CHEST

Effects of Compliance With the Early Management Bundle (SEP-1) on Mortality Changes Among Medicare Beneficiaries With Sepsis

A Propensity Score Matched Cohort Study

Sean R. Townsend, MD, FCCP; Gary S. Phillips, MAS; Reena Duseja, MD; Lemeneh Tefera, MD; Derek Cruikshank, PSM; Robert Dickerson, RRT, MSHSA; H. Bryant Nguyen, MD; Christa A. Schorr, DNP, RN; Mitchell M. Levy, MD, FCCP; R. Phillip Dellinger, MD, FCCP; William A. Conway, MD; Warren S. Browner, MD, MPH; and Emanuel P. Rivers, MD, MPH, FCCP

BACKGROUND: US hospitals have reported compliance with the SEP-1 quality measure to Medicare since 2015. Finding an association between compliance and outcomes is essential to gauge measure effectiveness.

RESEARCH QUESTION: What is the association between compliance with SEP-1 and 30-day mortality among Medicare beneficiaries?

STUDY DESIGN AND METHODS: Studying patient-level data reported to Medicare by 3,241 hospitals from October 1, 2015, to March 31, 2017, we used propensity score matching and a hierarchical general linear model (HGLM) to estimate the treatment effects associated with compliance with SEP-1. Compliance was defined as completion of all qualifying SEP-1 elements including lactate measurements, blood culture collection, broad-spectrum antibiotic administration, 30 mL/kg crystalloid fluid administration, application of vasopressors, and patient reassessment. The primary outcome was a change in 30-day mortality. Secondary outcomes included changes in length of stay.

RESULTS: We completed two matches to evaluate population-level treatment effects. In standard match, 122,870 patients whose care was compliant were matched with the same number whose care was noncompliant. Compliance was associated with a reduction in 30-day mortality (21.81% vs 27.48%, respectively), yielding an absolute risk reduction (ARR) of 5.67% (95% CI, 5.33-6.00; $P < .001$). In stringent match, 107,016 patients whose care was compliant were matched with the same number whose care was noncompliant. Compliance was associated with a reduction in 30-day mortality (22.22% vs 26.28%, respectively), yielding an ARR of 4.06% (95% CI, 3.70-4.41; $P < .001$). At the subject level, our HGLM found compliance associated with lower 30-day risk-adjusted mortality (adjusted conditional OR, 0.829; 95% CI, 0.812-0.846; $P < .001$). Multiple elements correlated with lower mortality. Median length of stay was shorter among cases whose care was compliant (5 vs 6 days; interquartile range, 3-9 vs 4-10, respectively; $P < .001$).

INTERPRETATION: Compliance with SEP-1 was associated with lower 30-day mortality. Rendering SEP-1 compliant care may reduce the incidence of avoidable deaths.

CHEST 2022; 161(2):392-406

KEY WORDS: compliance; length of stay; Medicare; mortality; propensity score matching; sepsis; sepsis bundles; septic shock; severe sepsis

FOR EDITORIAL COMMENT, SEE PAGE 303

Take-home Points

Study Question: What is the association between compliance with the Early Management Bundle, Severe Sepsis/Septic Shock (SEP-1) and 30-day mortality among Medicare beneficiaries?

Results: In this post hoc analysis of data submitted by US hospitals to Medicare, compliance with SEP-1 was associated with a significant reduction in 30-day mortality. Among patients matched by propensity score using standard matching criteria, there was a 5.67% absolute risk reduction (ARR) in 30-day mortality, and when applying the most stringent literature-based matching criteria, there was a 4.06% ARR.

Interpretation: If the relationship between SEP-1 compliance and 30-day mortality is causal, rendering SEP-1 compliant care likely reduces the incidence of avoidable deaths.

In the prelude to a national sepsis measure, hospitalization rates for sepsis more than doubled from 2000 to 2008.[1,2] During this time, sepsis was present in > 50% of US hospital deaths and was the costliest disease at $24 billion annually.[3,4] Congress encouraged the Centers for Medicare and Medicaid Services (CMS) to improve sepsis care after gripping testimony from Ciaran Staunton, father of Rory Staunton, a boy who died of sepsis.[5] CMS identified sepsis as a priority and in 2013 asked the Measure Applications Partnership, a multistakeholder group convened by the National Quality Forum and required by Congress to review quality measures for CMS, to review measures for possible inclusion in Medicare's Inpatient Quality Reporting Program.[6]

In 2014, the National Quality Forum reendorsed a measure developed by Henry Ford Hospital (Detroit, MI), which CMS incorporated as SEP-1, into the Inpatient Quality Reporting Program on October 1, 2015. Controversy remains regarding the efficacy of SEP-1, which may have limited compliance with the measure.[7] Using data from CMS, we examined the association between compliance with SEP-1 and 30-day mortality.

Study Design and Methods

Setting, Population, and Overview

We used chart-abstracted data from 1,312,024 patients obtained from October 1, 2015, to March 31, 2017, at 3,241 US hospitals. Hospitals submitted data from all or a predefined sample of patients with severe sepsis or septic shock. SEP-1 specifies an algorithm to ascertain compliance including a start time (time zero) set by suspicion of infection, two systemic inflammatory response syndrome criteria, and an organ dysfunction, or a provider's designation of time zero. Elements included measurement of lactate, obtaining blood cultures, broad-spectrum antibiotic administration, 30 mL/kg of crystalloid fluids for hypotension or lactate ≥ 4.0 mM, vasopressors for persistent hypotension, lactate remeasurement if elevated, and patient reassessment (see e-Table 1 for description of the SEP-1 measure). CMS' Clinical Data Abstraction Center randomly audits 600 hospitals annually for accuracy of abstracted data.[8]

Patients ≥ 18 years of age with a qualifying *International Classification of Diseases, Tenth Revision, Clinical Modification* code were eligible for SEP-1 if hospitalized < 120 days (see e-Table 2 for qualifying codes). Hospital personnel audited the sample to verify patients met clinical criteria and excluded those that did not (see e-Table 3 for SEP-1 clinical criteria). The SEP-1 dataset included demographics, timestamps, diagnosis codes, and limited clinical data. We obtained Medicare-verified dates of death (Fig 1, box 2). Compliance was defined as passing all eligible SEP-1 elements. The algorithm

ABBREVIATIONS: AOR_c = conditional adjusted OR; ARR = absolute risk reduction; ATT = average treatment effect among the treated; CMS = Centers for Medicare and Medicaid Services; HGLM = hierarchical generalized linear model; IQR = interquartile range; LPS = logit of the propensity score; NNT = number needed to treat; RR = relative risk; SMD = standardized mean difference

AFFILIATIONS: From the Division of Pulmonary, Critical Care Medicine (S. R. Townsend), California Pacific Medical Center, San Francisco, CA; the Department of Medicine (S. R. Townsend), University of California San Francisco School of Medicine, San Francisco, CA; retired from the Center for Biostatistics (G. S. Phillips), Department of Biomedical Informatics, The Ohio State University, Columbus, OH; the Center for Clinical Standards and Quality (R. Duseja), Centers for Medicare and Medicaid Services, Baltimore, MD; the Department of Emergency Medicine (L. Tefera), Alameda Health System, Oakland, CA; General Dynamics Information Technology (P. Cruikshank), Coralville, IA; Mathematica Policy Research (R. Dickerson), Princeton, NJ; the Division of Pulmonary, Critical Care, Hyperbaric, Allergy and Sleep Medicine (H. B. Nguyen), Loma Linda University, Loma Linda, CA; the Cooper Health System (C. A. Schorr and R. P. Dellinger), Camden, NJ; the Division of Pulmonary, Critical Care and Sleep Medicine (M. M. Levy), Rhode Island Hospital, Providence, RI; the Warren Alpert School of Medicine at Brown University (M. M. Levy), Providence, RI; the Department of Internal Medicine (W. A. Conway), Henry Ford Hospital, Detroit, MI; the Wayne State University (W. A. Conway and E. P. Rivers), Detroit, MI; the California Pacific Medical Center Research Institute (W. S. Browner), San Francisco, CA; the Department of Epidemiology and Biostatistics (W. S. Browner), University of California, San Francisco, CA; and the Department of Emergency Medicine and Surgery (E. P. Rivers), Henry Ford Hospital, Detroit, MI.

FUNDING/SUPPORT: This study was funded by the Centers for Medicare and Medicaid Services with respect to data and statistical support. A grant from the Gordon and Betty Moore Foundation to the Society of Critical Care Medicine (SCCM) supported the contributions of Mr Phillips. SCCM provided essential organizational and intellectual talent to the development, endorsement, and re-endorsement of SEP-1 since inception.

