shock supplemented by recorded diagnosis strings; the full code list is provided in the Supplementary Methods.

Patients admitted to a dedicated neurological or neurosurgical intensive care unit were excluded from both cohorts, because in this population a depressed level of consciousness, and hence the RASS, may reflect the underlying brain injury rather than the depth of pharmacological sedation. For patients with more than one qualifying ICU admission, only the first admission was analysed, so that each patient contributed a single observation. The derivation of both cohorts is summarized in Fig. 1.

Exposure

The exposure of interest was the sedation strategy applied during the first 48 hours of invasive mechanical ventilation. For each patient, we extracted all RASS scores documented during this window and calculated the median. Patients were classified as receiving deep sedation if the median RASS score was − 3 or lower, and as receiving light sedation if the median RASS score was between − 2 and 0 inclusive. The median was chosen, in preference to a single time point or the mean, because it summarizes the sustained level of sedation over the window while being robust to transient excursions, such as brief deepening for a procedure. Patients whose median RASS score was above 0 were predominantly agitated rather than sedated and were considered a clinically distinct group; together with patients whose median value fell between the deep and light strata, in the interval (− 3, − 2), they were not included in the primary comparison, so as to ensure a clear contrast between the two sedation strategies. Time zero for all analyses was defined as 48 hours after the initiation of invasive mechanical ventilation—the point at which the exposure was fully ascertained—so that follow-up began only after the exposure window had elapsed, thereby avoiding immortal time bias.

Outcomes

The primary outcome was 28-day all-cause mortality, measured from time zero. Secondary outcomes were the occurrence of delirium, defined as at least one positive Confusion Assessment Method for the ICU (CAM-ICU) [19] assessment after time zero and assessed in the MIMIC-IV cohort only, because CAM-ICU data were not uniformly recorded in eICU-CRD; ICU mortality; and in-hospital mortality. Vital status and the timing of death were obtained from the structured death and discharge records of each database.

Covariates

Baseline covariates were selected a priori as variables plausibly associated with both the choice of sedation depth and mortality, and were ascertained at or before time zero. They comprised age, sex, and race or ethnicity; the type of ICU; severity of illness expressed as the SOFA score and the Simplified Acute Physiology Score II (SAPS-II) [20], each taken as the worst value within the first 48 hours; the Charlson Comorbidity Index [21]; the worst values within the first 48 hours of serum lactate, serum creatinine, haematocrit, and white blood cell count; the maximum dose of vasopressors expressed as a norepinephrine-equivalent infusion rate [22]; and the presence of septic shock. For the primary MIMIC-IV analysis, the covariate set was extended to capture concomitant treatments administered during the exposure window that may influence both sedation depth and prognosis: the use of each of the principal sedative agents (propofol, midazolam, dexmedetomidine, and lorazepam), the use and cumulative duration of neuromuscular blocking agents, and the use of systemic corticosteroids.

In the eICU-CRD validation cohort, the records of continuously infused medications were found to substantially under-capture sedative and neuromuscular blocking agent administration, with documented exposure frequencies that were not clinically plausible; these variables were therefore not used. The propensity score model for the validation cohort was instead restricted to covariates that were reliably recorded in eICU-CRD: age, sex, the Acute Physiology and Chronic Health Evaluation (APACHE) IVa score [23], the worst values of lactate, creatinine, haematocrit, and white blood cell count, the use of vasopressors, the use of renal replacement therapy, and admission to a medical or medical-surgical ICU. Race or ethnicity was not included in the validation model because its recording is not standardized across the contributing eICU-CRD hospitals. The specific data elements and item identifiers used to construct every covariate in each database are listed in Supplementary Table S4.

Statistical analysis

Continuous variables are summarized as the mean with standard deviation or, for skewed distributions, as the median with interquartile range, and categorical variables as counts with percentages. Baseline differences between the deep and light sedation groups were quantified using the absolute standardized mean difference (SMD), with a value below 0.1 taken to indicate negligible imbalance.

The primary analysis estimated the effect of deep versus light sedation using inverse probability of treatment weighting (IPTW) [24]. A propensity score—the conditional probability of receiving deep sedation given the baseline covariates—was estimated by multivariable logistic regression including all covariates described above. Missing values, which were confined to the laboratory covariates and, in eICU-CRD, the APACHE IVa score, were imputed using chained equations with predictive mean matching [25]; the primary analysis used a single imputed dataset, and a prespecified sensitivity analysis (described below) repeated the entire analysis across five multiply imputed datasets with estimates combined by Rubin's rules. Stabilized weights targeting the average treatment effect in the whole population were then computed, and to limit the influence of extreme weights they were truncated at the 99th percentile. The positivity assumption was examined by inspecting the overlap of the estimated propensity score distributions between the deep and light sedation groups (Supplementary Figure S1). Covariate balance after weighting was assessed by recomputing the SMD for every covariate and by visual inspection of a Love plot (Supplementary Figure S2). The effect of deep sedation on each binary outcome was estimated in the weighted sample using a logistic regression model containing the exposure as the sole covariate, yielding an odds ratio (OR); the effect on time to death within 28 days was estimated using a weighted Cox proportional hazards model, yielding a hazard ratio (HR); and the absolute difference in 28-day mortality was estimated using a weighted linear model, yielding a risk difference (RD). The proportional hazards assumption underlying the Cox model was formally tested using scaled Schoenfeld residuals; because it was not satisfied (global test $P < 0.001$), the hazard ratio is interpreted as an average effect over the 28-day follow-up period, and the principal effect measures for the primary outcome are the odds ratio and the absolute risk difference, neither of which depends on this assumption. Confidence intervals for all weighted estimates were obtained using a robust (sandwich) variance estimator that accounts for the weighting.

Three sensitivity analyses were prespecified to probe the robustness of the primary estimate. First, as an alternative approach to confounding adjustment, 1:1 nearest-neighbour propensity score matching without replacement was performed using a caliper of 0.2 standard deviations of the logit of the propensity score, and balance was reassessed in the matched sample. Second, as a sensitivity analysis for the handling of missing data, the full IPTW analysis was repeated across five datasets multiply imputed by chained equations under a missing-at-random assumption, and the resulting estimates were combined using Rubin’s rules. Third, an E-value [26] was calculated for both the point estimate and the confidence limit closer to the null; the E-value expresses the minimum strength of association, on the risk ratio scale, that an unmeasured confounder would need to have with both the exposure and the outcome, over and above the measured covariates, to fully explain away the observed association.

Effect modification was examined in four prespecified subgroups: age (below 65 years versus 65 years or older), SOFA score category (6 or below, 7 to 10, and above 10), the presence of septic shock, and the use of neuromuscular blocking agents. The IPTW analysis of 28-day mortality was repeated within each subgroup, and statistical interaction was assessed by adding a multiplicative exposure-by-subgroup term to the weighted outcome model. External validation repeated the IPTW analysis of 28-day, ICU, and

in-hospital mortality in the eICU-CRD cohort using the parsimonious propensity score model described above, with multiple imputation applied in the same manner as the primary analysis.

All analyses were performed in R (version 4.5.2; R Foundation for Statistical Computing) using the WeightIt and cobalt packages for weighting and balance assessment, MatchIt for propensity score matching [27], survey and survival for weighted outcome models, mice for imputation by chained equations, and EValue for the sensitivity analysis. All tests were two-sided and a P value below 0.05 was considered statistically significant. Further technical detail is provided in the Supplementary Methods.

