Supplementary Online Content

Gupta S, Wang W, Hayek SS, et al; for the STOP-COVID Investigators. Association Between Early Treatment With Tocilizumab and Mortality Among Critically Ill Patients With COVID-19. Published October 20, 2020. *JAMA Intern Med.* doi:10.1001/jamainternmed.2020.6252

eMethods. Data Collection and Validation, Analyses and Covariates

eTable 1. List of Participating Sites

eTable 2. Definitions of Baseline Characteristics, Comorbidities, and Adverse Events

eTable 3. Corticosteroids Administered on ICU Admission

eTable 4. Causes of Death in Tocilizumab-Treated and Non–Tocilizumab-Treated Patients

eTable 5. Adverse Events in Tocilizumab-Treated and Non–Tocilizumab-Treated Patients Within the First 14 Days After ICU Admission

eFigure 1. Distribution of Tocilizumab Receipt According to the Number of Days After ICU Admission

eFigure 2. Standardized Differences Before and After Applying Inverse Probability Weighting

References.

This supplementary material has been provided by the authors to give readers additional information about their work.

eMethods. Data Collection and Validation, Analyses and Covariates

Data Collection and Validation
Data were collected using REDCap, a secure, HIPAA-compliant, web-based application. Wherever possible, data were captured using checkboxes rather manual entry to minimize keystroke errors. For data that required keystroke entry (e.g., laboratory values), we implemented validation ranges to flag potential errors in real-time. We also implemented automated data validation rules to flag errors in dates (e.g., if the date of death was entered as being before the date of ICU admission). In addition, all data were manually reviewed, and values that appeared incongruent or out of range were manually validated by confirming the accuracy of the data with the collaborator who entered it.

Target trial specification
We sought to emulate a hypothetical target trial in which eligible patients are adults (≥18 years old) with laboratory-confirmed COVID-19 who were admitted to a participating intensive care unit (ICU) between March 4 and May 10, 2020. Exclusion criteria were as follows:

1. Hospitalization for one week or longer prior to ICU admission.
2. Receipt of tocilizumab prior to ICU admission.
3. Enrollment in a placebo-controlled clinical trial involving tocilizumab or other inhibitors of interleukin-6 or the interleukin-6 receptor.
4. Receipt of an interleukin-6/interleukin-6 receptor antagonist other than tocilizumab during the first two days of ICU admission.
5. Liver dysfunction that precluded the ability to receive tocilizumab. Specifically, we excluded patients with an aspartate aminotransferase or alanine aminotransferase level greater than 500 U/L on ICU admission. Similar exclusion criteria for liver dysfunction were used in clinical trials involving tocilizumab (e.g., ClinicalTrials.gov NCT04320615).

In order to meet eligibility criteria, patients had to be admitted to the ICU for illness directly attributable to COVID-19. Patients were categorized according to receipt or no receipt of tocilizumab in the first two days of ICU admission. ICU day 1 was defined as the 24-hour period spanning from midnight to midnight on the day of ICU admission. Day 2 was defined as the subsequent 24-hour period following ICU day 1. Patients who received tocilizumab after the first two days of ICU admission were categorized in the tocilizumab non-treated group. Patients were followed until the first of hospital discharge, death, or June 12, 2020 – the date on which the study database for the current analysis was locked. All patients who remained hospitalized at last follow-up had a minimum of 28 days of follow-up.

The primary analysis compared the time to death among patients who received tocilizumab within two days following ICU admission to those who did not. Time to death was defined as the interval from ICU admission to death, censored at hospital discharge or the end of follow-up, whichever occurred first. Hazard ratios and 95% confidence intervals were estimated using a Cox model.

Inverse probability weighting
To adjust for confounding we fit a logistic regression model with tocilizumab receipt as the outcome conditional on the covariates listed below. These covariates were selected based on clinical judgment, as they were thought to be potentially associated with a clinician's decision to initiate treatment with tocilizumab and with mortality. We used these predicted probabilities to calculate stabilized inverse probability weights.[1] For tocilizumab-treated patients, stabilized weights were generated by taking the reciprocal of the estimated probability of treatment and

multiplying it by unadjusted overall probability of treatment. For tocilizumab non-treated patients, we took the reciprocal of the probability of no treatment and multiplied it by the unadjusted probability of no treatment. We used a robust (sandwich) variance estimator to account for potential replications of patients induced by inverse probability weighting, which results in conservative (wider) 95% CIs. We evaluated standardized differences across each measured covariate before and after applying the weighting (**eFigure 2**).

