

Multicenter target trial emulation to evaluate corticosteroids for sepsis stratified by predicted organ dysfunction trajectory

Received: 31 January 2024

Accepted: 30 April 2025

Published online: 13 May 2025



Suraj Rajendran¹, Zhenxing Xu², Weishen Pan², Chengxi Zang²,
Ilias Siempos^{3,4}, Lisa Torres³, Jie Xu⁵, Jiang Bian^{6,7},
Edward J. Schenck³✉ & Fei Wang²✉

Corticosteroids decrease the duration of organ dysfunction in sepsis and a range of overlapping and complementary infectious critical illnesses, including septic shock, pneumonia and the acute respiratory distress syndrome (ARDS). The risk and benefit of corticosteroids are not fully defined using the construct of organ dysfunction duration. This retrospective multicenter, proof-of-concept study aimed to evaluate the association between usage of corticosteroids and mortality of patients with sepsis, pneumonia and ARDS by emulating a target trial framework stratified by predicted organ dysfunction trajectory. The study employed a two staged machine learning (ML) methodology to first subphenotype based on organ dysfunction trajectory then predict this defined trajectory. Once patients were classified by predicted trajectory we conducted a target trial emulation. Our analysis revealed that the association between corticosteroid use and 28-day mortality varied by predicted trajectory and between cohorts. Our findings suggest that matching treatment strategies to empirically observed pathobiology may offer a more nuanced understanding of corticosteroid utility.

Infection is the most common cause of critical illness across the globe. Critically ill patients with infection, labeled as sepsis, present with organ dysfunction that manifest in overlapping and summative phenomena including circulatory, kidney and lung failure¹. These organ specific failures are subject to their own syndromic definitions, such as septic shock in the case of circulatory failure and the acute respiratory distress syndrome (ARDS) in the case of hypoxemic respiratory failure. However, patients with septic shock often meet the syndromic definition of ARDS and vice versa. Both septic shock and ARDS have been evaluated in numerous epidemiologic and clinical studies with a marked lack of breakthrough therapies^{2–4}. Moreover, the organ specific source of infection, e.g., pneumonia, and intra-abdominal infections have their own distinct

evidentiary base⁵. To reconcile these conflicting conceptual models of the causes of and consequences of sepsis, there has been a call to redefine critical illnesses as phenotypic representations of treatable traits¹.

Corticosteroids are an adjunctive therapy for selected patients with sepsis due to their vasoconstrictive, anti-inflammatory, and immunomodulatory effects^{6,7}. These effects lead to an empirical reduction in the duration of organ dysfunction in septic shock, and ARDS⁸. However, the benefits of corticosteroid treatment must be balanced against the risks of immunosuppression, hyperglycemia, and other adverse events^{9–11}. There remains an unmet need to establish a more precise target population within sepsis to improve the risk benefit profile of corticosteroid therapy^{8,12–14}.

¹Tri-Institutional Computational Biology & Medicine Program, Cornell University, Cornell, NY, USA. ²Division of Health Informatics, Department of Population Health Sciences, Weill Cornell Medicine, New York, NY, USA. ³Division of Pulmonary and Critical Care Medicine, Department of Medicine, New York-Presbyterian Hospital-Weill Cornell Medical Center, Weill Cornell Medicine, New York, NY, USA. ⁴First Department of Critical Care Medicine and Pulmonary Services, Evangelismos Hospital, National and Kapodistrian University of Athens Medical School, Athens, Greece. ⁵Department of Health Outcomes and Biomedical Informatics, College of Medicine, University of Florida, Gainesville, FL, USA. ⁶Department of Biostatistics and Health Data Science, School of Medicine, Indiana University, Indianapolis, IN, USA. ⁷Regenstrief Institute, Indianapolis, IN, USA. ✉e-mail: ejs9005@med.cornell.edu; few2001@med.cornell.edu

Given the landscape of overlapping syndromes and causes of sepsis, the development of a precision medicine framework is needed¹. Recent attempts to evaluate heterogeneity of corticosteroid treatment effects have used specific syndromes, e.g., septic shock^{15–17}, or specific infectious causes of critical illness, e.g., COVID-19^{18,19}. These analyses were focusing on specific syndromes and most of them, aside from an exploration of how steroids interact with the duration of vasoactive medications in phenotypes of pediatric multiple organ dysfunction, did not explicitly explore the effect of corticosteroids on the expected duration of organ dysfunction¹⁷. In addition, those analyses were mostly conducted on a single cohort, which can not reflect the broad population of patients with sepsis treated with steroids in clinical practice²⁰. To bridge this gap, there is a need to evaluate the real world effect of corticosteroids as used in clinical practice, which can be done with target trial emulation^{21,22}. More specifically, since the observed beneficial effect of corticosteroids in sepsis is through the amelioration of organ dysfunction, exploring treatment heterogeneity between subphenotypes defined by organ dysfunction trajectory is valuable²³.

We conducted a retrospective multicenter proof-of-concept study to explore the effects of corticosteroids in patients with sepsis, through a target trial emulation framework stratified by subphenotypes of predicted organ dysfunction trajectory (Fig. 1). Specifically, the objectives of this study were to (1) determine whether the effectiveness of corticosteroid administration on 28-day mortality is modified by stratification based on trajectory-based subphenotypes; and (2) evaluate the effectiveness of corticosteroids on secondary endpoints including the duration of ICU stay and the duration of mechanical ventilation. This real-world data will serve as a model to demonstrate the feasibility of a subphenotype stratified clinical trial based on mechanistic rationales that take the expected response to standard of care into account.

Results

Data cohorts

Three distinct data sources including Medical Information Mart for Intensive Care (MIMIC-IV)²⁴, eICU²⁵, and Critical Care Database for Advanced Research (CEDAR)²⁶ were used in our analysis. We combined MIMIC-IV and eICU to form the development cohort (eICU-MIMIC) for primary analysis, and left the CEDAR cohort for validation. More details on these cohorts can be found in the Methods, Data Source and Cohort.

Target trial emulation design

In this study, we conducted an intention-to-treat (ITT) analysis to assess the impact of corticosteroid treatment on sepsis outcomes in critically ill patients. Sepsis was defined using a modified Sepsis-3 criteria²⁷, which included suspected infection and administration of antibiotics alongside the collection of a body fluid culture specimen and a Sequential Organ Failure Assessment (SOFA) score of at least 2. The enrollment window was defined as the 24-h period after ICU admission. Patients aged 18 or older who met sepsis criteria within the enrollment window were included. The treatment group is defined as patients with initiation of hydrocortisone at ≥ 160 mg/day within 10 h before to 24 h after ICU admission, and the control group is defined as patients with no corticosteroids within the same window. Details on corticosteroid dosage and the inclusion-exclusion cascade are shown in Supplementary Figs. 1–3 and Supplementary Tables 3–5. The primary outcome was 28-day mortality from ICU admission. Secondary outcomes included time to ICU discharge and time to cessation of mechanical ventilation. For the latter, only patients on mechanical ventilation at baseline were included. We did not analyze the cessation of ventilation outcome in the CEDAR cohort due to an insufficient number of patients meeting the inclusion criteria. Patients were followed until death, loss of follow-up, or discharge. Further details on