Results

Cohort derivation

In MIMIC-IV, 10,297 ICU stays belonging to 9,199 unique patients met the criteria of sepsis, invasive mechanical ventilation initiated within 24 hours of admission, an ICU stay of at least 48 hours, and at least six RASS assessments during the first 48 hours of ventilation. After retaining only the first qualifying admission for each patient, 9,199 patients remained. Classification by the median RASS score over the first 48 hours of ventilation identified 2,977 patients with deep sedation, 5,833 with light sedation, 215 with predominant agitation, and 174 whose median value fell between the deep and light strata. After exclusion of the 389 patients who were agitated or of intermediate sedation depth, and of 333 patients admitted to a neurological or neurosurgical ICU, the primary analytic cohort comprised 8,477 patients, of whom 2,828 (33.4%) had received deep sedation and 5,649 (66.6%) light sedation (Fig. 1).

In eICU-CRD, 919 qualifying ICU stays were identified; after retention of the first admission per patient, 881 patients remained, comprising 237 with deep sedation, 627 with light sedation, 7 with predominant agitation, and 10 of intermediate sedation depth. After exclusion of the 17 patients who were agitated or of intermediate sedation depth, and of a further 20 patients admitted to a neurological intensive care unit, the validation cohort comprised 844 patients, of whom 226 (26.8%) had received deep sedation and 618 (73.2%) light sedation, drawn from 35 hospitals.

Baseline characteristics

Baseline characteristics of the primary cohort are shown in Table 1. Patients who received deep sedation differed substantially from those who received light sedation across nearly every domain. They were younger (mean age 60.7 versus 64.7 years) and slightly more often male (62.9% versus 57.4%). They were considerably more severely ill: the mean SOFA score was 9.6 versus 7.1 (SMD 0.68) and the mean SAPS-II was 48.7 versus 42.6 (SMD 0.42). Markers of circulatory and metabolic derangement were correspondingly more deranged in the deep sedation group, with a higher median serum lactate (3.0 versus 1.9 mmol/L, SMD 0.59), a higher median creatinine, a higher norepinephrine-equivalent vasopressor dose, and a greater frequency of septic shock (74.3% versus 66.7%). The comorbidity

burden, by contrast, was similar in the two groups (median Charlson Comorbidity Index 5 in both; SMD 0.08).

Table 1
Baseline characteristics of the primary cohort (MIMIC-IV), according to sedation strategy.

Characteristic	Light sedation (n = 5,649)	Deep sedation (n = 2,828)	SMD
Age, years, mean (SD)	64.7 (15.3)	60.7 (16.6)	0.25
Male sex, n (%)	3,244 (57.4)	1,779 (62.9)	0.11
Race or ethnicity, n (%)			0.24
White	3,741 (66.2)	1,563 (55.3)	
Black	483 (8.6)	278 (9.8)	
Hispanic	174 (3.1)	107 (3.8)	
Asian	148 (2.6)	78 (2.8)	
Other or unknown	1,103 (19.5)	802 (28.4)	
ICU type, n (%)			0.32
Coronary care unit	403 (7.1)	287 (10.1)	
Cardiac vascular ICU	1,509 (26.7)	497 (17.6)	
Medical ICU	1,337 (23.7)	942 (33.3)	
Medical / surgical ICU	693 (12.3)	337 (11.9)	
Surgical ICU	824 (14.6)	456 (16.1)	
Trauma surgical ICU	883 (15.6)	308 (10.9)	
Other	0 (0.0)	1 (0.0)	
SOFA score, mean (SD)	7.1 (3.1)	9.6 (4.1)	0.68
SAPS-II, mean (SD)	42.6 (12.8)	48.7 (16.4)	0.42
Charlson Comorbidity Index, median (IQR)	5 (3–7)	5 (3–7)	0.08
Septic shock, n (%)	3,769 (66.7)	2,102 (74.3)	0.17
Worst value within first 48 h			

Continuous variables are presented as mean (standard deviation, SD) or median (interquartile range, IQR); categorical variables as number (percentage). The standardized mean difference (SMD) is the crude, unweighted value; for categorical variables with more than two levels a single multilevel SMD is reported. SMDs for the medication rows correspond to the binary use indicator. Opioid use is shown for descriptive purposes only and was not included in the propensity score model; the cumulative duration of neuromuscular blockade, which was included as a covariate, is summarized in Supplementary Table S1. ICU, intensive care unit; SOFA, Sequential Organ Failure Assessment; SAPS-II, Simplified Acute Physiology Score II.

Characteristic	Light sedation (n = 5,649)	Deep sedation (n = 2,828)	SMD
Lactate, mmol/L, median (IQR)	1.9 (1.3–3.0)	3.0 (1.8–5.5)	0.59
Creatinine, mg/dL, median (IQR)	1.2 (0.8–1.8)	1.5 (1.0–2.7)	0.24
Haematocrit, %, mean (SD)	27.8 (5.3)	28.3 (6.3)	0.09
White cell count, ×10⁹/L, mean (SD)	16.2 (10.3)	17.8 (12.8)	0.14
Norepinephrine-equivalent, µg/kg/min, median (IQR)	0.10 (0–0.21)	0.20 (0–0.50)	0.14
Treatment during exposure window, n (%)			
Propofol	4,534 (80.3)	2,208 (78.1)	0.05
Midazolam	1,629 (28.8)	1,343 (47.5)	0.39
Dexmedetomidine	2,831 (50.1)	1,986 (70.2)	0.42
Lorazepam	1,046 (18.5)	341 (12.1)	0.18
Opioid	4,516 (79.9)	2,145 (75.8)	0.10
Neuromuscular blocking agent	622 (11.0)	1,191 (42.1)	0.75
Systemic corticosteroid	498 (8.8)	471 (16.7)	0.24

Continuous variables are presented as mean (standard deviation, SD) or median (interquartile range, IQR); categorical variables as number (percentage). The standardized mean difference (SMD) is the crude, unweighted value; for categorical variables with more than two levels a single multilevel SMD is reported. SMDs for the medication rows correspond to the binary use indicator. Opioid use is shown for descriptive purposes only and was not included in the propensity score model; the cumulative duration of neuromuscular blockade, which was included as a covariate, is summarized in Supplementary Table S1. ICU, intensive care unit; SOFA, Sequential Organ Failure Assessment; SAPS-II, Simplified Acute Physiology Score II.

Marked differences were also present in concomitant treatment. Deep sedation was accompanied far more often by the use of neuromuscular blocking agents (42.1% versus 11.0%; SMD 0.75), and among recipients the median cumulative duration of neuromuscular blockade was longer (36 versus 20 hours). Midazolam was used more frequently with deep sedation (47.5% versus 28.8%), as was dexmedetomidine (70.2% versus 50.1%); lorazepam, by contrast, was used less often with deep sedation (12.1% versus 18.5%), and the use of propofol and of opioids was broadly similar between groups. Systemic corticosteroids were administered to 16.7% of the deep sedation group and 8.8% of the light sedation group. Taken together, these differences indicate strong confounding by indication—deeper sedation was preferentially applied to more severely ill patients—and provide the rationale for the causal inference approach. Many covariates were substantially imbalanced before adjustment, most notably the use and cumulative duration of neuromuscular blockade, the SOFA score, and serum lactate (Table 1, Supplementary Table S1).