CORRESPONDENCE TO: Sean R. Townsend, MD, FCCP; email: townsesr@sutterhealth.org

DOI: https://doi.org/10.1016/j.chest.2021.07.2167

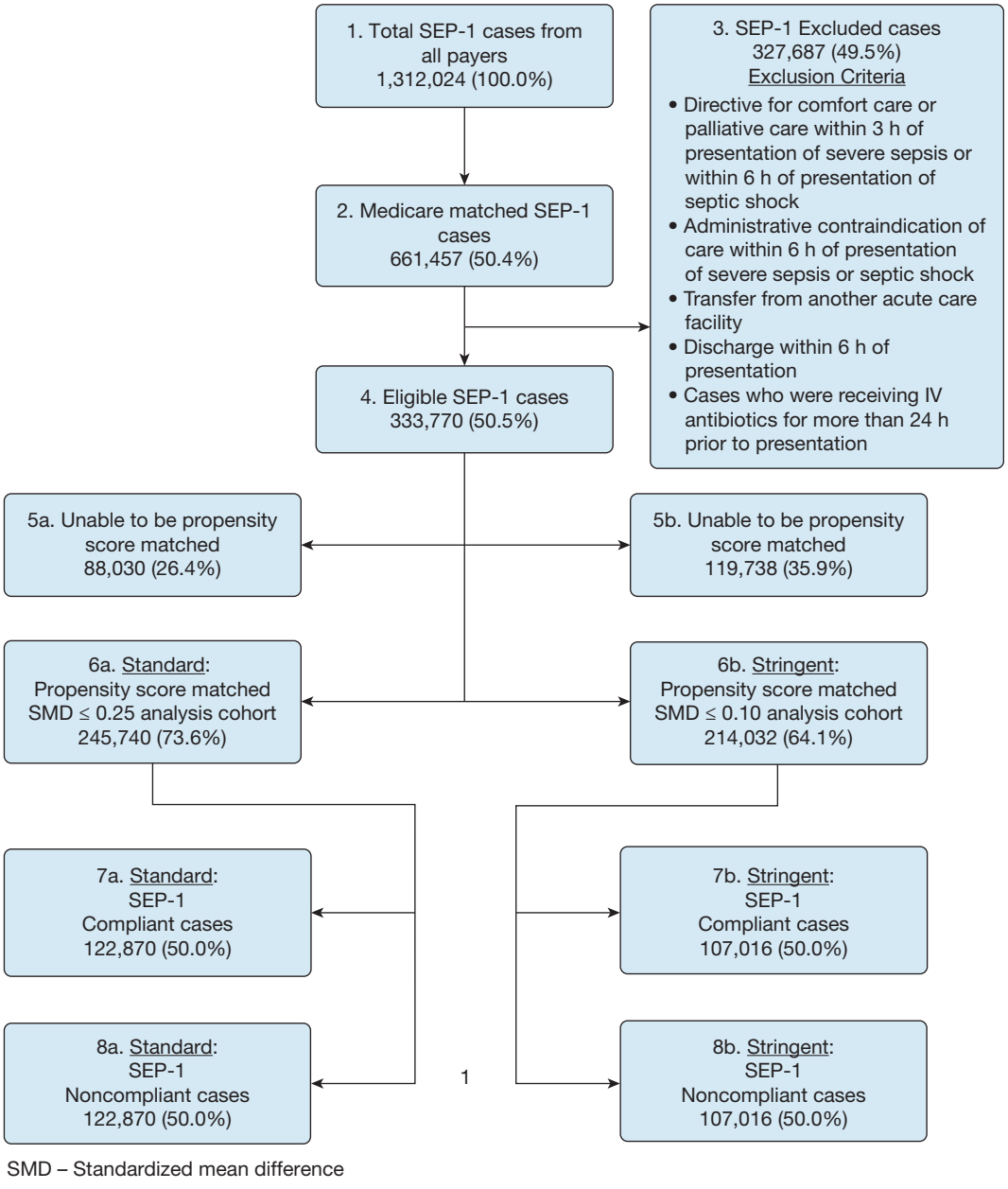

Figure 1 – *Consolidated Standards of Reporting Trials diagram. SMD = standardized mean difference.*

excluded 327,687 patients leaving 333,770 in the SEP-1 eligible cohort (Fig 1, boxes 3 and 4). The Third International Consensus Definitions for Sepsis and Septic Shock were not applied because SEP-1 antedated their release.[9]

In overview of our methods, we analyzed the association between compliance with SEP-1 and mortality. We used propensity score matching to estimate the marginal treatment effect, the average effect at the population level of moving a cohort of patients from untreated to treated,[10] specifically estimating the average treatment effect among the treated (ATT).[11] We also estimated the conditional treatment effect, the average effect at the subject level of moving a subject from untreated to treated,[10] using a hierarchical generalized linear model (HGLM). Mortality hereinafter refers to 30-day mortality unless otherwise stated. This study received a waiver of consent from the institutional review board for human research at Henry Ford Hospital, Detroit, Michigan (reference No. 12252).

Covariate Identification and Derivation

We sought to identify which variables were associated with both the exposure and the outcome, potentially confounding results. To inform our selection of variables, we relied on sepsis bundle studies (finding increasing compliance) and previously validated administrative sepsis mortality models (studying the model proposed by Darby et al[12] most closely). To ensure that these relationships were properly mapped, we organized the collected variables as a directed acyclic graph[13] (e-Fig 1) using a publicly available, web-based environment (http://dagitty.net). Studying the resultant graph, we determined that many of the variables that confounded this

relationship were related to the severity of illness. Variables included comorbidities, infection sources, acute organ failures, demographics, among others. Following contemporary practice, we selected true confounders known prior to, or just after, any SEP-1 algorithm failure point for inclusion in our models.[14-16]

Variables were modeled as indicator covariates unless otherwise noted (see e-Table 4 for variable descriptions). Demographics included age (continuous), sex, race, and ethnicity. Hospital characteristics included square root of bed count (continuous), accreditation status, and urban, rural, or critical access designation. Comorbidities were defined as per Elixhauser et al[17] and Quan et al.[18] Acute organ failures were defined as per Elias et al[19] (see e-Table 5 for ICD-9 to IDC-10 crosswalk). Infection sources (bacteremia, septicemia, fungal, peritoneal, heart, upper respiratory, lung, CNS, GI, skin, and other infections) were modeled as hierarchical infection categories by strength of the infection's association with unadjusted mortality as per Darby et al[12] and Iwashyna et al[20] (see e-Appendix 1 for additional methods, e-Table 6 for *International Classification of Diseases, Ninth Revision, Clinical Modification* to *International Classification of Diseases, Tenth Revision, Clinical Modification* crosswalk, and e-Table 7 for model parameter estimates). Severity of illness covariates included lactate range ($\leq$ 2.0, $>$ 2.0 and $<$ 4.0, and $\geq$ 4.0 mM), persistent hypotension, and septic shock.

Model Construction and Matching

We estimated the propensity score for compliance using an HGLM using a binomial family and a logit link function accounting for patients clustered within hospitals.[21,22] We calculated the intraclass correlation coefficient then added patient-level fixed effects (infection source, acute organ failures, comorbidities, age, race, ethnicity, sex, initial lactate range, persistent hypotension, septic shock, discharge quarter) and hospital-level fixed effects (bed count; accreditation status; urban, rural, or critical access designation), iteratively checking fit.[23] The variable defining each hospital was included in both levels as a random effect. A similar HGLM predictive mortality model estimated conditional treatment effects. Calibration was assessed by plotting observed vs predicted outcome.[12] Discrimination was assessed by exporting HGLM predictive probabilities then applying logistic regression.[24]

Once the propensity score was generated, compliant cases were matched random seed, 1:1, without replacement, nearest neighbor using the Mahalanobis distance (constituted by logit of the propensity score [LPS] and binary lactate variables) within propensity score calipers.[25] Given competing literature-based standards to assess variable balance, standard match targeted covariate absolute standardized mean differences (SMDs) of $\leq$ 0.25,[26,27] and stringent match SMDs of $\leq$ 0.10.[28,29] We calculated the treated-to-control variance ratio.[26,27]

Outcome and Sensitivity Analyses

ATT was estimated as the mean mortality difference between groups after matching. We partitioned matched cohorts into deciles by propensity score. We used the general estimating equation with independent working correlation accounting for matched pairs and robust SEs. Estimands included the absolute risk reduction (ARR), relative risk (RR), number needed to treat (NNT), and marginal OR.[29-31] We used adaptive Gaussian quadrature to approximate maximum likelihood estimation.[32] To estimate conditional treatment effects, the compliance parameter estimate was exponentiated as a conditional adjusted OR (AOR_c). We studied quarter 4, 2015, to quarter 2, 2016, for element-level analysis because data were fully abstracted. We assessed length of stay as medians with interquartile ranges (IQRs) using the Wilcoxon rank-sum test.