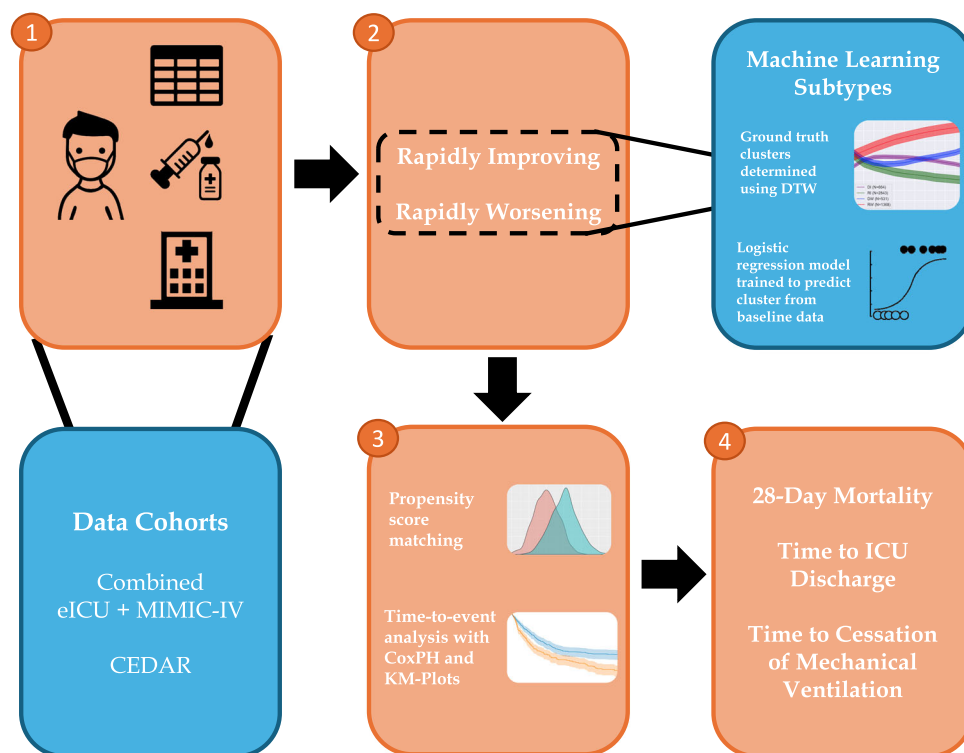


Fig. 1 | Overview of analyses conducted in study. Relevant covariates and patient information were extracted from multiple data cohorts. These patients were then stratified and processed as per trial emulation setup. Statistical methods were then used to determine relevant outcomes.

outcome definitions and follow-up are provided in Supplementary Fig. 4. More information about study design can be found in the Methods, Target Trial Emulation and Specification section.

To balance baseline covariates between treatment and control groups, we employed propensity score matching²⁸. Propensity scores were estimated using logistic regression (LR), with treatment assignment as the dependent variable and baseline covariates as independent variables. A Cox proportional hazards model was used to estimate the hazard ratio (HR) for different outcomes (28-day mortality, time to ICU discharge, and time to cessation of mechanical ventilation). Details of statistical analysis were provided in the Methods, Statistical Methods section, with various sensitivity analysis detailed in the Methods, Sensitivity Analyses section.

ML-based patient stratification

To address the heterogeneity of sepsis, we stratified patients into subgroups based on disease trajectories to investigate the differential effectiveness of corticosteroid treatment. We used the Rapidly Improving (RI) and Rapidly Worsening (RW) subphenotypes identified in the previous Xu et al. study²³, which were derived using Dynamic Time Warping (DTW) and hierarchical clustering. The RI patients were characterized by a decreasing SOFA score over time, while the RW patients displayed an increasing SOFA score over the same duration. The characteristics of the subphenotypes in eICU, MIMIC, and CEDAR are shown in Supplementary Fig. 5. For the eICU-MIMIC cohort, we identified 5038 RW patients and 11,239 RI patients. Consistent with prior studies, we found that the RI group was characterized by a lower in-hospital mortality rate despite a higher mean SOFA score at ICU admission (Supplementary Table 9 and Supplementary Fig. 5). Conversely, the RW group was associated with metabolic acidosis (high lactate), disseminated intravascular coagulation (a higher INR and low platelets) and worse outcomes despite a lower overall SOFA score at ICU admission (Supplementary Tables 9, 12 and 13). The derivation of these subphenotypes involves data in the potential exposure window, which cannot be used as a way for patient stratification to investigate heterogeneous treatment effects. To mitigate this issue, we stratified the cohorts by predictive models restricted to data available during the enrollment period. Specifically, we trained LR models to predict a patient's actual subphenotype assignments from enrollment period covariates such as lab tests, vital signs, and demographic information, with the complete list shown in Supplementary Text 1. These models were used to identify the RI and RW subphenotypes in target trial emulation. More details on the LR models can be found in the Methods, with their performance evaluations provided in Supplementary Text 1.

Characteristics of the cohorts

Demographic details of the eICU-MIMIC and CEDAR cohorts are provided in Supplementary Tables 6 and 7 respectively. The comorbidity burden of sepsis patients, as indicated by the Elixhauser index, in the eICU-MIMIC cohort who were treated with steroids differ from those who were untreated. The treated group showcased a higher median value of 6.0 compared to the untreated group's median of 5.0 ($p < 0.001$). This suggests that the patients who received steroids possibly had a more complex clinical profile. When examining the sources of infections, both groups predominantly suffered from septicemia bacteremia, with a higher proportion seen in the untreated group (45.8%), compared to the treated group (43.5%) ($p < 0.001$). Pneumonia was another significant source, affecting a larger fraction of the treated cohort (46.4%), compared to 28.8% in the untreated population ($p < 0.001$). Other sources of infection, including infections of the central nervous system, intra-abdominal region, skin soft tissue, and urinary tract, displayed varying proportions.

In the CEDAR cohort, the age of the treated cohort was notably younger, with a median age of 64.0 years, in contrast to the untreated

group's median age of 72.0 years ($p < 0.001$). In terms of sex distribution, 56.6% of the untreated patients were male, compared to 50.7% of the treated patients. The comorbidity burden was uniformly distributed with a median of 15.0 in both groups. In terms of clinical severity, the mechanical ventilation rates and SOFA scores at admission were similar across both groups. The distributions of the infection sources were consistent between the treated and untreated groups, with pneumonia and septicemia bacteremia as dominant sources.

Supplementary Table 8 summarizes the outcomes and baseline characteristics of the eICU-MIMIC and CEDAR cohorts. For the outcome of 28-day mortality, we analyzed the incidence and the time from ICU admission to death across different patient subgroups, including the overall cohort and the predicted RI and RW subphenotypes. In the eICU-MIMIC cohort, the mortality incidence was 17.7%, with an average time from ICU admission to death of 22.13 days. We also found that the RI subgroup displayed a lower mortality rate compared to the RW group in both cohorts (Supplementary Table 9). For the outcome of time to ICU discharge, the overall discharge incidence was 71.1% in the eICU-MIMIC cohort. Patients stayed in the ICU for an average duration of 11.83 days prior to discharge, and the RI group showed a higher discharge incidence compared to the RW subgroup. In the CEDAR cohort, the discharge incidence between the RI and RW subgroups are similar. For time to cessation of mechanical ventilation, the overall incidence of ventilation cessation was 79.1%, with an average duration of 8.70 days from ICU admission to the cessation of ventilation. In the eICU-MIMIC cohort, the incidence of ventilation cessation is comparable between RI and RW subgroups.