Primary outcome

Crude 28-day mortality was 36.2% (1,024 of 2,828) among patients who received deep sedation, compared with 14.5% (820 of 5,649) among those who received light sedation. After IPTW, all 30 covariates were well balanced between the groups, with every standardized mean difference reduced below the 0.1 threshold; the largest residual imbalance, for the cumulative duration of neuromuscular blockade, was 0.082 (Supplementary Table S1, Supplementary Figure S2). Weighting reduced the effective sample size to approximately 1,889 in the deep sedation group and 4,682 in the light sedation group.

In the weighted analysis, deep sedation was associated with a substantially higher risk of 28-day mortality, with an odds ratio of 2.37 (95% confidence interval [CI] 2.09 to 2.69); on the absolute scale, the adjusted risk difference was 15.4 percentage points (95% CI 13.2 to 17.7). A weighted Cox model gave an average hazard ratio of 2.22 (95% CI 2.00 to 2.47) over the 28-day period; because the proportional hazards assumption was not met (global test $P < 0.001$; Supplementary Methods), this is interpreted as a time-averaged effect rather than a constant one, and the inference for the primary outcome rests on the odds ratio and the risk difference. The IPTW-adjusted Kaplan-Meier curves showed an early separation in survival between the two strategies that persisted throughout follow-up (Fig. 2).

Secondary outcomes

Deep sedation was associated with worse results for every secondary outcome (Table 2). The occurrence of delirium was more frequent with deep sedation (crude incidence 59.3% versus 38.4%; IPTW-adjusted OR 1.84, 95% CI 1.65 to 2.05). ICU mortality was 29.4% with deep sedation versus 8.1% with light sedation (adjusted OR 2.79, 95% CI 2.40 to 3.24), and in-hospital mortality was 36.2% versus 12.2% (adjusted OR 2.72, 95% CI 2.39 to 3.11).

Table 2
Association between deep sedation and clinical outcomes in the primary cohort, sensitivity analyses, and the external validation cohort.

Outcome or analysis	Deep sedation, n/N (%)	Light sedation, n/N (%)	Effect estimate (95% CI)	P value
Primary cohort (MIMIC-IV), IPTW-adjusted				
28-day mortality	1,024/2,828 (36.2)	820/5,649 (14.5)	OR 2.37 (2.09–2.69)	< 0.001
28-day mortality	–	–	HR 2.22 (2.00–2.47)	< 0.001
28-day mortality	–	–	RD 15.4% (13.2–17.7)	< 0.001
In-hospital mortality	1,023/2,828 (36.2)	690/5,649 (12.2)	OR 2.72 (2.39–3.11)	< 0.001
ICU mortality	831/2,828 (29.4)	455/5,649 (8.1)	OR 2.79 (2.40–3.24)	< 0.001
Delirium	1,677/2,828 (59.3)	2,168/5,649 (38.4)	OR 1.84 (1.65–2.05)	< 0.001
Sensitivity analyses, 28-day mortality				
Propensity score matching (2,186 pairs)	–	–	OR 2.31 (2.00–2.65)	< 0.001
Multiple imputation (5 datasets, pooled)	–	–	OR 2.37 (2.09–2.70)	< 0.001
E-value (point estimate / lower 95% CI)	–	–	2.45 / 2.25	–
External validation (eICU-CRD, n = 844)				
28-day mortality	85/226 (37.6)	130/618 (21.0)	OR 2.01 (1.42–2.87)	< 0.001

All effect estimates compare deep sedation with light sedation (reference). For binary outcomes the effect is an odds ratio (OR) from a weighted or matched logistic model; for 28-day mortality the hazard ratio (HR) from a weighted Cox model and the absolute risk difference (RD) from a weighted linear model are additionally shown. n/N (%) denotes the number of events, the number at risk, and the crude percentage; a dash indicates that an event count is not applicable to that row. Confidence intervals (CI) for weighted estimates use a robust variance estimator. The E-value quantifies the minimum association an unmeasured confounder would need with both exposure and outcome to explain away the result. The proportional hazards assumption underlying the Cox model was not satisfied (Supplementary Methods); the hazard ratio is therefore an average effect over the 28-day period, and the odds ratio and risk difference are the principal measures for the primary outcome. IPTW, inverse probability of treatment weighting; ICU, intensive care unit; eICU-CRD, eICU Collaborative Research Database.

Outcome or analysis	Deep sedation, n/N (%)	Light sedation, n/N (%)	Effect estimate (95% CI)	P value
28-day mortality, multiple imputation	–	–	OR 1.96 (1.37–2.79)	< 0.001
In-hospital mortality	87/226 (38.5)	138/618 (22.3)	OR 1.94 (1.37–2.75)	< 0.001
ICU mortality	69/226 (30.5)	85/618 (13.8)	OR 2.37 (1.61–3.49)	< 0.001

All effect estimates compare deep sedation with light sedation (reference). For binary outcomes the effect is an odds ratio (OR) from a weighted or matched logistic model; for 28-day mortality the hazard ratio (HR) from a weighted Cox model and the absolute risk difference (RD) from a weighted linear model are additionally shown. n/N (%) denotes the number of events, the number at risk, and the crude percentage; a dash indicates that an event count is not applicable to that row. Confidence intervals (CI) for weighted estimates use a robust variance estimator. The E-value quantifies the minimum association an unmeasured confounder would need with both exposure and outcome to explain away the result. The proportional hazards assumption underlying the Cox model was not satisfied (Supplementary Methods); the hazard ratio is therefore an average effect over the 28-day period, and the odds ratio and risk difference are the principal measures for the primary outcome. IPTW, inverse probability of treatment weighting; ICU, intensive care unit; eICU-CRD, eICU Collaborative Research Database.

Sensitivity analyses

The association between deep sedation and 28-day mortality was robust across all three sensitivity analyses (Table 2). Propensity score matching yielded 2,186 matched pairs, with 642 deep sedation patients left unmatched because no light sedation patient lay within the caliper; after matching, all 30 covariates were balanced, every standardized difference falling below 0.1 (Supplementary Table S5). The odds ratio for 28-day mortality in the matched sample was 2.31 (95% CI 2.00 to 2.65), consistent with the primary IPTW estimate. Repeating the full IPTW analysis across five multiply imputed datasets and pooling the results by Rubin's rules yielded an essentially identical estimate for 28-day mortality (pooled OR 2.37, 95% CI 2.09 to 2.70). The E-value for the point estimate was 2.45, and the E-value for the lower confidence limit was 2.25; an unmeasured confounder would therefore need to be associated with both deep sedation and 28-day mortality by a risk ratio of at least 2.45, beyond the 30 measured covariates, to fully account for the observed association.

Subgroup analyses

The association between deep sedation and 28-day mortality was present in all prespecified subgroups but differed significantly in magnitude according to the severity of illness (Table 3, Fig. 3). The strongest association was observed among patients with a low SOFA score: in patients with a SOFA score of 6 or below, the odds ratio was 4.39 (95% CI 3.49 to 5.54), compared with 1.94 (95% CI 1.59 to 2.37) in patients with a SOFA score of 7 to 10 and 1.61 (95% CI 1.30 to 2.01) in patients with a SOFA score above 10 (P for interaction < 0.001). A parallel gradient was seen for neuromuscular blockade: the association was stronger in patients who did not receive a neuromuscular blocking agent (OR 2.69, 95% CI 2.33 to 3.11) than in those who did, for whom deep sedation is in large part obligatory (OR 1.68, 95% CI 1.30 to

2.17; P for interaction < 0.001). The association did not differ significantly by the presence of septic shock (no shock OR 2.83, 95% CI 2.26 to 3.56; septic shock OR 2.22, 95% CI 1.91 to 2.59; P for interaction = 0.08) or by age (below 65 years OR 2.44, 95% CI 2.00 to 2.97; 65 years or older OR 2.37, 95% CI 2.00 to 2.81; P for interaction = 0.83).