Regarding sensitivity analyses, we assessed whether clustering of patients within hospitals was accounted for by our HGLM repeating matches using preferential within-cluster matching, which matches within the same then other clusters.[21,22] We also assessed whether Mahalanobis distance improved on propensity score matching, repeating matches using the propensity score alone. We performed quantitative bias analysis to derive an unmeasured confounder capable of nullifying our estimates. Analyses were performed using SAS 9.3 (SAS Institute Inc).

Results

Covariate Identification and Derivation

We selected true confounders from our causal directed acyclic graph for our models (e-Fig 1), and Table 1 depicts their distribution. The average age was 73.5 years. The commonest comorbidity was hypertension (70.9%), then fluid and electrolyte disorders (65.4%), and then diabetes (40%). The commonest acute organ failure was renal failure (47.9%) then respiratory failure (31.3%). The commonest infections were lung (34.3%), septicemia alone (24.5%), and genitourinary (24.3%) (Table 1). Unadjusted mortality was 26.7%.

Model Output and Match Results

The intraclass correlation coefficients between hospital and compliance and hospital and mortality were 0.145 and 0.29, respectively. Factors most strongly associated with compliance were initial lactate value, discharge quarter, persistent hypotension, septic shock, critical access hospital status, among others (see e-Table 8 for compliance model parameter estimates). The mortality model associations were similar excepting discharge quarter and critical access hospital status (see e-Table 9 for mortality model parameter estimates). The compliance and mortality models showed good discrimination with C-statistics of 0.835 and 0.788, respectively. For the compliance model, the calibration plot's slope and intercept were 1.02 and 0.008, respectively, and the mortality model's slope and intercept were 1.02 and 0.004, respectively (e-Fig 2).

Regarding match results, the caliper was set at 0.40 and 0.22 times the pooled estimate of the SD of the LPS for standard match and stringent match, respectively. Not all patients could be matched, leaving 245,740 (73.6%) and 214,032 (64.1%) in the standard match and stringent match analysis cohorts, respectively (Fig 1, boxes 6a and 6b). Observed characteristics among the patients whose care was compliant with SEP-1 were similar to their

TABLE 1] Case Characteristics of the SEP-1 Eligible Cohort Across Compliance and Noncompliance With the SEP-1 Measure

Variable	All Cases (N = 333,770)	Compliant (n = 140,504; 42.1%)	Noncompliant (n = 193,266; 57.9%)
Hospital certification bed count			
Mean ± SD	308.9 ± 273.4	293.2 ± 259.5	320.3 ± 282.6
Median	240	231	249
Interquartile range	138-390	130-375	144-405
Hospital accreditation			
AOA/HFAP	13,580 (4.1)	5,598 (4.0)	7,982 (4.1)
CIHQ	2,026 (0.6)	1,052 (0.7)	974 (0.5)
DNV	31,305 (9.4)	12,620 (9.0)	18,685 (9.7)
NONE	9,661 (2.9)	3,554 (2.5)	6,107 (3.2)
TJC	277,198 (83.0)	117,680 (83.8)	159,518 (82.5)
Critical access hospital	5,152 (1.5)	2,255 (1.6)	2,897 (1.5)
Urban hospital location	253,458 (75.9)	106,039 (75.5)	147,419 (76.3)
Hospital region			
New England	11,807 (3.5)	4,987 (3.6)	6,820 (3.5)
Northeast	27,725 (8.3)	11,250 (8.0)	16,475 (8.5)
Mideast	47,391 (14.2)	20,151 (14.3)	27,240 (14.1)
Southeast	75,951 (22.8)	31,902 (22.7)	44,049 (22.8)
Great Lakes	57,063 (17.1)	22,697 (16.1)	34,366 (17.8)
South	38,893 (11.7)	16,260 (11.6)	22,633 (11.7)
Midwest	15,950 (4.8)	6,753 (4.8)	9,197 (4.8)
Mountain West	9,096 (2.7)	4,002 (2.9)	5,094 (2.6)
Southwest	38,842 (11.6)	17,973 (12.8)	20,869 (10.8)
Northwest	11,052 (3.3)	4,529 (3.2)	6,523 (3.4)
Discharge quarter			
Quarter 4, 2015	51,785 (15.5)	18,158 (12.9)	33,627 (17.4)
Quarter 1, 2016	54,858 (16.4)	22,154 (15.8)	32,704 (16.9)
Quarter 2, 2016	53,003 (15.9)	23,754 (16.9)	29,249 (15.1)
Quarter 3, 2016	53,936 (16.2)	21,764 (15.5)	32,172 (16.7)
Quarter 4, 2016	57,953 (17.4)	25,402 (18.1)	32,551 (16.8)
Quarter 1, 2017	62,235 (18.6)	29,272 (20.8)	32,963 (17.1)
Patient Hispanic ethnicity (yes)	18,756 (5.6)	7,527 (5.4)	11,229 (5.8)
Patient sex (male)	169,804 (50.9)	72,224 (51.4)	97,580 (50.5)
Patient race			
American Indian/Alaska Native	2,210 (0.6)	864 (0.6)	1,346 (0.7)
Asian	6,983 (2.1)	3,397 (2.4)	3,586 (1.9)
Black	38,313 (11.5)	14,529 (10.4)	23,784 (12.3)
Native Hawaiian/Pacific Islander	941 (0.3)	452 (0.3)	489 (0.2)
Unable to determine	11,706 (3.5)	4,920 (3.5)	6,786 (3.5)
White	273,617 (82.0)	116,342 (82.8)	157,275 (81.4)
Patient age at admission, y			
Mean ± SD	73.5 ± 13.0	73.6 ± 13.0	73.4 ± 13.0
Median	74	75	74
Interquartile range	66-83	66-83	66-83

(Continued)

TABLE 1] *(Continued)*

Variable	All Cases (N = 333,770)	Compliant (n = 140,504; 42.1%)	Noncompliant (n = 193,266; 57.9%)
Length of stay, d			
Mean	7.8 ± 7.4	7.1 ± 6.4	8.3 ± 8.1
Median	6	5	6
Interquartile range	3-10	3-9	4-10
Persistent hypotension (yes)	16,666 (5.0)	5,008 (3.6)	11,658 (6.0)
Initial lactate level			
Not collected	44,838 (13.4)	0 (0.0)	44,838 (23.2)
≤ 2.0 mM	99,121 (29.7)	57,833 (41.2)	41,288 (21.4)
> 2.0 and < 4.0 mM	133,659 (40.1)	70,426 (50.1)	63,233 (32.7)
≥ 4.0 mM	56,152 (16.8)	12,245 (8.7)	43,907 (22.7)
Severe sepsis	263,238 (78.9)	120,479 (85.7)	142,759 (73.9)
Septic shock	70,532 (21.1)	20,037 (14.3)	50,495 (26.1)
Mortality	88,998 (26.7)	30,444 (21.7)	58,554 (30.3)
Severe sepsis	63,510 (24.1)	23,791 (19.8)	39,719 (27.8)
Septic shock	25,488 (36.1)	6,653 (33.2)	18,835 (37.3)
Site of infection			
Bacteremia	160 (0.1)	55 (0.1)	105 (0.0)
CNS	849 (0.2)	331 (0.2)	518 (0.3)
Fungal	4,814 (1.4)	1,887 (1.3)	2,927 (1.5)
GI	3,712 (1.1)	1,302 (0.9)	2,410 (1.3)
Genitourinary	81,089 (24.3)	35,425 (25.2)	45,664 (23.6)
Heart	2,539 (0.8)	980 (0.7)	1,559 (0.8)
Lung	114,391 (34.3)	50,878 (36.2)	63,513 (32.9)
Unknown	1,238 (0.4)	535 (0.4)	703 (0.4)
Other	29,596 (8.9)	12,600 (9.0)	16,996 (8.8)
Peritoneal	6,778 (2.0)	1,726 (1.2)	5,052 (2.6)
Septicemia only	81,682 (24.5)	31,871 (22.7)	49,811 (25.8)
Soft tissue	6,131 (1.8)	2,547 (1.8)	3,584 (1.8)
Upper respiratory tract	791 (0.2)	367 (0.3)	424 (0.2)
Organ failures			
Cardiovascular	33,297 (10.0)	11,580 (8.2)	21,717 (11.2)
Hematologic	12,410 (3.7)	3,471 (2.5)	8,939 (4.6)
Metabolic	91,398 (27.4)	34,173 (24.3)	57,225 (29.6)
Neurologic	32,501 (9.7)	12,457 (8.9)	20,044 (10.4)
Renal	159,946 (47.9)	62,215 (44.3)	97,731 (50.6)
Respiratory	104,296 (31.3)	39,230 (27.9)	65,066 (33.7)
Comorbidities			
AIDS	995 (0.3)	426 (0.3)	569 (0.3)
Alcohol abuse	10,440 (3.1)	3,901 (2.8)	6,539 (3.4)
Chronic blood loss anemia	3,851 (1.2)	1,410 (1.0)	2,441 (1.3)
Chronic pulmonary disease	113,087 (33.9)	49,357 (35.1)	63,730 (33.0)
Coagulopathy	49,797 (14.9)	18,726 (13.3)	31,071 (16.1)
Congestive heart failure	102,126 (30.6)	39,890 (28.4)	62,236 (32.2)
Deficiency anemias	108,137 (32.4)	45,025 (32.1)	63,112 (32.7)