Survival analysis on outcomes

After obtaining balanced covariates with propensity score matching, we computed the hazard ratios (HRs) for the different outcomes across the different cohorts and stratifications. The standard mean difference (SMD) of covariates before and after balancing are depicted in Supplementary Figs. 6, 7, and 8. Supplementary Tables 10 and 11 show the median and interquartile range values of all covariates before and after balancing for treated and untreated groups. Kaplan Meier plots for all cohorts and outcomes with respect to RI and RW are shown in Fig. 2 and Supplementary Fig. 9.

28-day mortality. Figure 3 shows the HRs of the 28-day mortality for patients treated with steroids versus those who were not. An HR of less than 1 implies a decreased risk of 28-day mortality associated with steroid treatment. For the eICU-MIMIC cohort, the overall HR was 1.10, suggesting a negative impact of steroids on 28-day mortality. The RI subgroup had an increased risk of mortality with steroid treatment while the RW group showed a decreased risk. In the CEDAR cohort (Supplementary Fig. 10), the HRs were higher across both the entire cohort and the RI, RW subgroups compared to the eICU-MIMIC cohort, indicating a general potentially increased risk of 28-day mortality associated with steroid use.

Time to ICU discharge. To maintain consistency with the mortality outcome (where $HR < 1$ indicates a benefit from steroids), we reverse-coded the favorable endpoints (ICU discharge and ventilator cessation) by swapping the numerator and denominator of the HR, such that a resultant $HR < 1$ represents a shorter time to these events in the treated group (Fig. 4). In the eICU-MIMIC cohort, the overall HR suggested no effect to discharge associated with steroid use. The RI subgroup showed an association between steroid use and a longer time to ICU discharge, whereas the RW subgroup showed steroid use was associated with a shorter time to ICU discharge. In the CEDAR cohort the overall cohort had a slightly prolonged time to discharge with steroid use (Supplementary Fig. 11). The RI subgroup had an HR significantly greater than 1, suggesting steroid use is associated

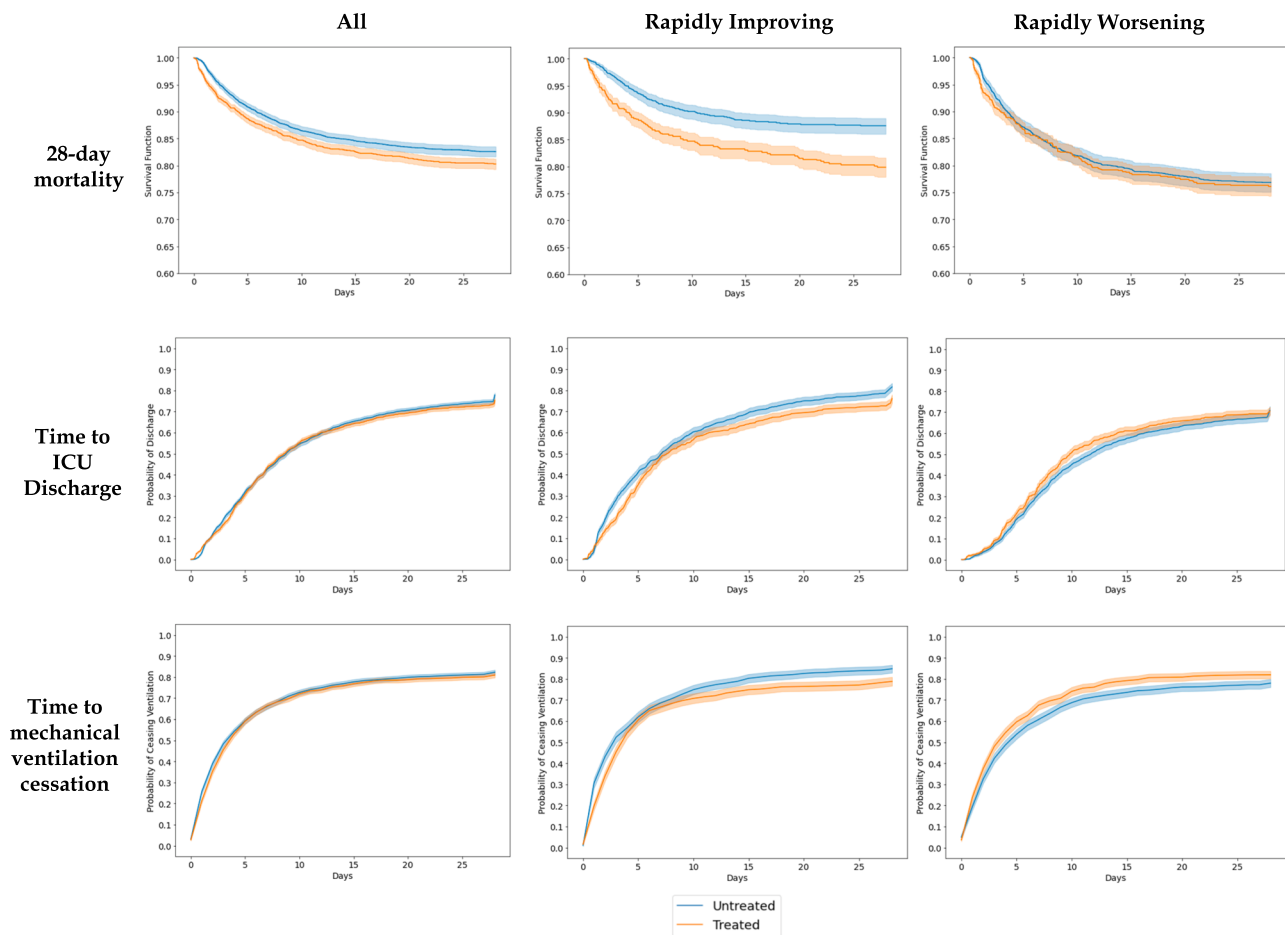


Fig. 2 | Kaplan–Meier plots for eICU-MIMIC. Different outcomes for both treated and untreated patients in the entire cohort and RI, RW subphenotypes are depicted. Kaplan Meier plots were based on propensity-score matched cohorts. Bands around the Kaplan–Meier curves indicate 95% confidence intervals.

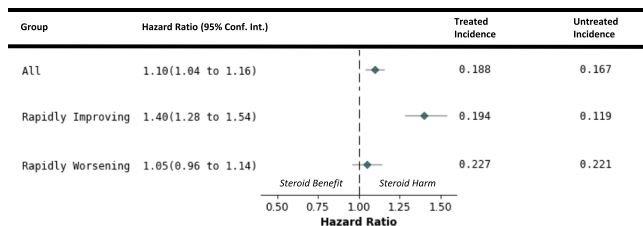


Fig. 3 | Hazard ratios and cumulative incidence for 28-day mortality in the eICU-MIMIC cohort. Forest plot shows the hazard ratios (center of the bar) with 95% confidence intervals for each patient stratification. Dashed vertical line represents a hazard ratio of 1. Cumulative incidences for treated and untreated patients are shown on the right. $n = 13704$ for All patients, $n = 4102$ for RI patients, $n = 4802$ for RW patients. Source data are provided as a Source Data file.