Table 3
Subgroup analysis of the association between deep sedation and 28-day mortality (MIMIC-IV).

Subgroup	Deep, n	Light, n	OR (95% CI)	P for interaction
Overall	2,828	5,649	2.37 (2.09–2.69)	–
Age				
< 65 years	1,568	2,608	2.44 (2.00–2.97)	0.83
≥ 65 years	1,260	3,041	2.37 (2.00–2.81)	
SOFA score				
≤ 6	693	2,633	4.39 (3.49–5.54)	< 0.001
7–10	962	2,197	1.94 (1.59–2.37)	
> 10	1,173	819	1.61 (1.30–2.01)	
Septic shock				
Absent	726	1,880	2.83 (2.26–3.56)	0.08
Present	2,102	3,769	2.22 (1.91–2.59)	
Neuromuscular blockade				
Not used	1,637	5,027	2.69 (2.33–3.11)	< 0.001
Used	1,191	622	1.68 (1.30–2.17)	

Each odds ratio (OR) was estimated by inverse probability of treatment weighting applied within the subgroup, comparing deep sedation with light sedation (reference). The P value for interaction was obtained from a multiplicative exposure-by-subgroup term in the weighted outcome model and is shown once per subgroup. Deep, n and Light, n denote the number of patients in each sedation group within the subgroup. CI, confidence interval; SOFA, Sequential Organ Failure Assessment.

External validation

Baseline characteristics of the eICU-CRD validation cohort are presented in Supplementary Table S2. As in the primary cohort, patients who received deep sedation were more severely ill, with a higher median APACHE IVa score (101 versus 90) and a higher median serum lactate (3.4 versus 2.2 mmol/L). Other indices, however, did not follow this pattern: vasopressor use and renal replacement therapy were somewhat less frequent in the deep sedation group of the validation cohort (Supplementary Table S2), in contrast to the primary cohort. This reflects differences in case mix and in documentation practice

across the contributing hospitals, and is consistent with the incomplete capture of continuously infused treatments in eICU-CRD discussed below. Crude 28-day mortality was 37.6% (85 of 226) with deep sedation and 21.0% (130 of 618) with light sedation. The parsimonious propensity score model achieved good balance on all ten covariates after weighting, with all standardized mean differences below 0.1 (Supplementary Table S3, Supplementary Figure S3). In the weighted analysis, deep sedation was associated with increased 28-day mortality, with an odds ratio of 2.01 (95% CI 1.42 to 2.87); the multiply imputed estimate was 1.96 (95% CI 1.37 to 2.79). The associations for ICU mortality (OR 2.37, 95% CI 1.61 to 3.49) and in-hospital mortality (OR 1.94, 95% CI 1.37 to 2.75) were concordant. The direction and approximate magnitude of the association in the independent multicentre cohort therefore reproduced the findings of the primary analysis (Table 2).

Discussion

In two large, independent cohorts of mechanically ventilated adults with sepsis, deep sedation during the first 48 hours of invasive ventilation was associated with a substantially higher risk of death. In the MIMIC-IV primary cohort, after inverse probability of treatment weighting had balanced 30 baseline covariates, the odds of 28-day mortality were more than twice as high among patients who received deep sedation as among those who received light sedation (odds ratio 2.37), and the absolute risk of death at 28 days was 15.4 percentage points higher. Deep sedation was also associated with a higher incidence of delirium and with higher ICU and in-hospital mortality. The association with 28-day mortality persisted across propensity score matching and multiple imputation, was not accounted for by the measured covariates, and would require an unmeasured confounder associated with both deep sedation and death by a risk ratio of at least 2.45 to be nullified (E-value 2.45); it was reproduced in direction and approximate magnitude in an independent cohort drawn from 35 hospitals (odds ratio 2.01). This analysis provides sepsis-specific evidence that the depth of early sedation is associated with survival.

The direction of this association is concordant with a substantial body of work conducted predominantly in general or mixed ICU populations, in which early deep sedation and greater sedation intensity during the first hours and days of ventilation have been shown to predict delayed extubation and increased mortality [6–8], a relationship confirmed in a systematic review and meta-analysis [9]. randomised trials of dexmedetomidine-based early sedation and of nonsedation as compared with light sedation have not demonstrated a mortality benefit from any single agent or from the avoidance of sedation altogether [28, 29]; they have, however, consistently shown that light sedation is achievable in most ventilated patients without harm, and no randomised trial has tested the contrast between sustained deep and sustained light sedation against a mortality endpoint. Our estimate is nevertheless larger than the mortality associations reported in most previous cohorts, and several features of the present study are likely to contribute. The exposure was defined as sustained deep sedation—a median RASS of − 3 or lower maintained across a full 48-hour window—rather than any single deep measurement or a graded measure of sedation intensity, and therefore identifies a more extreme and more durable phenotype. The comparison was deliberately restricted to clearly deep versus clearly light sedation, with patients of intermediate depth excluded, which sharpens the contrast. And the population was confined to patients

with sepsis, in whom the consequences of prolonged ventilation, immobility, and delirium may be amplified. We nonetheless interpret the absolute magnitude with caution, because residual confounding, discussed below, remains a plausible partial explanation; the direction of the association is more secure than its precise size.

A notable and internally consistent pattern emerged from the subgroup analyses. The association between deep sedation and mortality was strongest among the least severely ill patients—the odds ratio was 4.39 in those with a SOFA score of 6 or below, falling to 1.94 at a score of 7 to 10 and to 1.61 above 10 (P for interaction < 0.001)—and a parallel gradient was seen for neuromuscular blockade, the association being considerably stronger in patients who did not receive a neuromuscular blocking agent (odds ratio 2.69) than in those who did (1.68; P for interaction < 0.001). By contrast, the association did not differ significantly with age or with the presence of septic shock. The two significant gradients are clinically coherent and point in the same direction. In a patient who is not profoundly unstable and who is not receiving a neuromuscular blocker, deep sedation is largely discretionary: it reflects a choice rather than an unavoidable requirement, and its attendant harms—prolonged ventilation, immobility, delirium, and haemodynamic depression—are incurred without a competing clinical necessity. In the sickest patients, deep sedation is frequently obligatory, demanded by severe respiratory failure, refractory shock, or the use of neuromuscular blockade [30]; in that setting mortality is driven substantially by the underlying illness, the scope for sedation depth to be modified is narrower, and its independent contribution is correspondingly harder to discern. The observation that the association is most pronounced precisely where deep sedation is most avoidable is, moreover, difficult to reconcile with the association being driven principally by unmeasured illness severity, which would be expected to exert its strongest confounding effect among the most severely ill.