(Continued)

TABLE 1] *(Continued)*

Variable	All Cases (N = 333,770)	Compliant (n = 140,504; 42.1%)	Noncompliant (n = 193,266; 57.9%)
Depression	45,665 (13.7)	19,894 (14.2)	25,771 (13.3)
Diabetes with chronic complications	72,176 (21.6)	29,548 (21.0)	42,628 (22.1)
Diabetes without chronic complications	61,547 (18.4)	25,896 (18.4)	35,651 (18.5)
Drug abuse	7,223 (2.2)	3,060 (2.2)	4,163 (2.2)
Fluid and electrolyte disorders	218,179 (65.4)	88,210 (62.8)	129,969 (67.3)
Hypertension	236,558 (70.9)	100,108 (71.3)	136,450 (70.6)
Hypothyroidism	61,692 (18.5)	26,225 (18.7)	35,467 (18.4)
Liver disease	19,099 (5.7)	6,901 (4.9)	12,198 (6.3)
Lymphoma	5,912 (1.8)	2,539 (1.8)	3,373 (1.8)
Metastatic cancer	14,775 (4.4)	5,822 (4.1)	8,953 (4.6)
Obesity	51,505 (15.4)	20,646 (14.7)	30,859 (16.0)
Other neurologic disorders	67,194 (20.1)	29,319 (20.9)	37,875 (19.6)
Paralysis	31,207 (9.4)	13,430 (9.6)	17,777 (9.2)
Peptic ulcer disease excluding bleeding	4,914 (1.5)	1,750 (1.3)	3,164 (1.6)
Peripheral vascular disease	35,393 (10.6)	14,066 (10.0)	21,327 (11.0)
Psychoses	15,916 (4.8)	7,207 (5.1)	8,709 (4.5)
Pulmonary circulation disease	6,361 (1.9)	2,388 (1.7)	3,973 (2.1)
Renal failure	102,247 (30.6)	40,503 (28.8)	61,744 (32.0)
Rheumatoid arthritis/collagen vascular diseases	15,456 (4.6)	6,637 (4.7)	8,819 (4.6)
Solid tumor without metastasis	13,853 (4.2)	5,757 (4.1)	8,096 (4.2)
Valvular disease	30,745 (9.2)	12,451 (8.9)	18,294 (9.5)
Weight loss	55,562 (16.7)	21,558 (15.3)	34,004 (17.6)

Values are No. (%) or as otherwise indicated. AOA = American Osteopathic Association; cert = certification; CIQH = Center for Improvement in Healthcare Quality; DNV = Det Norske Veritas; HFAP = Healthcare Facilities Accreditation Program; Q = quarter; TJC = The Joint Commission.

matched counterparts (see e-Tables 10 and 13 for standard match and stringent match distributions, respectively). In standard match and stringent match, all variables achieved absolute SMDs ≤ 0.25 and 0.10 and variance ratios between 0.5 and 1.3 and 0.7 and 1.2, respectively (Fig 2; see e-Tables 11 and 14 for standard match and stringent match SMDs, respectively). Unmatched patients' covariate distributions are shown in e-Tables 12 and 15, respectively.

The propensity score performed well: the lowest decile in standard match and stringent match included patients with propensity score ranges of 1.4% to 24.3% and 1.4% to 27.0%, whereas the highest decile ranged from 77.8% to 98.5% and 76.2% to 98.3%, respectively (see e-Tables 16 and 17 for standard match and stringent match distributions by deciles, respectively). The score was inversely correlated with both the severity of illness and the frequency of underlying comorbidities. For example, in standard match and stringent match, acute organ failure percentages were highest in the first decile and progressively decreased (e-Tables 16, 17). The propensity score in both matches was associated with lower mortality across all deciles whether the actual care was compliant with SEP-1 (Fig 3).

Marginal and Conditional Treatment Effects and Length of Stay

Regarding marginal treatment effects and ATT, for standard match, compliance with SEP-1 was associated with a reduction in population-averaged mortality (21.81% vs 27.48%, respectively), yielding an ARR of 5.67% (95% CI, 5.33-6.00) and RR of 0.794 (95% CI, 0.783-0.805) (Table 2). These results were significantly decreased within each decile (Fig 3). We estimated the NNT of 17.65 (95% CI, 16.66-18.76) and marginal OR of 0.736 (95% CI, 0.723-0.750). For stringent match, compliance with SEP-1 was associated with a reduction in mortality (22.22% vs 26.28%, respectively), yielding an ARR of 4.06% (95% CI, 3.70-4.41) and RR of 0.846

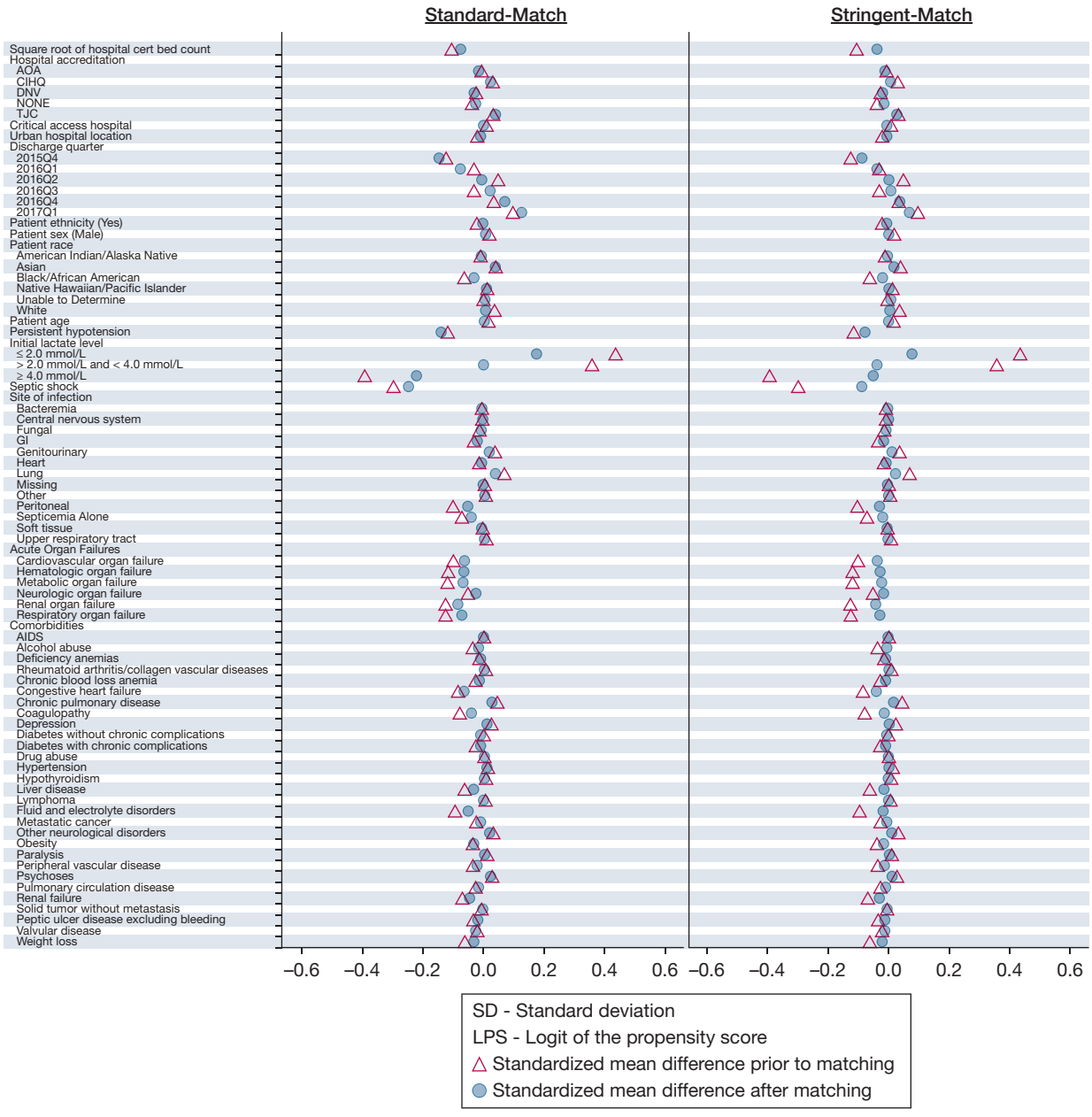

Figure 2 – *Standard match and stringent match standardized mean differences before and after matching where matching proceeded 1:1, without replacement, greedy nearest neighbor based on a random seed using Mahalanobis distance within propensity score calipers set at 0.40 and 0.22 SD of the logit of the propensity score (LPS), respectively. Standard match targets standardized mean differences in the range between −0.25 and 0.25, and stringent match targets between −0.10 and 0.10. Mahalanobis distance was defined by the LPS and three binary lactate variables (≤ 2.0, > 2.0 and < 4.0, ≥ 4.0 mM). AOA = American Osteopathic Association; cert = certification; CIQH = Center for Improvement in Healthcare Quality; DNV = Det Norske Veritas; HFAP = Healthcare Facilities Accreditation Program; Q = quarter; TJC = The Joint Commission.*

(95% CI, 0.833-0.858) (Table 2). These results were significantly decreased within each decile (Fig 3). We estimated the NNT of 24.6 (95% CI, 22.66-27.05) and marginal OR of 0.802 (95% CI, 0.786-0.818).