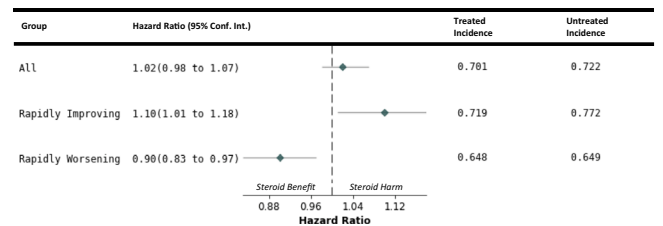


Fig. 4 | Hazard ratios and cumulative incidence for time to ICU discharge outcome in eICU-MIMIC cohort. Forest plot shows the hazard ratios (center of the bar) and 95% confidence intervals for each patient stratification. Dashed vertical line represents a hazard ratio of 1. Cumulative incidences for treated and untreated patients are shown on the right. $n = 13702$ for All patients, $n = 4112$ for RI patients, $n = 4818$ for RW patients. Source data are provided as a Source Data file.

with a longer time to discharge, while the RW subgroup showed no significant association.

Time to cessation of mechanical ventilation. Similar to time to ICU discharge, we reverse-coded the endpoints for cessation of mechanical ventilation. As such, for time to cessation of mechanical ventilation, an HR less than 1 indicates that steroid use is associated with a shorter duration of ventilation for treated patients compared to those untreated (Fig. 5). For the eICU-MIMIC cohort, the overall HR was 1.03 (95% CI: 0.98–1.08), suggesting no significant association between steroid use and ventilation cessation time. The RI subgroup had a longer duration of ventilation associated with steroid use, whereas the RW subgroup had a shorter duration of ventilation associated with steroid use.

Our sensitivity analyses (Supplementary Figs. 12–18) agree with the results of our primary analyses, where we also observed relatively increased negative associations for RI patients with steroid use compared to RW patients, and the incidence of 28-day mortality for RW patients was higher than that of RI patients. Supplementary Figs. 17 and 18 show the effects of including source of infection as a covariate to our balancing method. For MIMIC and CEDAR, we make the assumption that the source of infection was defined at baseline as diagnosis timepoints were not available in these databases. These results show that inclusion of source of infection does not impact the effects seen without source of infection. The directionality difference in terms of steroid benefit between the RI and RW groups are still retained. The sensitivity analysis using ICU-free days

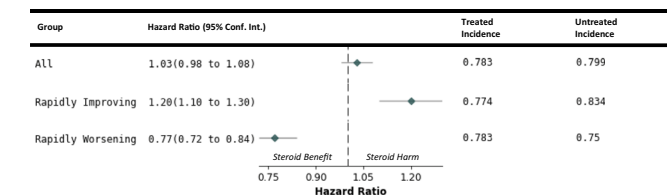


Fig. 5 | Hazard ratios and cumulative incidence for time to cessation of mechanical ventilation outcome in eICU-MIMIC cohort. Forest plot shows the hazard ratios (center of the bar) and 95% confidence intervals for each patient stratification. Dashed vertical line represents a hazard ratio of 1. Cumulative incidences for treated and untreated patients are shown on the right. $n = 9576$ for All patients, $n = 2986$ for RI patients, $n = 3744$ for RW patients. Source data are provided as a Source Data file.

(ICU-FDs) and ventilation-free days (VFDs) (Supplementary Text 2) shows that in the RW subphenotype, corticosteroid-treated patients had more ICU-FDs and VFDs compared to untreated patients, while such difference is not seen in the RI subphenotype²⁹. In addition, we estimated an Optimal Individual Treatment Regime with details described in Supplementary Text 3, which essentially builds a predictive model using baseline covariates along with the treatment indicator as predictors and 28-day mortality as the outcome³⁰. We found that it was challenging for the model to achieve satisfactory performance (AUROC of 0.5746 ± 0.0038), which highlights the inherent complexity of this problem. Despite its performance, we found that a greater proportion of patients classified as RW would be optimally treated with corticosteroids compared to RI patients according to the model, which is consistent with our main analysis. We also conducted a set of sensitivity analysis to investigate immortal time bias, where all patients were cloned to both treatment and control groups and censored when they deviated³¹. As shown in Supplementary Table 15, the results were consistent with our primary findings, displaying the same directionality of effects between the RI and RW subphenotypes. Finally, to assess the potential impact of unmeasured confounding, we calculated E -values using the formula developed by Tyler J. VanderWeele³², which ranged from 1.13 to 2.01 in our primary analysis. This suggests that unmeasured confounders of moderate strength could potentially explain away the observed associations (Supplementary Table 16 and Supplementary Text 4).

Discussion

We examined the effects of corticosteroids in patients with sepsis inclusive of ARDS, pneumonia, and septic shock, through a target trial emulation (TTE) framework with the EHR data from two cohorts (eICU-MIMIC and CEDAR). Across both cohorts our model found that corticosteroids treatment was associated with a small increase in 28-day mortality compared to untreated controls (eICU MIMIC, HR 1.10, CI: 1.04–1.16). There was no significant increase in duration of mechanical ventilation (eICU MIMIC, HR 1.03, CI: 0.98–1.08) or length of stay in the ICU (eICU MIMIC, HR 1.02, CI: 0.98–1.07). To explore the potential population heterogeneity and its impact on treatment effects, we leveraged the model developed in Xu et al. to identify organ dysfunction derived subphenotypes for sepsis patients and conduct TTE within each of them for treatment effect estimation²³.

Through the analysis, we found that steroid treatment was associated with higher mortality for patients within the RI subphenotype in both cohorts. This may be due to several mechanisms. For example, improving patients with sepsis may already have shifted to an adaptive immune response³³. This is conceptually supported by a secondary analysis of the VANISH clinical trial that showed that patients with an adaptive immune transcriptomic profile were harmed by steroids compared to patients with an innate immune profile³⁴. Moreover, the

risk of secondary infection, and delirium may outweigh the relative vasopressor effect in this population.

In the eICU-MIMIC cohort, corticosteroid use in the RW group was not significantly associated with the change of 28-day mortality. However, in the CEDAR cohort, corticosteroid use in the RW group was associated with increased 28-day mortality (HR: 1.24, 95% CI: 1.05–1.46), albeit not as much increase as in the RI group. For secondary outcomes, the RW group in eICU-MIMIC showed associations between corticosteroid use and shorter durations of mechanical ventilation and ICU stay. In CEDAR, similar associations were observed for ICU stay. These results validated the results from recent relevant trials^{10,34–37}. The finding that steroids benefit patients who are predicted to have worsening organ dysfunction is consistent with the proposed biologic mechanism for the action of steroids in critically ill patients as RW patients may benefit from the vasopressor effect of systemic corticosteroid to improve tissue perfusion and ameliorate visceral organ dysfunction³⁸. The steroid effect on immune cell death through activation of the apoptotic pathways may help limit the dysregulated innate inflammatory response and potentially disrupt the beneficial adaptive immune response. These hypotheses should be tempered, however, by the existing evidence that the adrenocortical candidate gene expression did not predict steroid response in randomized clinical trials^{16,39}. Moreover, traditional markers of suppression of the hypothalamic pituitary axis such as serum cortisol, do not reliably predict the response to corticosteroids⁴⁰. Lastly, there is mounting evidence that pulmonary sources of sepsis may benefit from corticosteroids compared to other sources⁵.