The reproduction of the association in the eICU-CRD cohort, assembled from 35 hospitals, addresses the principal weakness of the single-centre primary analysis and supports the generalizability of the finding beyond the practice of one institution. The estimate in the validation cohort was somewhat smaller than in the primary cohort (odds ratio 2.01 versus 2.37). This attenuation is unsurprising and has several plausible explanations: the propensity score model in eICU-CRD was necessarily restricted to ten covariates, because the records of continuously infused sedatives and neuromuscular blocking agents were too incomplete to be used, so that confounding by co-treatment could not be addressed; sepsis was operationalized differently in the two cohorts—the clinical Sepsis-3 definition in MIMIC-IV and administrative codes supplemented by diagnosis strings in eICU-CRD—which identify overlapping but not identical populations; case mix and sedation practice differed across the contributing hospitals; and vital status in eICU-CRD is ascertained only to hospital discharge, so that the 28-day mortality outcome in the validation cohort effectively reflects death before discharge rather than death from any cause at 28 days. A related point concerns the primary cohort: 28-day mortality slightly exceeded in-hospital mortality in the light sedation group (820 versus 690 deaths) but not in the deep sedation group (1,024 versus 1,023), because patients who received light sedation were more likely to survive to discharge, and a small number then died out of hospital within the 28-day window. Such post-discharge deaths are captured in MIMIC-IV through its linkage to state vital-records data, so that the 28-day endpoint remains

complete in the primary cohort. Despite these differences, the direction and approximate magnitude of the association were preserved.

Several mechanisms could link sustained deep sedation to mortality. Deep sedation delays spontaneous awakening and breathing trials and prolongs the duration of mechanical ventilation [31], extending exposure to ventilator-associated complications. It enforces immobility and so contributes to ICU-acquired weakness, which early mobilization—difficult to deliver in a deeply sedated patient—can mitigate [32]. Deeply sedated patients are at higher risk of delirium, which was more frequent with deep sedation in this study and which is itself associated with increased mortality and long-term cognitive impairment [12, 33]. The sedative and analgesic agents required to achieve deep sedation have dose-dependent cardiovascular depressant effects that may increase vasopressor requirements in a circulation already compromised by sepsis. Deep sedation may also obscure the clinical examination and delay recognition of evolving organ dysfunction. These pathways are biologically plausible, but they were not formally tested as mediators here, and the contribution of each remains to be established.

The study has several strengths. It is large, comprises more than nine thousand patients across two databases, and provides external validation in an independent multicentre population. The analysis was structured around pre-specified design elements—eligibility, exposure definition, start of follow-up, and outcomes—and time zero was placed at the end of the exposure window, so that follow-up began only after the exposure had been ascertained, avoiding immortal time bias. Confounding was addressed with a causal-inference approach and, in the primary cohort, a covariate set that included not only demographic factors and validated measures of illness severity but also the concomitant use of individual sedative agents, neuromuscular blockade, and corticosteroids. The robustness of the primary estimate was examined with three pre-specified sensitivity analyses, and effect modification was explored in four pre-specified subgroups.

Several limitations must be weighed against these strengths. First, the study is observational, and causal interpretation is constrained by the possibility of residual confounding; this consideration also bears on the magnitude of the estimated effect. Although thirty covariates—including validated severity scores and the concomitant use of individual sedative agents, neuromuscular blockade, and corticosteroids—were balanced by weighting, unmeasured factors such as the severity of hypoxaemia, ventilator settings, the trajectory of illness across the exposure window, premorbid frailty, and decisions about the goals of care could have influenced both the depth of sedation and the risk of death. The effect estimates are larger than those reported in most previous cohorts, and we interpret their absolute magnitude conservatively. Part of this difference is attributable to design: the exposure was defined as sedation sustained across a full 48-hour window rather than a single measurement or a graded intensity score, and the comparison was restricted to clearly deep versus clearly light sedation, so that a more extreme and more durable phenotype was contrasted. Residual confounding is nonetheless likely to contribute, and the E-value indicates that an unmeasured confounder would need to be associated with both deep sedation and mortality by a risk ratio of at least 2.45, beyond the measured covariates, to explain the association away. We therefore regard the direction and the qualitative consistency of the association

across cohorts, methods, and subgroups as more secure than its precise magnitude, and the subgroup estimates—in particular the large odds ratio in the least severely ill patients—as hypothesis-generating rather than as precise measures of effect.

Second, the exposure was operationalized through the median Richmond Agitation-Sedation Scale score, which reflects the observed depth of sedation rather than the intent or intensity of sedative treatment. In most patients a low score is the consequence of sedative administration, but a depressed level of consciousness may also arise from sepsis-associated encephalopathy or other neurological causes independent of sedation. Patients admitted to dedicated neurological or neurosurgical units were excluded to mitigate this, and the propensity score model included the use of each major sedative agent, so that the contrast is adjusted for documented pharmacological exposure; residual misclassification of the exposure cannot, however, be excluded, and would be expected to attenuate rather than to exaggerate the association. The exposure was, moreover, summarized as a single fixed value and was not modelled as a time-varying process; sedation depth in fact fluctuates and is continually adjusted in response to the patient's condition, and a time-varying exposure with time-varying confounding would require g-methods beyond the scope of this analysis [34].

Third, by design the cohort was restricted to patients who survived and remained in the ICU for at least 48 hours, since the exposure window and time zero required this; the findings therefore apply to patients who survive the first 48 hours of ventilation and do not address sedation decisions made earlier. Fourth, the proportional hazards assumption was not satisfied; the hazard ratio is reported only as a time-averaged summary, and inference for the primary outcome rests on the odds ratio and the risk difference, neither of which depends on this assumption. Fifth, delirium was ascertained from documented CAM-ICU assessments, and deeply sedated patients are frequently unarousable and therefore cannot be assessed; the delirium comparison is subject to ascertainment bias and is reported as exploratory only. Sixth, the validation cohort was smaller, was adjusted with a reduced covariate set because continuously infused medications were incompletely recorded, and captured mortality only to hospital discharge; the validation analysis should therefore be read as a test of the consistency of the direction and approximate magnitude of the association, not as an independent replication of equivalent rigour. Finally, outcomes were limited to mortality within 28 days and to the index hospitalization; functional, cognitive, and longer-term outcomes were not available.

Within these limitations, the findings have clinical relevance. They extend to patients with sepsis a principle already embodied in current guidelines and in structured ICU care bundles—that sedation should be no deeper than the patient's condition requires [10, 35]—and they are consistent with the broader practice of analgesia-first sedation, daily assessment of sedation depth, paired awakening and breathing trials, and early mobilization [36]. The subgroup findings indicate where this message is most actionable: the greatest apparent harm, and therefore the greatest opportunity, lies with patients who are not profoundly unstable and who are not receiving neuromuscular blockade, in whom deep sedation is most often a discretionary choice rather than a clinical necessity; in patients for whom deep sedation is genuinely obligatory the clinical priorities differ. These observational results cannot establish causation

and should not be read as a mandate. Rather, they strengthen the rationale for a randomised trial of sedation-depth targets conducted specifically in mechanically ventilated patients with sepsis, which would provide the definitive test.

Conclusions

In two large, independent cohorts of mechanically ventilated adults with sepsis, deep sedation during the first 48 hours of invasive ventilation was associated with substantially higher 28-day mortality, higher ICU and in-hospital mortality, and a higher incidence of delirium (assessed as an exploratory outcome). The association was consistent across propensity score matching and multiple imputation, was reproduced in an independent multicentre cohort, and was most pronounced in patients in whom deep sedation was least likely to be clinically obligatory. Although the observational design precludes a definitive causal conclusion and residual confounding cannot be excluded, these findings support the avoidance of unnecessary deep sedation in mechanically ventilated patients with sepsis and provide a clear rationale for a randomised trial of sedation-depth targets in this population.