Regarding conditional, subject-level treatment effects, among the 333,770 patients in the SEP-1 eligible cohort, compliance was associated with lower observed and adjusted mortality (OR, 0.636; 95% CI, 0.626-0.647 and AOR_c, 0.829; 95% CI, 0.812-0.846, respectively) (Table 3). We estimated the effect of passing individual SEP-1 elements and bundles (e-Table 18 clarifies the SEP-1 bundle framework). The severe sepsis 3-h bundle (AOR_c, 0.803; 95% CI, 0.779-0.828) and each element were associated with decreased adjusted mortality. Repeat lactate collection (AOR_c, 0.885; 95% CI, 0.851-0.921) and crystalloid fluid

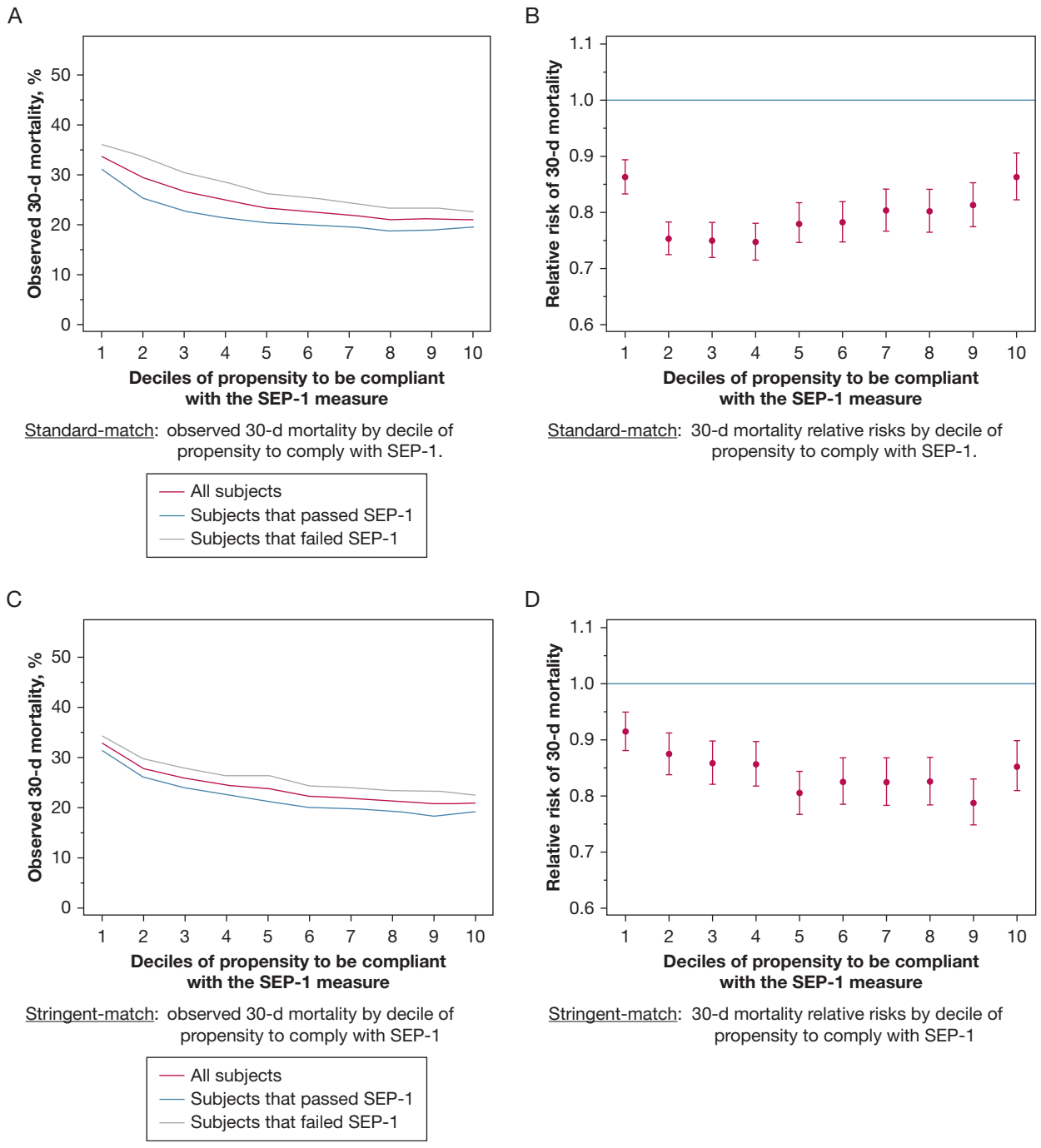

Figure 3 – *A-D, Standard match and stringent match outcomes across deciles of propensity to comply with SEP-1. Matching proceeded 1:1, without replacement, greedy nearest neighbor based on a random seed using Mahalanobis distance within propensity score calipers. Calipers were set at 0.40 and 0.22 SDs of the logit of the propensity score for standard match and stringent match, respectively. Mahalanobis distance was determined by the logit of the propensity score and three binary lactate variables (≤ 2.0, > 2.0 and < 4.0, ≥ 4.0 mM). Standard match targets standardized mean differences in the range between −0.25 and 0.25, and stringent match targets between −0.10 and 0.10.*

administration (AOR_c, 0.915; 95% CI, 0.855-0.980) were associated with decreased adjusted mortality. Vasopressor use was associated with higher mortality (AOR_c, 1.317; 95% CI, 1.126-1.541), and repeat assessment was not significant (AOR_c, 1.012; 95% CI, 0.920-1.114) (Table 3).

Median length of stay was shorter among all compliant cases (5 vs 6 days; IQR, 3-9 vs 4-10;

TABLE 2] Standard Match and Stringent Match 30-Day Mortality by Compliance Propensity Deciles

Match	Decile	Compliance Propensity Range (%)	No. of Compliant Cases	Compliant 30-d Mortality Rate (%)	No. of Noncompliant Cases	Noncompliant 30-d Mortality Rate (%)	Absolute 30-d Mortality Difference (%)	Relative Risk	Relative Risk 95% CI
Standard match	1	1.40-24.30	12,287	31.20	12,287	36.18	4.98	0.862	0.833-0.893
	2	24.30-32.93	12,287	25.35	12,287	33.69	8.34	0.752	0.724-0.782
	3	32.93-40.18	12,287	22.85	12,287	30.50	7.65	0.749	0.719-0.781
	4	40.19-46.80	12,287	21.35	12,287	28.63	7.28	0.746	0.713-0.779
	5	46.80-52.81	12,287	20.51	12,287	26.31	5.80	0.780	0.745-0.816
	6	52.81-58.64	12,287	19.97	12,287	25.56	5.58	0.782	0.746-0.818
	7	58.64-64.39	12,287	19.61	12,287	24.43	4.83	0.803	0.766-0.841
	8	64.40-70.45	12,287	18.73	12,287	23.38	4.66	0.801	0.763-0.841
	9	70.46-77.76	12,287	19.05	12,287	23.46	4.42	0.812	0.773-0.852
	10	77.76-98.49	12,287	19.53	12,287	22.66	3.13	0.862	0.821-0.905
	Total	1.40-98.49	122,870	21.81	122,870	27.48	5.67	0.794	0.783-0.805
Stringent match	1	1.40-27.02	10,702	31.43	10,702	34.39	2.95	0.914	0.880-0.950
	2	27.02-35.01	10,701	26.04	10,701	29.79	3.75	0.874	0.838-0.912
	3	35.01-41.25	10,702	23.97	10,702	27.93	3.96	0.858	0.821-0.898
	4	41.25-46.74	10,702	22.66	10,702	26.48	3.82	0.856	0.817-0.897
	5	46.74-51.99	10,701	21.26	10,701	26.44	5.18	0.804	0.767-0.844
	6	51.99-57.29	10,702	20.11	10,702	24.38	4.27	0.825	0.784-0.867
	7	57.29-62.72	10,702	19.83	10,702	24.06	4.23	0.824	0.783-0.867
	8	62.72-68.79	10,701	19.34	10,701	23.44	4.10	0.825	0.784-0.869
	9	68.79-76.21	10,702	18.37	10,702	23.33	4.96	0.787	0.747-0.830
	10	76.21-98.27	10,701	19.23	10,701	22.56	3.33	0.853	0.809-0.898
	Total	1.40-98.27	107,016	22.22	107,016	26.28	4.06	0.846	0.833-0.858

Standard match and stringent match 30-d mortality by compliance propensity deciles are shown where matching proceeded 1:1, without replacement, greedy nearest neighbor based on a random seed using Mahalanobis distance within propensity score calipers set at 0.40 and 0.22 SD of the logit of the propensity score, respectively. Mahalanobis distance was defined by the logit of the propensity score and three binary lactate variables (≤ 2.0, > 2.0 and < 4.0, ≥ 4.0 mM).