This research compliments the prior works exploring trajectories in the context of critical illness and sepsis. Specifically, in a secondary analysis of a multicenter randomized clinical trial evaluating balanced crystalloids compared to normal saline⁴¹, Bhavani and colleagues demonstrated that patterns of vital sign trajectories interacted with treatment effects⁴². The patients with lowering temperature, heart rates, and systolic blood pressure in an eight-hour period benefited from balanced crystalloids. In another study focusing on pediatric multiple organ dysfunction, it was shown that corticosteroid treatment interacted with trajectory motif defined subphenotypes by either increasing or decreasing the number of vasoactive agent free days¹⁷. These studies were association analysis which identified the subphenotypes first using the entire trajectory and then investigated the interactions between treatment and outcomes for each subphenotype. They cannot infer the causal treatment effect because they utilized post-treatment information when identifying the subphenotypes. Our work makes a step forward as only information available within the enrollment window was used for determining the subphenotype assignment. One potential concern is that the treatment is associated with the organ dysfunction trajectory, thus patient stratification by predicting the trajectory induced subphenotypes may introduce potential bias. However, we do not have perfect prediction performance, and propensity score matching was adopted to balance the treated and untreated groups, which mitigated the impact of such bias to a certain extent. We will continue developing methodologies for further mitigating such bias in our future research.

As a comparison, we also explored methods for estimating individualized treatment effects with the “optimal individual model” presented by Pirracchio et al.³⁰, which built a predictive model taking patient baseline covariates and the dichotomous treatment indicator as inputs to predict the outcome (e.g., mortality). We found it was challenging for the model to achieve satisfactory performance (AUROC 0.5746 ± 0.0038 for 28-day mortality prediction, see Supplementary Text 3 for details), which highlighted the inherent complexity of this problem. Meanwhile, the ITE approach produced results directionally consistent with our primary findings.

Our findings suggest several avenues for future investigation. First, our study regarding the evaluation of the effectiveness of corticosteroids in patients with sepsis was conducted through a trial-emulated

framework with real-world data, which could be important complementary evidence to the findings of RCTs^{10,43}. Second, we derived differential hazard ratios based on different predicted organ function trajectory patterns. This suggests that the risk benefit profile of corticosteroids could vary by the expected progression history of the patient. Third, the ML we developed for stratifying patients can facilitate timely trial enrollment or interventions before disease trajectory is established.

This study is not without limitations. First, we used a modification of the Sepsis-3 framework to identify patients for this study examining the trajectory of organ dysfunction. The individual components of the SOFA score may not fully capture the scope of organ failure in sepsis⁴⁴, and thus the subphenotypes derived on SOFA score trajectories may not fully encompass the complexity of organ failure progression. Second, we used real-world EHR data from ICUs, which may be subject to inherent biases such as missing data, measurement errors, and variability in data collection practices across different institutions. Third, we also assume that corticosteroids do not play an influential role in group enrollment as with the trial emulation setup, treatment could occur a few hours before group assignment and may affect predicted trajectory subgroups. This is a limitation of real-world data, as in RCTs, steroids and cohort determination would occur simultaneously. To limit the bias of this we restricted the baseline period to 24 h limiting the potential for steroids to influence predicted trajectory. The sensitivity analysis where we only included CS administration during baseline further mitigates the potential effect that receiving CS before RI/RW group assignment could have. For example, the decrease in norepinephrine doses and shock resolution were not seen in the ADRENAL trial until two to three days after corticosteroid administration¹⁰. Fourth, the study's observational design implies that despite rigorous propensity score matching, residual confounding cannot be completely ruled out. Although we made significant efforts to balance confounders across treatment groups, some unmeasured or unknown confounders might still influence the observed results (Supplementary Table 16 and Supplementary Text 4). Future studies should aim to identify and adjust for additional confounders, possibly through instrumental variable analyses or randomized controlled trials, to confirm these associations³². Lastly, the performance of our subphenotype prediction model could have been better. In the future, discrimination between RI and RW groups could be enhanced through testing steroid responsive transcriptomics or protein pathway enrichment by group⁴⁵. The differences observed between the development and validation cohorts highlight the potential variability in the association between corticosteroid use and mortality in the RW group. In the eICU-MIMIC cohort, corticosteroid use was not significantly associated with increased mortality in the RW group, whereas in the CEDAR cohort, it was associated with a statistically significant increase in 28-day mortality. These discrepancies may be attributed to differences in patient populations, clinical practices, unmeasured confounders, or sample sizes between the cohorts. The CEDAR cohort, being a single-center dataset, may have specific characteristics or practices influencing the outcomes.

Our study provides a proof-of-concept view of the role of corticosteroids in sepsis management, in several overlapping syndromic disease states. We identified differential effects when stratified by predicted organ dysfunction trajectory based subphenotypes. These findings highlight the need for further validation but suggest the possibility of more personalized treatment strategies in sepsis, informed by subphenotype rather than syndrome. Our future research will focus on validating these predicted trajectory subphenotypes in independent cohorts, in randomized clinical trials as well as exploring their biological basis.

Methods

Patient consent and human subjects

Approval for use of the MIMIC-IV, eICU, and WC-CEDAR databases was obtained from the Institutional Review Board at Weill Cornell Medicine

(protocols 1811019761 and 1706018287). All raw data in MIMIC-IV and eICU are publicly available through PhysioNet; we satisfied PhysioNet's research credentialing requirements. CEDAR data were de-identified in accordance with the HIPAA Safe Harbor standard prior to transfer. Because this study comprised secondary analysis of existing de-identified data and involved no interaction with human subjects, the IRB granted a waiver of informed consent. All data used were anonymized prior to analysis.

Data source and cohort

Three distinct data sources were used in our study, including MIMIC-IV, eICU, and CEDAR. The MIMIC-IV database was derived from Beth Israel Deaconess Medical Center, which is a teaching hospital of Harvard Medical School in Boston, Massachusetts with 673 licensed beds, including 493 medical/surgical beds, 77 critical care beds, and 62 OB/GYN beds. This database covered a decade of admissions between 2008 and 2019²⁴. The eICU database was built based on multi-center data from patients who were admitted to one of 335 units at 208 hospitals located throughout the US between 2014 and 2015²⁵. Both MIMIC-IV and eICU databases are publicly available data sources. The CEDAR database was built on New York-Presbyterian/Weill Cornell Medical Center (NYP/WCMC), which included 862 beds in total²⁶. This database is private and includes ICU admissions dating from 2001 to 2020. We combined the MIMIC-IV and eICU databases together in our primary analysis as an eICU-MIMIC cohort and validated them on the private database (CEDAR).