Declarations

Ethics approval and consent to participate

The MIMIC-IV and eICU Collaborative Research Database projects were approved by the institutional review boards of the Massachusetts Institute of Technology and, for MIMIC-IV, the Beth Israel Deaconess Medical Center, with a waiver of the requirement for individual informed consent because the data are de-identified. The present secondary analysis used only publicly available, fully de-identified data and therefore required no additional ethical approval. The investigators completed the mandated human-subjects research training and signed the data use agreements before obtaining access.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Funding

This study received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Author Contribution

L.S. and Y.L. conceived and designed the study. L.S. performed the data extraction, statistical analysis, and interpretation, and drafted the manuscript. G.L. and Y.Y. contributed to data extraction and to the analysis and interpretation of the data. C.X. and H.T. contributed to the interpretation of the results and to the critical revision of the manuscript for important intellectual content. Y.L. supervised the study and critically revised the manuscript. All authors read and approved the final manuscript and agree to be accountable for all aspects of the work.

Acknowledgement

The authors thank the investigators and institutions responsible for developing and maintaining the MIMIC-IV and eICU-CRD databases and for making them available to the research community.

Data Availability

The MIMIC-IV and eICU-CRD datasets analysed in the present study are not publicly redistributable but are available to credentialed researchers on the PhysioNet platform (https://physionet.org) following completion of the required training and execution of the data use agreement. The analysis code (SQL and R) needed to reproduce the study from these databases has been deposited in the Zenodo repository and is openly available at https://doi.org/10.5281/zenodo.20252219

References

1. Singer, M. et al. The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). *JAMA* **315**, 801–810 (2016).
2. Rudd, K. E. et al. Global, regional, and national sepsis incidence and mortality, 1990–2017: analysis for the Global Burden of Disease Study. *Lancet* **395**, 200–211 (2020).
3. Kress, J. P., Pohlman, A. S., O'Connor, M. F. & Hall, J. B. Daily interruption of sedative infusions in critically ill patients undergoing mechanical ventilation. *N Engl. J. Med.* **342**, 1471–1477 (2000).
4. Strøm, T., Martinussen, T. & Toft, P. A protocol of no sedation for critically ill patients receiving mechanical ventilation: a randomised trial. *Lancet* **375**, 475–480 (2010).
5. Treggiari, M. M. et al. randomised trial of light versus deep sedation on mental health after critical illness. *Crit. Care Med.* **37**, 2527–2534 (2009).
6. Shehabi, Y. et al. Early intensive care sedation predicts long-term mortality in ventilated critically ill patients. *Am. J. Respir Crit. Care Med.* **186**, 724–731 (2012).
7. Balzer, F. et al. Early deep sedation is associated with decreased in-hospital and two-year follow-up survival. *Crit. Care*. **19**, 197 (2015).

8. Shehabi, Y. et al. Sedation intensity in the first 48 hours of mechanical ventilation and 180-day mortality: a multinational prospective longitudinal cohort study. *Crit. Care Med.* **46**, 850–859 (2018).
9. Stephens, R. J. et al. Practice patterns and outcomes associated with early sedation depth in mechanically ventilated patients: a systematic review and meta-analysis. *Crit. Care Med.* **46**, 471–479 (2018).
10. Devlin, J. W. 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.* **46**, e825–e873 (2018).
11. Sessler, C. N. et al. The Richmond Agitation-Sedation Scale: validity and reliability in adult intensive care unit patients. *Am. J. Respir Crit. Care Med.* **166**, 1338–1344 (2002).
12. Pandharipande, P. P. et al. Long-term cognitive impairment after critical illness. *N Engl. J. Med.* **369**, 1306–1316 (2013).
13. Sonneville, R. et al. Potentially modifiable factors contributing to sepsis-associated encephalopathy. *Intensive Care Med.* **43**, 1075–1084 (2017).
14. Hernán, M. A. & Robins, J. M. Using big data to emulate a target trial when a randomised trial is not available. *Am. J. Epidemiol.* **183**, 758–764 (2016).
15. Johnson, A. E. W. et al. MIMIC-IV, a freely accessible electronic health record dataset. *Sci. Data*. **10**, 1 (2023).
16. Pollard, T. J. et al. The eICU Collaborative Research Database, a freely available multi-center database for critical care research. *Sci. Data*. **5**, 180178 (2018).
17. von Elm, E. et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *Lancet* **370**, 1453–1457 (2007).
18. Vincent, J. L. et al. The SOFA (Sepsis-related Organ Failure Assessment) score to describe organ dysfunction/failure. *Intensive Care Med.* **22**, 707–710 (1996).
19. Ely, E. W. et al. Delirium in mechanically ventilated patients: validity and reliability of the Confusion Assessment Method for the Intensive Care Unit (CAM-ICU). *JAMA* **286**, 2703–2710 (2001).
20. Le Gall, J. R., Lemeshow, S. & Saulnier, F. A new Simplified Acute Physiology Score (SAPS II) based on a European/North American multicenter study. *JAMA* **270**, 2957–2963 (1993).
21. Charlson, M. E., Pompei, P., Ales, K. L. & MacKenzie, C. R. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *J. Chronic Dis.* **40**, 373–383 (1987).
22. Goradia, S. et al. Vasopressor dose equivalence: a scoping review and suggested formula. *J. Crit. Care*. **61**, 233–240 (2021).
23. Zimmerman, J. E., Kramer, A. A., McNair, D. S. & Malila, F. M. Acute Physiology and Chronic Health Evaluation (APACHE) IV: hospital mortality assessment for today's critically ill patients. *Crit. Care Med.* **34**, 1297–1310 (2006).
24. Austin, P. C. & Stuart, E. A. Moving towards best practice when using inverse probability of treatment weighting (IPTW) using the propensity score to estimate causal treatment effects in observational

studies. *Stat. Med.* **34**, 3661–3679 (2015).

25. van Buuren, S. & Groothuis-Oudshoorn, K. mice: multivariate imputation by chained equations in R. *J. Stat. Softw.* **45**, 1–67 (2011).
26. VanderWeele, T. J. & Ding, P. Sensitivity analysis in observational research: introducing the E-value. *Ann. Intern. Med.* **167**, 268–274 (2017).
27. Ho, D. E., Imai, K., King, G. & Stuart, E. A. MatchIt: nonparametric preprocessing for parametric causal inference. *J. Stat. Softw.* **42**, 1–28 (2011).
28. Shehabi, Y. et al. Early sedation with dexmedetomidine in critically ill patients. *N Engl. J. Med.* **380**, 2506–2517 (2019).
29. Olsen, H. T. et al. Nonsedation or light sedation in critically ill, mechanically ventilated patients. *N Engl. J. Med.* **382**, 1103–1111 (2020).
30. Papazian, L. et al. Neuromuscular blockers in early acute respiratory distress syndrome. *N Engl. J. Med.* **363**, 1107–1116 (2010).
31. Girard, T. D. et al. Efficacy and safety of a paired sedation and ventilator weaning protocol for mechanically ventilated patients in intensive care (Awakening and Breathing Controlled trial): a randomised controlled trial. *Lancet* **371**, 126–134 (2008).
32. Schweickert, W. D. et al. Early physical and occupational therapy in mechanically ventilated, critically ill patients: a randomised controlled trial. *Lancet* **373**, 1874–1882 (2009).
33. Salluh, J. I. F. et al. Outcome of delirium in critically ill patients: systematic review and meta-analysis. *BMJ* **350**, h2538 (2015).
34. Robins, J. M., Hernán, M. A. & Brumback, B. Marginal structural models and causal inference in epidemiology. *Epidemiology* **11**, 550–560 (2000).
35. Pun, B. T. et al. Caring for critically ill patients with the ABCDEF bundle: results of the ICU Liberation Collaborative in over 15,000 adults. *Crit. Care Med.* **47**, 3–14 (2019).
36. Reade, M. C. & Finfer, S. Sedation and delirium in the intensive care unit. *N Engl. J. Med.* **370**, 444–454 (2014).