TABLE 3] Element-Level Unadjusted and Adjusted Conditional Treatment Effects Based on the Hierarchical Generalized Linear Model Logistic Regression Model

Bundle: Treatment Section and Elements	No. of Eligible Cases	Pass Rate (%)	Compliant 30-d Mortality (%)	Noncompliant 30-d Mortality (%)	Conditional Adjusted OR	Conditional Adjusted OR 95% CI	*P* Value
Complete SEP-1 bundle[a]	333,770	42.1	21.7	30.3	0.829	0.812-0.864	< .001
Severe sepsis 3 h: initial lactate level	159,646	86.0	26.2	32.0	0.772	0.743-0.802	< .001
Severe sepsis 3 h: antibiotic administration	137,252	88.5	25.8	29.0	0.844	0.798-0.892	< .001
Severe sepsis 3 h: blood culture	121,454	90.0	25.3	30.8	0.867	0.827-0.908	< .001
Severe sepsis 3-h bundle	159,646	68.5	25.3	30.8	0.803	0.779-0.828	< .001
Severe sepsis 6-h bundle: repeat lactate level	74,349	62.6	27.0	26.9	0.885	0.851-0.921	< .001
Shock 3-h bundle: crystalloid fluid administration	24,357	62.2	34.1	34.8	0.915	0.855-0.980	.011
Shock 6 h: vasopressors	5,332	77.3	39.3	29.1	1.317	1.126-1.541	< .001
Shock 6 h: reassessment	9,931	38.1	38.0	36.5	1.012	0.920-1.114	.807
Shock 6 h: vasopressors and reassessment	4,122	42.5	40.8	38.3	1.014	0.879-1.169	.846
Shock 6-h bundle	11,141	34.0	38.0	35.3	1.048	0.955-1.149	.326

[a]Data inclusive from quarter 4, 2015, to quarter 1, 2017; data in all other rows represent quarter 4, 2015, to quarter 2, 2016.

$P < .001$) and among cases of compliant survivors (5 vs 6 days; IQR, 3-8 vs 4-9; $P < .001$), respectively.

Sensitivity Analyses

All sensitivity match analyses revealed associations similar to standard match and stringent match (see e-Appendix 2 for additional methods and e-Table 19 for comparison details). The quantitative bias analysis suggested that our standard match and stringent match results would be robust unless the unmeasured confounder increased the odds of mortality by > 2.00 and 1.45 times, respectively (see e-Tables 20 and 21 for calculation details).

Discussion

Compliance with SEP-1 was associated with a significant reduction in mortality after adjusting for factors that influence the likelihood of compliance. Sorting patients into deciles by their propensity to comply, we observed lower mortality among compliant cases at each gradation, suggesting SEP-1 is broadly associated with lower mortality regardless of likelihood of compliance, severity of illness, or type of sepsis syndrome. Each bundle element that achieved statistical significance was associated with lower mortality, except vasopressors. Median length of stay was shorter among compliant cases.

In observational studies, unlike randomized controlled trials, treatment selection may be influenced by patients and providers, which may result in biased effect estimates. Directly analyzing the entire SEP-1 eligible population as groups of compliant and noncompliant cases would bias estimates of treatment effect because these groups would differ in clinical characteristics. Researchers often turn to multivariable regression to adjust for these differences.[29] These models typically estimate the change in odds of an outcome when moving a subject from untreated to treated, known as conditional estimates, reported as ORs.[10] For a national program addressing a prevalent disease (eg, SEP-1), population-averaged, or marginal, effect estimates are preferable.[11] Matching patients on observed characteristics by propensity score minimizes selection bias and yields marginal effect estimates.[33] After matching, the distribution of known covariates among groups is balanced, allowing for direct comparison of outcomes (eg, difference in group means). These

measures of association are preferred over conditional ORs, which are noncollapsible (ie, stratum-specific and not equivalent to marginal estimates).[30,31]

After matching, it is necessary to assess variable balance. Two standards have been advanced, absolute SMDs < 0.25 and 0.10,[26-28,34] which in our report are termed standard match and stringent match, respectively. Standard match yielded a 5.67% ARR compared with 4.06% for stringent match, with calipers set at 0.40 and 0.22 SD of LPS, respectively. Rosenbaum and Rubin[35] reported matching at 0.40 vs 0.20 SD of LPS eliminated 95% and 99%, respectively, of bias attributable to measured confounders.[36] The temptation to cure inexact matching by increasing the reduction in bias from 95% to 99% with tighter calipers should be dampened by the loss of 31,708 patients' information and the introduction of bias because of incomplete matching, which may be substantial.[37] Because researchers must weigh these considerations, we presented results from both matches. We also reported a conditional estimate that considered information from all 333,770 patients, finding an association between SEP-1 compliance and lower mortality (AOR_c, 0.829; 95% CI, 0.812-0.846).

Our results accord with other studies noting an association between compliance with sepsis bundles and decreased mortality, including those from New York after Rory's Regulations required hospitals to follow evidence-informed protocols.[38-41] Levy et al[38] studied 113,380 patients who had a protocol initiated and found decreased in-hospital mortality (28.8% vs 24.4%; $P <$ .001). This 4.4% ARR corresponds to our findings, and Levy et al[38] also found compliance with each bundle element lowered risk-adjusted odds of mortality, except vasopressor application.

The correlation between vasopressor application and death in observational studies has been long regarded as a marker of severity of illness.[42] Levy et al[38] found the association insignificant (AOR_c, 1.03; 95% CI, 0.97-1.10), whereas in our analysis it reached significance (AOR_c, 1.317; 95% CI, 1.126-1.541). A likely explanation is that patients with profound shock received vasopressors, whereas those with milder shock avoided them. The wide difference in observed mortality in our study among those compliant and noncompliant with vasopressor therapy (39.3% vs 29.1%) supports this reasoning. Moreover, the odds of death should be greater reporting 30-day vs in-hospital mortality.

Researchers commonly expect that more severely ill patients in observational trials are more likely to receive interventions.[43] All-or-none measurement upends this understanding because more severely ill patients qualify for more bundle elements, creating more opportunities to fail the measure. This explains the inverse association observed between the propensity score and mortality (Table 2). All-or-none measurement also means that many patients may have received very good care, passing nearly all qualifying elements, but they remain noncompliant. This narrows the observed mortality difference when comparing compliant and noncompliant patients.[44]

To our knowledge, our report is the largest study of SEP-1 to date, and each quarter > 99% of US hospitals report SEP-1 data to CMS.[45] An analysis of reporting hospitals found those that provide timely care on other measures (eg, time to ECG for myocardial infarction) were associated with improved SEP-1 performance, lending the measure construct validity as a marker of timely sepsis care.[46] Extending these findings, Bauer et al[47] studied 33,000 cases formally abstracted for SEP-1 at 12 Cleveland Clinic hospitals and found compliance was associated with lower risk-adjusted in-hospital mortality, lower risk-adjusted incidence of readmission, increased discharge to home, and decreased length of stay.

Notwithstanding these favorable studies, some reports reached heterogenous conclusions. A meta-analysis by Pepper et al[48] examined 17 sepsis bundle studies and found "available studies support the notion that antibiotic- and fluid-focused sepsis bundles like SEP-1 improve survival…," but also that evidence "…is lacking for the specificity required in [CMS'] Sepsis Bundle (SEP-1)." The meta-analysis, which excluded results from New York, called for high-quality evidence that the antibiotic timing and fluid dosing in SEP-1 improved outcomes compared with other choices. Our study of SEP-1 program data adds to the literature, manifestly bolstering the specific interventions in SEP-1.