Target trial emulation and specification

This study employs an ITT analysis. We defined infectious critical illness through a modification of the Sepsis-3 criteria²⁷. The criteria required a suspected infection, indicated by the administration of antibiotics and the collection of a body fluid culture specimen. We required any total SOFA score of at least 2, instead of an increase of 2 over a baseline, as having an infectious critical illness. The enrollment window was defined as the 24-h period after ICU admission. The eligibility criteria for inclusion in the study were as follows: patients aged 18 or older at enrollment window (baseline), meet our infectious critical illness criteria within the enrollment window, no history of suspected infection and no prescription of corticosteroids before the enrollment window. The inclusion-exclusion cascade for the patients in our study is shown in Supplementary Figs. 1–3 for the three cohorts. Two treatment strategies were compared in this study. Strategy 0 (control group) involved no initiation of any corticosteroids drug within a window of 10 h before to 24 h after ICU admission, while Strategy 1 (treatment group) involved the initiation of hydrocortisone at a dose of at least 160 mg per day during that same window. A hydrocortisone dosage of 160 mg is comparable to 40 mg prednisone or 32 mg methylprednisolone. We compute cumulative milligram dosing of hydrocortisone within the enrollment window, and if a patient received at least 160 mg per day hydrocortisone equivalent, they are denoted as having corticosteroid exposure. We consider corticosteroids including Prednisolone, Prednisone, Hydrocortisone, Dexamethasone, and Methylprednisolone. A table showing the breakdown of corticosteroid usage for treated patients is shown in Supplementary Table 1. Patients who received less than 160 mg of hydrocortisone equivalent within the window were not included in the treated group. In accordance with ITT analyses, patients who received 160 mg per day after the enrollment window were not censored and stayed assigned to their initial treatment group. Baseline characteristics of the patients assigned to different treatment strategies were balanced through propensity matching to mimic the randomization process. We handled immortal time bias by doing the following: (1) Time zero for treated patients was set to time of treatment to remove immortal time between eligibility and exposure; (2) cloning all patients who died at the enrollment window (on Day 1) to both treatment

strategies. For (2), we conducted sensitivity analyses where patients who died within the enrollment window were excluded³¹.

Multiple types of baseline covariates including vital signs, laboratory measurements, and demographics were selected. The vital signs included heart rate, mean arterial pressure, respiratory rate, oxygen saturation (SpO₂), systolic arterial blood pressure (Systolic ABP), and temperature. These variables were selected as they provide crucial information about the patient's physiological status and are routinely monitored in ICU settings. Laboratory measurements comprised a broad range of biochemical, hematological, and physiological parameters. Biochemical measurements included albumin, alanine aminotransferase (ALT), aspartate aminotransferase (AST), bilirubin, blood urea nitrogen, chloride, creatinine, C-reactive protein (CRP), glucose, lactate, and sodium. Hematological measurements included bands, hemoglobin, international normalized ratio (INR), platelet count, and white blood cell count. Physiological parameters included the fraction of inspired oxygen (FiO₂), Glasgow Coma Scale score, arterial oxygen partial pressure (PaO₂), and urine output. These laboratory measurements were selected as they provide a comprehensive overview of the patient's organ function and metabolic status. The SOFA score subcomponents, including respiration score, cardiovascular score, central nervous system (CNS) score, liver score, coagulation score, and renal score, were also included. These scores provide a quantifiable measure of organ dysfunction, vasopressor administration, and mechanical ventilation, which are key aspects of sepsis. Demographic data included age, sex, and body mass index (BMI). Age and BMI were binned to facilitate analysis, with BMI bin ranges defined according to the World Health Organization guidelines. In addition, The Elixhauser Comorbidity Index was used to evaluate comorbidities based on past medical history⁴⁶. Regarding preprocessing those covariates, we removed some abnormal values if they were above the 99th percentile and used the median value to fill in missing values. Missingness of covariates are shown in Supplementary Table 2. If there are multiple values within the 24-h enrollment window, the worst clinical condition values were selected.

The study examined one primary outcome: 28-day mortality from ICU admission. Two secondary outcomes were also studied: time to ICU discharge from ICU admission, and time to cessation of mechanical ventilation from ICU admission. Time to mechanical ventilation cessation was defined as a period of 24 h without either invasive or non-invasive ventilation support. For time to cessation of mechanical ventilation, only patients who were on mechanical ventilation at baseline were included. For secondary outcomes, death was considered as a competing risk⁴⁷. Each patient was followed from his/her baseline until the day of his/her death, loss to follow-up, or discharge, whichever occurred first (Supplementary Fig. 4). The study aimed to estimate the observational analog of the ITT effect, providing a robust analysis of the potential impact of corticosteroid treatment on sepsis outcomes. More details about target trial emulation specifications are shown in Supplementary Tables 3–5.

Statistical methods

To balance the distribution of baseline covariates between the treatment and control groups, the study employed a propensity score matching technique²⁸. Propensity scores were estimated using a LR model, with the treatment assignment as the dependent variable and the baseline covariates as independent variables. The study implemented a nearest-neighbor matching algorithm based on the propensity scores, using a 1:4 matching ratio with repetition⁴⁸. This approach allowed each treated patient to be matched with up to four control patients. The Mahalanobis distance metric was used to measure the similarity in propensity scores between treated and control patients. A caliper value, set as the median absolute deviation of the propensity scores, was used to define the maximum allowable difference in propensity scores for a match. This caliper was adjusted as

necessary during the iterative matching process to ensure optimal balance in the covariates between the treatment and control groups. The process was repeated until as many covariates as possible were adequately balanced, as indicated by a SMD below 0.1⁴⁹.

This study evaluated the associations between corticosteroid use and the time-to-event outcomes (mortality, time to ICU discharge, and time to cessation of mechanical ventilation within 28 days after ICU admission) in the treatment and control groups using a Cox proportional hazards model. This model allowed for the estimation of the hazard ratios of outcomes between the treatment and control groups, providing a measure of the effect of the treatment⁵⁰. The study also adjusted the follow-up time for different patient outcomes⁴⁷. In the analysis of 28-day mortality, patients who were discharged before the 28-day mark had their time-to-event set to the maximum of 28 days. This adjustment assumes the best-case scenario for these patients, that they would have survived if they had stayed in the ICU for the full 28 days. Conversely, for the outcomes of time to ICU discharge and time to cessation of mechanical ventilation, patients who died before the 28-day mark had their duration set to 28 days. This adjustment assumes the worst-case scenario for these patients, that they would not have been discharged or had their ventilation ceased if they had survived to the end of the 28-day period. Based on this experimental setup, a HR of less than 1 is associated with benefits from steroids on the specific outcome. 95% confidence intervals were estimated for each HR measurement.

Stratified analyses

To address the heterogeneity of sepsis, we stratified patients into subgroups based on disease trajectories to investigate the differential effectiveness of corticosteroid treatment. We use the RI and RW subphenotypes identified from the previous Xu et al. study²³, where patient with sepsis were identified using data in the first 6 h of ICU admission and subphenotypes were obtained based on the full 72 h data of ICU admission using DTW and hierarchical clustering methods. DTW is a computational algorithm used to measure the similarity between pairwise temporal sequences that may vary in speed or length. Unlike standard distance measures, DTW can align sequences that are out of phase by warping the time axis to find an optimal match between data points. Following Xu et al., we applied DTW to align and compare patients' SOFA score trajectories over time. By calculating the DTW distance between each pair of patient trajectories, we created a distance matrix that reflects the similarity of organ dysfunction progression between patients. Hierarchical clustering was then performed on this distance matrix to group patients into subphenotypes with similar organ dysfunction trajectories, specifically identifying RI and RW groups. Like Xu et al., we used data from the first 72 h of admission to get ground truth labels subphenotype assignment. Once we obtained these subphenotypes, we used them as the ground truth label to train a set of LR models with the baseline patient characteristics. We chose LR models because they are easy to interpret. Two LR models were trained, one to predict RI (labels were RI vs Not RI) and one to predict RW (labels were RW vs Not RW). All variables used in training the model are shown in Supplementary Text 1. Both L1 and L2 penalties were applied to the LR models with a mix ratio of 0.5, under the Elastic Net methodology, in order to prevent overfitting and improve model generalization⁵¹. Class weights were used to mitigate the impact of class imbalance. The models were trained on the eICU-MIMIC cohort and validated on both the eICU-MIMIC and CEDAR cohorts. The trained models generated a probability for each patient belonging to RI or RW. We used a cutoff of 0.6 to make the final prediction.