Figures

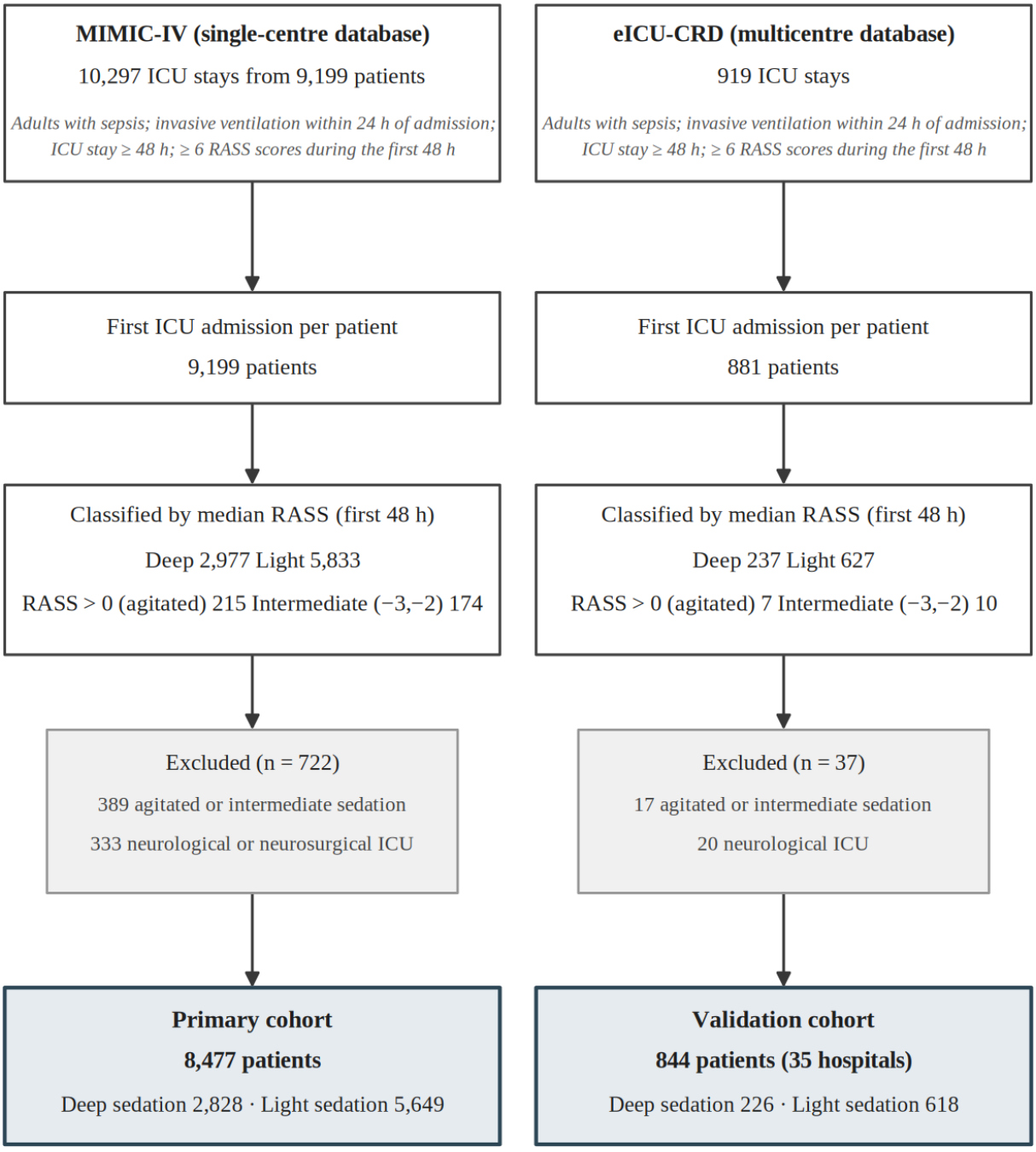


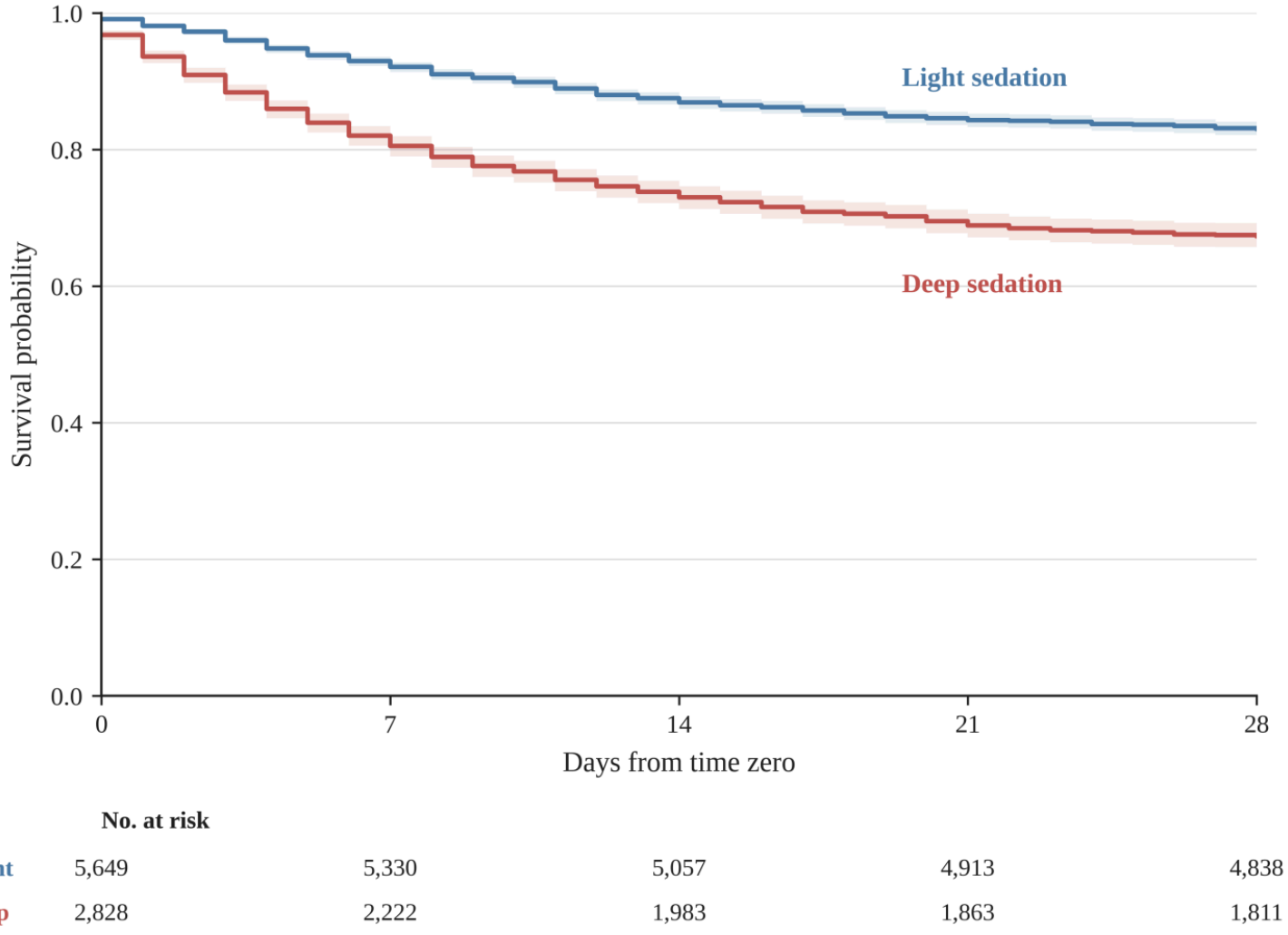
Figure 1

Derivation of the study cohorts.