A study by Baghdadi et al[49] examined 6,404 patients in the University of California system reaching varied conclusions. The study found that all-or-none SEP-1 care (modified as just lactate and blood culture collection, and antibiotic and fluid administration) was not associated with improved outcomes, but that multiple elements reduced mortality or decreased vasopressor days. The study defined community-acquired and hospital-onset cohorts, whereas we analyzed a single cohort. The improved outcomes with multiple bundle elements may have represented an early suggestion that a substantially larger sample evaluating

all SEP-1 elements would prove beneficial. The study raises pertinent questions about whether SEP-1 should measure hospital-onset sepsis, an avenue for future research.

Rhee et al[50] examined 851 SEP-1 cases from seven US hospitals and found higher unadjusted in-hospital mortality among noncompliant cases (OR, 1.82; 95% CI, 1.19-2.80), but no difference after adjusting for clinical characteristics (AOR_c, 1.94; 95% CI, 0.85-2.80). The model included whether patients presented with vague symptoms, which introduce delays and are associated with higher mortality.[51] Adjusting for vague symptoms is unnecessary because SEP-1 abstractors set time zero only after physicians document suspicion of infection.[52] Therefore, failures caused by delays cannot be attributed to vague symptoms because the clock was started only after symptoms were explicit. Because this adjustment biases results toward the null, reconsideration after removing the covariate would augment the literature.

Finally, Barbash et al[53] studied 54,000 cases from 11 hospitals in the University of Pittsburgh system. The time series analysis found improvement in compliance with certain process elements but no change in outcomes. In Barbash et al,[53] in-hospital mortality was low at 4.5% before SEP-1 and 4% after the inception of SEP-1, despite median ages compatible with a Medicare population (72 and 71 years of age, respectively). Our study of Medicare beneficiaries found average 30-day mortality to be 26.7%. The low mortality rates in Barbash et al[53] limit the study's generalizability to the SEP-1 population.

Regarding costs and the burden of abstraction on hospitals, Buchman et al[54] found Medicare's costs for sepsis admissions and skilled nursing care exceeded $41.5 billion annually—Medicare's most costly health care condition. Given this estimate, the burden of SEP-1 abstraction is contextually appropriate. Hospitals may submit 20% of their cases each quarter.[52] Abstractors spend 30 to 120 min abstracting each chart.[55,56] In the unusual circumstance that a hospital accrued 300 sepsis cases per quarter, abstraction would require less than one-quarter full-time employee.

Our study has limitations. Confounding may bias results in an observational study. Two studies have reported variability in abstracting time zero, which may impact compliance. The first study provided minimal training to nonprofessionals before assessing interrater reliability.[57] A second study found fair agreement initially but attained perfect reliability after improvement efforts.[55] CMS and key stakeholders are studying a revised time zero candidate, and CMS continues to audit hospitals for accuracy. We cannot report on balancing measures in detail (eg, the possibility of increased antibiotic usage, harm from crystalloid fluid administration). However, between 2012 and 2017, including after SEP-1 launched, inpatient US antibiotic usage remained stable[58] and multidrug-resistant organism and *Clostridium difficile* infections decreased.[59,60] A 30 mL/kg crystalloid infusion was associated with lower 30-day mortality in our study, lower in-hospital mortality in New York,[38] and at Kaiser Northern California, decreased 30-day mortality among patients with heart failure or chronic kidney disease.[41] Because SEP-1 selects patients from coded cases, one concern may be that coding inaccuracies influenced our results. Abstractors review each case to verify it meets clinical criteria for sepsis and exclude those that do not.

Interpretation

Our nationwide evaluation of SEP-1 found that overall compliance with the measure, with individual bundle elements, and within each decile of propensity to comply, was associated with a significant reduction in mortality at 30 days. Among all patients and survivors, we found median length of stay was decreased. If the relationship between compliance and mortality is causal, rendering SEP-1 compliant care may reduce the incidence of avoidable deaths.

Acknowledgments

Author contributions: S. R. T. had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis. S. R. T., W. S. B., and G. S. P. contributed to conception and design. S. R. T., D. C., G. S. P., E. P. R., L. T., R. Duseja, and W. S. B. contributed to analysis and interpretation of the data. S. R. T., G. S. P., W. S. B., R. Duseja, L. T., R. P. D., M. M. L., C. A. S., and H. B. N. contributed to drafting of the article. S. R. T., E. P. R., R. P. D., M. M. L., W. A. C., and W. S. B. contributed to critical revision of the article for important intellectual content. S. R. T., E. P. R., R. Duseja, G. S. P., and W. S. B. contributed to final approval of the article. D. C. and G. S. P. contributed to statistical expertise. R. Duseja and L. T. contributed to obtaining of funding. S. R. T., R. Dickerson, R. Duseja, L. T., E. P. R., and W. A. C. contributed to administrative, technical, or logistic support. S. R. T., D. C., G. S. P., and E. P. R. contributed to collection and assembly of data.

Financial/nonfinancial disclosures: None declared.

Role of sponsors: The sponsor had no role in the design of the study, the collection and analysis of the data, or the preparation of the manuscript.

Additional information: The e-Appendixes, e-Figures, and e-Tables can be found in the Supplemental Materials section of the online article.

References

1. Hall MJ, Williams SN, DeFrances CJ, Golosinskiy A. Inpatient care for septicemia or sepsis: a challenge for patients and hospitals. *NCHS Data Brief*. 2011;(62):1-8.
2. Liu V, Escobar GJ, Greene JD, et al. Hospital deaths in patients with sepsis from 2 independent cohorts. *JAMA*. 2014;312(1):90-92.
3. Gaieski DF, Edwards JM, Kallan MJ, Carr BG. Benchmarking the incidence and mortality of severe sepsis in the United States. *Crit Care Med*. 2013;41(5):1167-1174.
4. Torio C (AHRQ), Moore B (Truven Health Analytics). National inpatient hospital costs: the most expensive conditions by payer, 2013. HCUP Statistical Brief #204. Agency for Healthcare Research and Quality; 2016.
5. CSPAN. Senate Health Committee: father of Rory Staunton, who died of sepsis. Accessed November 8, 2019, https://www.c-span.org/video/?c4465991/senate-health-committee-father-rory-staunton-died-sepsis.
6. National Quality Forum. MAP 2015 considerations for implementing measures in federal programs. Draft for public comment. Accessed November 8, 2019, https://www.qualityforum.org/Setting_Priorities/Partnership/Draft_Programmatic_Deliverable_Report.aspx.
7. Moreira H, Sinert R. How effective is the early management bundle for severe sepsis/septic shock? *JAMA Intern Med*. 2020;180(5):716-717.
8. Department of Health and Human Services. Chart-abstracted data validation. Accessed November 8, 2019, https://qualitynet.org/inpatient/data-management/chart-abstracted-data-validation.
9. Seymour CW, Liu VX, Iwashyna TJ, et al. Assessment of clinical criteria for sepsis: for the Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3) [published correction appears in *JAMA*. 2016;315(20):2237]. *JAMA*. 2016;315(8):762-774.
10. Greenland S. Interpretation and choice of effect measures in epidemiologic analyses. *Am J Epidemiol*. 1987;125(5):761-768.
11. Imbens G. Nonparametric estimation of average treatment effects under exogeneity: a review. *Review of Economics and Statistics*. 2004;86(1):4-29.
12. Darby JL, Davis BS, Barbash IJ, Kahn JM. An administrative model for benchmarking hospitals on their 30-day sepsis mortality. *BMC Health Serv Res*. 2019;19(1):221.
13. Lederer DJ, Bell SC, Branson RD, et al. Control of confounding and reporting of results in causal inference studies. Guidance for authors from editors of respiratory, sleep, and critical care journals [published correction appears in *Ann Am Thorac Soc*. 2019;16(2):283]. *Ann Am Thorac Soc*. 2019;16(1):22-28.
14. Biondi-Zoccai G, Romagnoli E, Agostoni P, et al. Are propensity scores really superior to standard multivariable analysis? *Contemp Clin Trials*. 2011;32(5):731-740.
15. Austin PC, Grootendorst P, Anderson GM. A comparison of the ability of different propensity score models to balance measured variables between treated and untreated subjects: a Monte Carlo study. *Stat Med*. 2007;26(4):734-753.
16. Brookhart MA, Schneeweiss S, Rothman KJ, Glynn RJ, Avorn J, Stürmer T. Variable selection for propensity score models. *Am J Epidemiol*. 2006;163(12):1149-1156.
17. Elixhauser A, Steiner C, Harris DR, Coffey RM. Comorbidity measures for use with administrative data. *Med Care*. 1998;36(1):8-27.
18. Quan H, Sundararajan V, Halfon P, et al. Coding algorithms for defining comorbidities in ICD-9-CM and ICD-10 administrative data. *Med Care*. 2005;43(11):1130-1139.
19. Elias KM, Moromizato T, Gibbons FK, Christopher KB. Derivation and validation of the acute organ failure score to predict outcome in critically ill patients: a cohort study. *Crit Care Med*. 2015;43(4):856-864.
20. Iwashyna TJ, Odden A, Rohde J, et al. Identifying patients with severe sepsis using administrative claims: patient-level validation of the angus implementation of the international consensus conference definition of severe sepsis. *Med Care*. 2014;52(6):e39-e43.
21. Arpino B, Mealli F. The specification of the propensity score in multilevel observational studies. *Computational Statistics & Data Analysis*. 2011;55(4):1770-1780.
22. Arpino B, Cannas M. Propensity score matching with clustered data. An application to the estimation of the impact of caesarean section on the Apgar score. *Stat Med*. 2016;35(12):2074-2091.
23. Ene M, Leighton EA, Blue GL, Bell BA. Multilevel models for categorical data using SAS® PROC GLIMMIX: the basics. Accessed April 5, 2020. https://support.sas.com/resources/papers/proceedings15/3430-2015.pdf.
24. Kiernan K. Insights into using the GLIMMIX procedure to model categorical outcomes with random effects. Accessed April 5, 2020. https://www.sas.com/content/dam/SAS/support/en/sas-global-forum-proceedings/2018/2179-2018.pdf.
25. Rubin DB, Thomas N. Combining propensity score matching with additional adjustments for prognostic covariates. *J Am Stat Assoc*. 2000;95(450):573-585.
26. Rubin DB. Using propensity scores to help design observational studies: application to the tobacco litigation. *Health Services & Outcomes Research Methodology*. 2001;2:169-188.
27. Stuart EA. Matching methods for causal inference: a review and a look forward. *Stat Sci*. 2010;25(1):1-21.
28. Austin PC. Balance diagnostics for comparing the distribution of baseline covariates between treatment groups in propensity-score matched samples. *Stat Med*. 2009;28(25):3083-3107.
29. Austin PC, Grootendorst P, Normand SL, Anderson GM. Conditioning on the propensity score can result in biased estimation of common measures of treatment effect: a Monte Carlo study. *Stat Med*. 2007;26(4):754-768.
30. Austin PC. An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate Behav Res*. 2011;46(3):399-424.
31. Austin PC. The performance of different propensity score methods for estimating marginal odds ratios. *Stat Med*. 2007;26(16):3078-3094.
32. Capanu M, Gönen M, Begg CB. An assessment of estimation methods for generalized linear mixed models with binary outcomes. *Stat Med*. 2013;32(26):4550-4566.
33. Austin PC. The performance of different propensity score methods for estimating marginal hazard ratios. *Stat Med*. 2013;32(16):2837-2849.
34. Normand ST, Landrum MB, Guadagnoli E, et al. Validating recommendations for coronary angiography following acute myocardial