Sensitivity analyses

To ensure the robustness of our results, we performed the following sensitivity analyses. (1) Strategy 1 (treatment group) is changed to initiation of hydrocortisone at a dose of 200 mg per day within the enrollment window; (2) Steroid exposure period is limited to 0–24 h

post ICU admission; (3) Exclusion of patients who die during the exposure period (Day 1) instead of cloning; (4) Excluding laboratory values with high percentage of untested patients (ALT, AST, Albumin, Bands, Troponin T, CRP); (5) Inclusion of source of infection as a covariate for 28-day mortality. This can only be performed on the eICU dataset as date of diagnoses are not provided by other cohorts and hence, we could not confirm whether diagnoses were given at baseline. (6) Include patients whose corticosteroid treatment was initiated during the enrollment period but did not reach 160 mg hydrocortisone equivalent within that window. These patients were identified based on medication orders indicating the clinician's intention to administer corticosteroids, even if the full dosage was completed after the enrollment window. By including these patients, we aimed to capture those who were intended to be treated but did not meet the strict criteria of receiving the full dosage within the enrollment period. This change only added patients in the CEDAR cohort and was primarily patients who were treated with hydrocortisone. (7) To account for the high mortality rate in our study population and its potential impact on the interpretation of ICU length of stay and mechanical ventilation duration, we performed a sensitivity analysis using ICU-FDs and VFDs over a 28-day period (Supplementary Text 2 and Supplementary Table 14).

Our primary analysis tried to minimize the immortal time bias by setting time zero for treated patients at the time of treatment initiation and cloning patients who died during the enrollment window. We conducted an additional set of sensitivity analysis to confirm that our results were not materially impacted by the choice of time zero by defining time zero uniformly at ICU admission for both the treatment and control groups. Under this approach, individuals were cloned to both groups and then censored when they deviated from their assigned strategy. This method ensures that both treatment and control groups have the same starting time point, thereby eliminating any additional immortal time granted to one arm over the other. We implemented this procedure and repeated the target trial emulation under the new time zero definition. The results are shown in Supplementary Table 15.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The de-identified data utilized in this study for the development cohort (eICU and MIMIC-IV) can be accessed upon the approval of a formal proposal and the execution of a Data Access Agreement via Physio Net (<https://physionet.org/>). The CEDAR data are derived from WCM's electronic health record. Accordingly, while a de-identified version exists and can be made freely available upon request to confirm derivation of our aggregate variables and the analytic approach, we are unable to provide direct access due to institutional restrictions around providing unmanaged access to sensitive patient data. Qualified requesters seeking to validate our methodology can make a request for access to arch-support@med.cornell.edu, which will receive a response within ten business days. A fee may be assessed for provision of a secure computing environment if the requester is unable to demonstrate their ability to adhere to our institutional standards for data storage. Please send all other questions to the corresponding author, who will direct them accordingly. Source data supporting the finding within this paper are provided within this paper and in the Supplementary Information. Source data are provided with this paper.

Code availability

The source code pertinent to this research is publicly accessible. The primary repository is hosted on <https://github.com/surajraj99/Corticosteroids-in-Patients-with-Critical-Illness>. Source code for

determining subphenotypes can be found in https://github.com/xuzhenxing2019/sepsis_subphenotype.

References

- Maslove, D. M. et al. Redefining critical illness. *Nat. Med.* **28**, 1141–1148 (2022).
- Martin-Loeches, I., Diaz, E. & Valles, J. Risks for multidrug-resistant pathogens in the ICU. *Curr. Opin. Crit. Care* **25**, 433–441 (2019).
- Cecconi, M., Evans, L., Levy, M. & Rhodes, A. Sepsis and septic shock. *Lancet* **392**, 75–87 (2018).
- Rhee, C. et al. Prevalence, underlying causes, and preventability of sepsis-associated mortality in US acute care hospitals. *JAMA Netw. Open* **2**, e187571 (2019).
- Heming, N. et al. APROCCHSS investigators and CRICS-TRIGGERSEP network. Hydrocortisone plus fludrocortisone for community acquired pneumonia-related septic shock: a subgroup analysis of the APROCCHSS phase 3 randomised trial. *Lancet Respir. Med.* **12**, 366–374 (2024).
- Patel, G. P. & Balk, R. A. Systemic steroids in severe sepsis and septic shock. *Am. J. Respir. Crit. Care Med.* **185**, 133–139 (2012).
- Weil, M. H. Current concepts on the management of shock. *Circulation* **16**, 1097–1105 (1957).
- Villar, J. et al. Dexamethasone treatment for the acute respiratory distress syndrome: a multicentre, randomised controlled trial. *Lancet Respir. Med.* **8**, 267–276 (2020).
- Marik, P. E. The role of glucocorticoids as adjunctive treatment for sepsis in the modern era. *Lancet Respir. Med.* **6**, 793–800 (2018).
- Venkatesh, B. et al. Adjunctive glucocorticoid therapy in patients with septic shock. *N. Engl. J. Med.* **378**, 797–808 (2018).
- Rochwerg, B. et al. Corticosteroids in sepsis: an updated systematic review and meta-analysis. *Crit. Care Med.* **46**, 1411–1420 (2018).
- Wu, J. Y. et al. Efficacy and safety of adjunctive corticosteroids in the treatment of severe community-acquired pneumonia: a systematic review and meta-analysis of randomized controlled trials. *Crit. Care* **27**, 274 (2023).
- Steinberg, K. P. et al. Efficacy and safety of corticosteroids for persistent acute respiratory distress syndrome. *N. Engl. J. Med.* **354**, 1671–1684 (2006).
- Dequin, P. F. et al. Hydrocortisone in severe community-acquired pneumonia. *N. Engl. J. Med.* **388**, 1931–1941 (2023).
- Bosch, N. A. et al. Comparative effectiveness of fludrocortisone and hydrocortisone vs hydrocortisone alone among patients with septic shock. *JAMA Intern. Med.* **183**, 451–459 (2023).
- Antcliffe, D. B. et al. Transcriptomic signatures in sepsis and a differential response to steroids. From the VANISH randomized trial. *Am. J. Respir. Crit. Care Med.* **199**, 980–986 (2019).
- Sanchez-Pinto, L. N., Stroup, E. K., Pendergrast, T., Pinto, N. & Luo, Y. Derivation and validation of novel phenotypes of multiple organ dysfunction syndrome in critically ill children. *JAMA Netw. Open* **3**, e209271 (2020).
- RECOVERY Collaborative Group. et al. Dexamethasone in hospitalized patients with Covid-19. *N. Engl. J. Med.* **384**, 693–704 (2021).
- Verhoeff, P. A. et al. Analysis of protein biomarkers from hospitalized COVID-19 patients reveals severity-specific signatures and two distinct latent profiles with differential responses to corticosteroids. *Crit. Care Med.* **51**, 1697–1705 (2023).
- Sherman, R. E. et al. Real-world evidence - what is it and what can it tell us? *N. Engl. J. Med.* **375**, 2293–2297 (2016).
- Hernán, M. A., Wang, W. & Leaf, D. E. Target trial emulation: a framework for causal inference from observational data. *JAMA* **328**, 2446–2447 (2022).
- Dang, L. E. et al. A causal roadmap for generating high-quality real-world evidence. *JCTS* **7**, e212 (2023).
- Xu, Z. et al. Sepsis subphenotyping based on organ dysfunction trajectory. *Crit. Care* **26**, 197 (2022).