Flow diagram showing the derivation of the MIMIC-IV primary cohort and the eICU-CRD external validation cohort. In MIMIC-IV, 10,297 qualifying intensive care unit (ICU) stays from 9,199 unique patients were reduced to one admission per patient; classification by the median Richmond Agitation-

Sedation Scale (RASS) score over the first 48 hours of invasive mechanical ventilation identified deep, light, agitated, and intermediate-depth groups; after exclusion of agitated and intermediate-depth patients (n = 389) and of patients admitted to a neurological or neurosurgical ICU (n = 333), the primary cohort comprised 8,477 patients. The eICU-CRD cohort was derived in parallel, with the same exclusion of agitated, intermediate-depth, and neurological-ICU patients, and comprised 844 patients from 35 hospitals.

RASS, Richmond Agitation-Sedation Scale; ICU, intensive care unit; MIMIC-IV, Medical Information Mart for Intensive Care IV; eICU-CRD, eICU Collaborative Research Database.



No. at risk					
Light	5,649	5,330	5,057	4,913	4,838
Deep	2,828	2,222	1,983	1,863	1,811

Figure 2

Inverse probability of treatment weighting-adjusted Kaplan-Meier curves for 28-day survival, MIMIC-IV cohort.

Survival curves for the deep and light sedation groups in the primary cohort, estimated after inverse probability of treatment weighting, with follow-up beginning at time zero (48 hours after the initiation of invasive mechanical ventilation). The survival curves separate within the first days after time zero and remain separated throughout the 28-day follow-up. A weighted Cox model gave an average hazard ratio

for death of 2.22 (95% confidence interval 2.00 to 2.47) over this period; the relative hazard was not constant over time (proportional hazards assumption not met; see Supplementary Methods).

Shaded bands represent 95% confidence intervals. The numbers of patients at risk, shown below the time axis, are unweighted counts. Survival was defined by the same 28-day all-cause mortality endpoint used throughout; deaths recorded within the 28-day window but at or before time zero were assigned a survival time of zero.

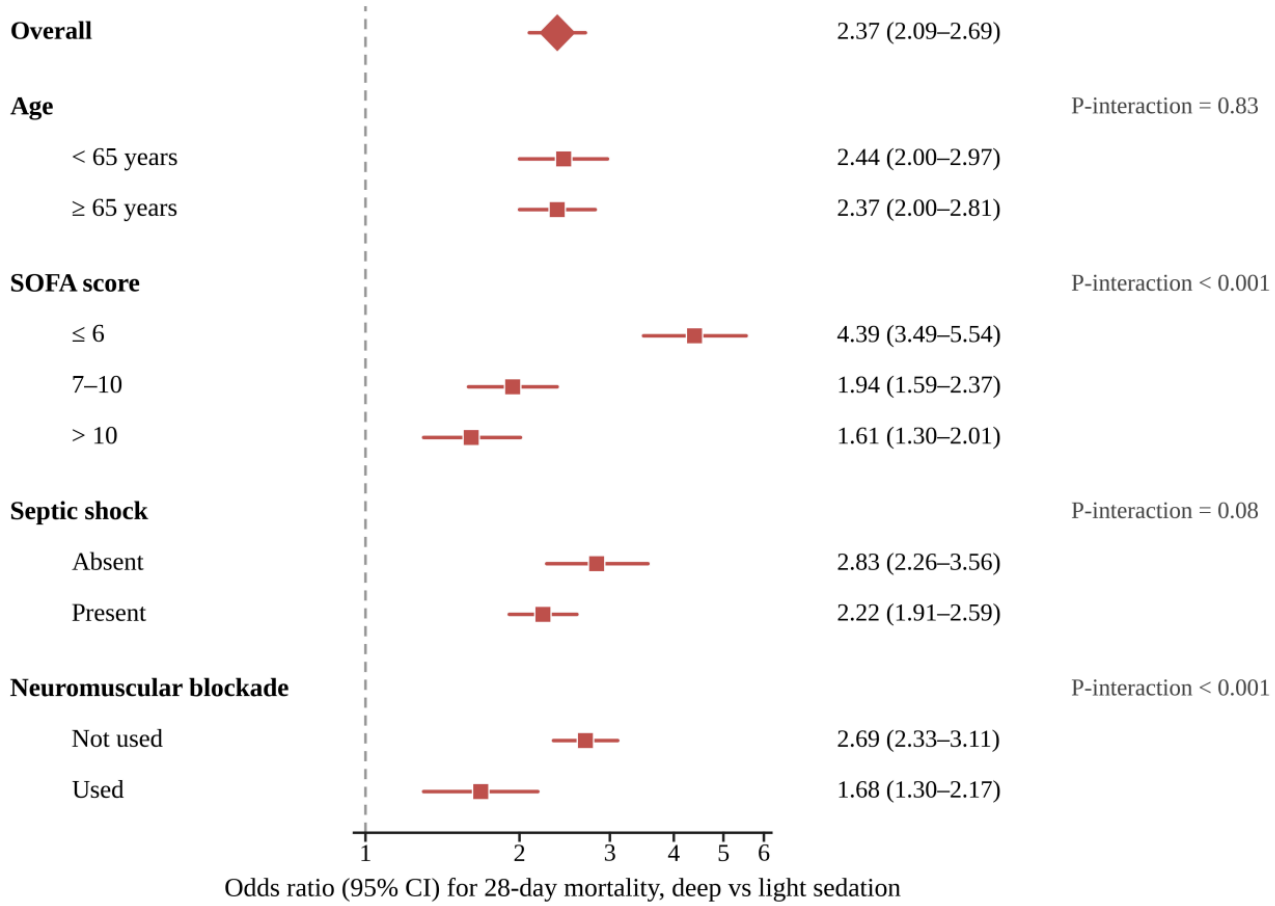


Figure 3

Association between deep sedation and 28-day mortality across prespecified subgroups, MIMIC-IV cohort.

Forest plot of the inverse probability of treatment weighting-adjusted odds ratio for 28-day mortality with deep sedation, relative to light sedation, within prespecified subgroups defined by age, Sequential Organ Failure Assessment (SOFA) score category, the presence of septic shock, and the use of neuromuscular blocking agents. The association is strongest among patients with a lower SOFA score and among patients not receiving neuromuscular blockade (both P for interaction < 0.001).

The diamond represents the overall estimate and squares the subgroup point estimates; horizontal lines represent 95% confidence intervals. The vertical reference line is drawn at an odds ratio of 1.0; estimates to

its right indicate higher mortality with deep sedation. OR, odds ratio; SOFA, Sequential Organ Failure Assessment.

Supplementary Files

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

- SupplementarymaterialsSR.docx
- STROBEchecklistSR.docx