infarction in the elderly: a matched analysis using propensity scores. *J Clin Epidemiol*. 2001;54(4):387-398.

35. Rosenbaum PR, Rubin DB. Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *Am Stat*. 1985;39(1):33-38.
36. Austin PC. Optimal caliper widths for propensity-score matching when estimating differences in means and differences in proportions in observational studies. *Pharm Stat*. 2011;10(2):150-161.
37. Rosenbaum PR, Rubin DB. The bias due to incomplete matching. *Biometrics*. 1985;41(1):103-116.
38. Levy MM, Gesten FC, Phillips GS, et al. Mortality changes associated with mandated public reporting for sepsis. The results of the New York State initiative. *Am J Respir Crit Care Med*. 2018;198(11):1406-1412.
39. Seymour CW, Gesten F, Prescott HC, et al. Time to treatment and mortality during mandated emergency care for sepsis. *N Engl J Med*. 2017;376(23):2235-2244.
40. Leisman DE, Doerfler ME, Ward MF, et al. Survival benefit and cost savings from compliance with a simplified 3-hour sepsis bundle in a series of prospective, multisite, observational cohorts. *Crit Care Med*. 2017;45(3):395-406.
41. Liu VX, Morehouse JW, Marelich GP, et al. Multicenter implementation of a treatment bundle for patients with sepsis and intermediate lactate values. *Am J Respir Crit Care Med*. 2016;193(11):1264-1270.
42. Beck V, Chateau D, Bryson GL, et al. Timing of vasopressor initiation and mortality in septic shock: a cohort study. *Crit Care*. 2014;18(3):R97.
43. Larzelere RE, Kuhn BR, Johnson B. The intervention selection bias: an underrecognized confound in intervention research. *Psychol Bull*. 2004;130(2):289-303.
44. Thompson BT, Schoenfeld D. Usual care as the control group in clinical trials of nonpharmacologic interventions. *Proc Am Thorac Soc*. 2007;4(7):577-582.
45. Department of Health and Human Services. November 14 sepsis efforts at Bellevue Hospital and SEP-1 early management bundle, severe sepsis/septic shock: v5.0b through v5.2a analysis result. Accessed April 21, 2021, https://qualitynet.cms.gov/files/5d379ab34b6d1a256059f2df?filename=IQR_Webinar_11142017.pdf.
46. Barbash IJ, Davis B, Kahn JM. National performance on the Medicare SEP-1 sepsis quality measure. *Crit Care Med*. 2019;47(8):1026-1032.
47. Bauer SR, Han X, Wang XF, Blonsky H, Reddy AJ. Association between compliance with the sepsis quality measure (SEP-1) and hospital readmission. *Chest*. 2020;158(2):608-611.
48. Pepper DJ, Sun J, Cui X, Welsh J, Natanson C, Eichacker PQ. Antibiotic- and fluid-focused bundles potentially improve sepsis management, but high-quality evidence is lacking for the specificity required in the Centers for Medicare and Medicaid Service's sepsis bundle (SEP-1). *Crit Care Med*. 2019;47(10):1290-1300.
49. Baghdadi JD, Brook RH, Uslan DZ, et al. Association of a care bundle for early sepsis management with mortality among patients with hospital-onset or community-onset sepsis. *JAMA Intern Med*. 2020;180(5):707-716.
50. Rhee C, Filbin MR, Massaro AF, et al. Compliance with the national SEP-1 quality measure and association with sepsis outcomes: a multicenter retrospective cohort study. *Crit Care Med*. 2018;46(10):1585-1591.
51. Filbin MR, Lynch J, Gillingham TD, et al. Presenting symptoms independently predict mortality in septic shock: importance of a previously unmeasured confounder. *Crit Care Med*. 2018;46(10):1592-1599.
52. Department of Health and Human Services. Hospital inpatient specifications manuals; version 5.8 - specifications manual for discharges 07/01/20 - 12/31/20. Accessed May 28, 2020, https://www.qualitynet.org/inpatient/specifications-manuals.
53. Barbash IJ, Davis BS, Yabes JG, Seymour CW, Angus DC, Kahn JM. Treatment patterns and clinical outcomes after the introduction of the Medicare sepsis performance measure (SEP-1). *Ann Intern Med*. 2021;174(7):927-935.
54. Buchman TG, Simpson SQ, Sciarretta KL, et al. Sepsis among Medicare beneficiaries: 1. The burdens of sepsis, 2012-2018. *Crit Care Med*. 2020;48(3):276-288.
55. Bauer SR, Gonet JA, Rosario RF, Griffiths LA, Kingery T, Reddy AJ. Inter-rater agreement for abstraction of the early management bundle, severe sepsis/septic shock (SEP-1) quality measure in a multi-hospital health system. *Jt Comm J Qual Patient Saf*. 2019;45(2):108-111.
56. Aaronson EL, Filbin MR, Brown DF, Tobin K, Mort EA. New mandated Centers for Medicare and Medicaid Services requirements for sepsis reporting: caution from the field. *J Emerg Med*. 2017;52(1):109-116.
57. Rhee C, Brown SR, Jones TM, et al. Variability in determining sepsis time zero and bundle compliance rates for the centers for Medicare and Medicaid services SEP-1 measure. *Infect Control Hosp Epidemiol*. 2018;39(8):994-996.
58. Baggs J, Kazakova S, Hatfield KM, et al. 2891. Trends in inpatient antibiotic use in US Hospitals, 2012–2017. *Open Forum Infect Dis*. 2019;6(suppl 2):S79.
59. Jernigan JA, Hatfield KM, Wolford H, et al. Multidrug-resistant bacterial infections in U.S. hospitalized patients, 2012-2017. *N Engl J Med*. 2020;382(14):1309-1319.
60. Guh AY, Mu Y, Winston LG, et al. Trends in U.S. burden of *Clostridioides difficile* infection and outcomes. *N Engl J Med*. 2020;382(14):1320-1330.