24. Johnson, A. E. W. et al. MIMIC-IV, a freely accessible electronic health record dataset. *Sci. Data* **10**, 1 (2023).
25. Pollard, TomJ. et al. The eICU collaborative research database, a freely available multi-center database for critical care research. *Sci. Data* **5**, 1–13 (2018).
26. Schenck, E. J. et al. Critical care Database for Advanced Research (CEDAR): an automated method to support intensive care units with electronic health record data. *J. Biomed. Inform.* **118**, 103789 (2021).
27. Singer, M. et al. The third international consensus definitions for sepsis and septic shock (Sepsis-3). *JAMA* **315**, 801–810 (2016).
28. Cohen, K. et al. Risk of persistent and new clinical sequelae among adults aged 65 years and older during the post-acute phase of SARS-CoV-2 infection: retrospective cohort study. *BMJ* **376**, e068414 (2022).
29. Schoenfeld, D. A., Bernard, G. R. & ARDS Network. Statistical evaluation of ventilator-free days as an efficacy measure in clinical trials of treatments for acute respiratory distress syndrome. *Crit. Care Med.* **30**, 1772–1777 (2002).
30. Pirracchio, R. et al. Assessment of machine learning to estimate the individual treatment effect of corticosteroids in septic shock. *JAMA Netw. Open* **3**, e2029050 (2020).
31. Hernán, M. A., Sauer, B. C., Hernández-Díaz, S., Platt, R. & Shrier, I. Specifying a target trial prevents immortal time bias and other self-inflicted injuries in observational analyses. *J. Clin. Epidemiol.* **79**, 70–75 (2016).
32. VanderWeele, T. J. & Ding, P. Sensitivity analysis in observational research: introducing the E-value. *Ann. Intern. Med.* **167**, 268–274 (2017).
33. Neyton, L. P. A. et al. Host and microbe blood metagenomics reveals key pathways characterizing critical illness phenotypes. *Am. J. Respir. Crit. Care Med.* Advance online publication. <https://doi.org/10.1164/rccm.202308-1328OC> (2024).
34. Gordon, A. C. et al. Effect of early vasopressin vs norepinephrine on kidney failure in patients with septic shock: the VANISH randomized clinical trial. *JAMA* **316**, 509–518 (2016).
35. Rygård, S. L. et al. Low-dose corticosteroids for adult patients with septic shock: a systematic review with meta-analysis and trial sequential analysis. *Intensive Care Med.* **44**, 1003–1016 (2018).
36. Pirracchio, R. et al. Patient-level meta-analysis of low-dose hydrocortisone in adults with septic shock. *NEJM Evid.* **2**. <https://doi.org/10.1056/evidoa2300034> (2023).
37. Sprung, C. L. et al. Hydrocortisone therapy for patients with septic shock. *N. Engl. J. Med.* **358**, 111–124 (2008).
38. Lamontagne, F. et al. Corticosteroid therapy for sepsis: a clinical practice guideline. *BMJ* **362**, k3284 (2018).
39. Cohen, J. et al. The relationship between adrenocortical candidate gene expression and clinical response to hydrocortisone in patients with septic shock. *Intensive Care Med.* **47**, 974–983 (2021).
40. Cohen, J. et al. Plasma cortisol, aldosterone, and ascorbic acid concentrations in patients with septic shock do not predict treatment effect of hydrocortisone on mortality. A nested cohort study. *Am. J. Respir. Crit. Care Med.* **202**, 700–707 (2020).
41. Semler, M. W. et al. Balanced crystalloids versus saline in critically ill adults. *N. Engl. J. Med.* **378**, 829–839 (2018).
42. Bhavani, S. V., Robichaux, C., Verhoef, P. A., Churpek, M. M. & Coopersmith, C. M. Using trajectories of bedside vital signs to identify COVID-19 subphenotypes. *Chest* **165**, 529–539 (2024).
43. Annane, D. et al. Hydrocortisone plus fludrocortisone for adults with septic shock. *N. Engl. J. Med.* **378**, 809–818 (2018).
44. Moreno, R. et al. The sequential organ failure assessment (SOFA) score: has the time come for an update? *Crit. Care* **27**, 15 (2023).
45. Wong, H. R. et al. Combining prognostic and predictive enrichment strategies to identify children with septic shock responsive to corticosteroids. *Crit. Care Med.* **44**, e1000–e1003 (2016).
46. Quan, H. et al. Coding algorithms for defining comorbidities in ICD-9-CM and ICD-10 administrative data. *Med. Care* **43**, 1130–1139 (2005).
47. Brock, G. N., Barnes, C., Ramirez, J. A. & Myers, J. How to handle mortality when investigating length of hospital stay and time to clinical stability. *BMC Med. Res. Methodol.* **11**, 144 (2011).
48. Bottigliengo, D. et al. Oversampling and replacement strategies in propensity score matching: a critical review focused on small sample size in clinical settings. *BMC Med. Res. Methodol.* **21**, 256 (2021).
49. Liu, R., Wei, L. & Zhang, P. A deep learning framework for drug repurposing via emulating clinical trials on real-world patient data. *Nat. Mach. Intell.* **3**, 68–75 (2021).
50. Lin, D. Y. & Wei, L. J. The robust inference for the cox proportional hazards model. *J. Am. Stat. Assoc.* **84**, 1074–1078 (1989).
51. De Mol, C., De Vito, E. & Rosasco, L. Elastic-net regularization in learning theory. *J. Complex.* **25**, 201–230 (2009).

Acknowledgements

The authors would like to acknowledge the support from National Science Foundation awards (nos. 1750326 and 2212175), National Institute of Health awards (nos. RF1AG072449, R01AG080624, R01AG076448, RF1AG084178, R01AG076234, R01AG080991, and R01MH 124740), National Institute of Health award NHLBI K23 HL151876-01A1, and National Institute of General Medical Sciences award number K23 GM151730-01 for this study.

Author contributions

S.R., E.S., F.W., and Z.X. conceived the study. S.R., E.S., Z.X., W.P., and C.Z. conceived the method and designed the algorithmic techniques. S.R. wrote the codes and performed the computational analysis with input from Z.X., W.P., C.Z. All the authors (S.R., Z.X., W.P., C.Z., I.S., L.T., J.X., J.B., E.S., and F.W.) read the paper and suggested edits. F.W. supervised the project.

Competing interests

E.S. received personal fees from Axle Informatics outside of stated work. The remaining authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-025-59643-z>.

Correspondence and requests for materials should be addressed to Edward J. Schenck or Fei Wang.

Peer review information *Nature Communications* thanks Nelson Sanchez-Pinto, Jean-François Timsit, and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025