Sensitivity analyses

We conducted four prespecified sensitivity analyses and one post hoc analysis. First, we kept discharged patients in the risk set until June 12, 2020, the date of last follow-up, since Cox models assume non-informative censoring. Second, we included the covariates below in a traditional unweighted Cox model. Third, to eliminate the potential for immortal time bias, we performed a nested target trial emulation analysis[2,3] in which we categorized eligible patients as having received tocilizumab or not on ICU day 1, and we repeated the process for eligible patients on ICU day 2. Our final estimates were obtained by pooling the data from the emulation of the nested target trials on ICU days 1 and 2, using inverse probability weighting as described above. Patients receiving treatment only appeared in the pooled dataset up to and including the day that treatment was initiated. For example, a patient who received tocilizumab on ICU day 1 did not have a corresponding observation on ICU day 2. A patient who received tocilizumab on ICU day 2, meanwhile, appeared as both a tocilizumab non-treated patient on ICU day 1 and as a tocilizumab-treated patient on ICU day 2. Fourth, we excluded patients who had any of the following critical values or events on the day of ICU admission, since such patients may not have received tocilizumab due to having been deemed too ill to benefit from it: arterial pH < 7.0; arterial lactate >10 mmol/L; receipt of 4 or more vasopressors; or cardiac arrest. Fifth, in a post hoc analysis we repeated the primary analysis and included the number of pre-COVID ICU beds (<50, 50-99, or ≥100) at each site in the model, as we previously found this variable to be an important predictor of death in critically ill patients with COVID-19.[4]

Subgroup analyses

We used similar methods as the primary analysis described above to assess the effect of tocilizumab on time to death across the following prespecified subgroups: age (<60 versus ≥60 years); sex; days from symptom onset to ICU admission (≤3 versus >3); degree of hypoxemia on ICU admission (mechanically ventilated with a PaO2:FiO2 ratio <200 versus mechanically ventilated with a PaO2:FiO2 ratio ≥200 or not mechanically ventilated); vasopressors received on ICU admission (≥1 versus 0); and receipt of corticosteroids on ICU admission (yes/no). We compared differences among subgroups by adding product (“interaction”) terms between the subgroup variable and the tocilizumab group into the outcome model. Because of multiple comparisons, findings for subgroup analyses should be interpreted as exploratory. Analyses were performed using SAS software version 9.4 (SAS Institute).

List of model covariates

A. Baseline covariates

1. Age: 18-49; 50-59; 60-69; ≥70
2. Male sex
3. Race: White versus Non-White/Other/Unknown
4. Ethnicity: Non-Hispanic versus Hispanic/Unknown
5. Body mass index (kg/m^2): <25; 25-29; 30-34; 35-39.9; ≥40; missing
6. Hypertension
7. Diabetes mellitus
8. Coronary artery disease
9. Congestive heart failure
10. Current smoker
11. Active malignancy
12. Home medications (each included separately and included as a binary variable): angiotensin converting enzyme inhibitor; angiotensin II receptor blocker; statin
13. Days from symptom onset to ICU admission: 0 to 3; >3

B. Severity-of-illness covariates assessed on ICU admission

1. Renal and liver components of the Sequential Organ Failure Assessment (SOFA) score:[5]

	Categories		
	0	1	2[a]
SOFA Renal (Cr, UOP, RRT, and ESRD)	Cr<1.2 mg/dl	Cr 1.2-1.9	Cr ≥2 or UOP<500[b] or acute RRT or ESRD
SOFA Liver (Bilirubin, mg/dl)	<1.2	1.2-1.9	≥2

Abbreviations: Cr, creatinine (mg/dl); ESRD, end stage renal disease; RRT, renal replacement therapy; UOP, urine output.
[a]The renal and liver components of the SOFA score were binned due to low frequency of events in categories "3" and "4".
[b]If the UOP was missing, the category was assigned according to the Cr

2. PaO_2:FiO_2 ratio as follows: not ventilated; ventilated and PaO_2:FiO_2 ratio ≥200; ventilated and PaO_2:FiO_2 ratio <200; ventilated and PaO_2:FiO_2 ratio missing
3. Vasopressors: 0, 1, or ≥2
4. Fever (>38ºC)
5. White blood cell count (per mm^3): <4000, 4000–11,900, ≥12,000; missing
6. Inflammation. Three mutually exclusive categories were created: inflamed, non-inflamed, or missing. Inflamed was defined as at least one of the following on ICU days 1 or 2: C-reactive protein >100 mg/L, interleukin-6 >80 pg/ml, or ferritin >1,000 ng/mL. Non-inflamed was defined as at least one value that was below the threshold and no value that was above the threshold for the above parameters. Missing was defined as all three values being missing. The above thresholds were chosen based on prior studies.[6-8]
7. Concurrent therapies (each assessed individually): hydroxychloroquine; azithromycin; corticosteroids; therapeutic anticoagulation; prone positioning; neuromuscular blockade

Missing Data

The renal and liver components of the SOFA score were categorized as "0" if missing.[9-11] Otherwise, missing data were not imputed. Rather, we created a separate missing category for each covariate that had missing data, since data may not have been missing at random. Further, the missingness of a variable could have clinical relevance (e.g., a healthier patient may not have certain physiologic or laboratory values assessed as frequently), which could affect treatment decisions.

eTable 1. List of participating sites

Northeast
Beth Israel Deaconess Medical Center
Brigham and Women's Faulkner Hospital
Brigham and Women's Hospital
Cooper University Health Care
Hackensack Meridian Health Hackensack University Medical Center
Hackensack Mountainside Hospital
Johns Hopkins Hospital
Kings County Hospital Center
Lowell General Hospital
Massachusetts General Hospital
MedStar Georgetown University Hospital
Montefiore Medical Center
Mount Sinai
Newton Wellesley Hospital
New York-Presbyterian Queens Hospital
New York-Presbyterian/Weill Cornell Medical Center
New York University Langone Hospital
Rutgers/New Jersey Medical School
Rutgers/Robert Wood Johnson Medical School
Temple University Hospital
Jefferson Health
Tufts Medical Center
United Health Services Hospitals
University of Pennsylvania Health System
University of Pittsburgh Medical Center
Westchester Medical Center
Yale University Medical Center
South
Baylor College of Medicine, Houston
Baylor University Medical Center/Baylor Scott White and Health
Duke University Medical Center
Mayo Clinic, Florida
Memphis VA Medical Center
Methodist University Hospital
Ochsner Medical Center
Tulane Medical Center
University of Alabama-Birmingham Hospital
University of Florida Health-Gainesville
University of Florida Health-Jacksonville
University of Miami Health System
University of North Carolina Hospitals
University of Texas Southwestern Medical Center
University of Virginia Health System
Midwest
Barnes-Jewish Hospital
Cook County Health
Froedtert Hospital
Indiana University Health Methodist Hospital
Mayo Clinic, Rochester
Northwestern Memorial Hospital
Promedica Health System
Rush University Medical Center
University Hospitals Cleveland Medical Center
University of Chicago Medical Center
University of Illinois Hospital and Health Sciences System
University of Kentucky Hospital
University of Michigan Hospital
University of Oklahoma Health Sciences Center
West
Loma Linda University Medical Center
Mayo Clinic, Arizona
Oregon Health and Science University Hospital
Renown Health
Stanford Healthcare
University of California-Davis Medical Center
University of California-Los Angeles Medical Center
University of California-San Diego Medical Center
University of California-San Francisco Medical Center
UCHealth University of Colorado
University Medical Center of Southern Nevada
University of Washington Medical Center

eTable 2. Definitions of baseline characteristics, comorbidities, and adverse events

Baseline Characteristics	
Home medications	Medications that the patient was taking at home within 1 week prior to admission. Does not include those started at an outside hospital if the patient was transferred.
Coexisting Conditions	
Active malignancy	Per chart review; active malignancy (other than non-melanoma skin cancer) treated in the past year. Defined as cancer of the lung, breast, colorectal, prostate, gastric, pancreatic, melanoma, ovarian, brain, or other
Congestive heart failure	Per chart review; heart failure with preserved versus reduced ejection fraction
Coronary artery disease	Per chart review; any history of angina, myocardial infarction, or coronary artery bypass graft surgery
Diabetes mellitus	Per chart review; insulin or non-insulin dependent
Hypertension	Per chart review
Smoking	Per chart review; does not include vaping or smoking of non-tobacco products. Current smoker versus non-smoker/former smoker
Adverse Events	
Secondary infection	Per chart review; suspected or confirmed new infection other than COVID-19 that developed after admission to the ICU. Pneumonia (including ventilator-associated), urosepsis, biliary sepsis, bacteremia, other
Arrhythmia	Per chart review; includes atrial fibrillation/flutter, ventricular tachycardia (sustained versus non-sustained), and ventricular fibrillation
Thrombotic event	Per chart review; deep venous thrombosis, pulmonary embolism, stroke, other

eTable 3. Corticosteroids administered on ICU admission

	Tocilizumab (N=433)	No Tocilizumab (N=3491)
Treated with any Steroid (%)	81 (18.7)	440 (12.6)
Methylprednisolone Equivalent Daily Dose (median, IQR)	64 (40–120)	80 (40–120)
Betamethasone		
No. (%)	1 (1.2)	0 (0.0)
Median Daily Dose, mg (IQR)	12 (12–12)	N/A
Dexamethasone		
No. (%)	12 (14.8)	28 (6.4)
Median Daily Dose, mg (IQR)	14 (10–20)	10 (10–20)
Fludrocortisone (%)		
No. (%)	0 (0.0)	1 (0.2)
Median Daily Dose, mg (IQR)	N/A	0.2 (0.2–0.2)
Hydrocortisone		
No. (%)	11 (13.6)	60 (13.6)
Median Daily Dose, mg (IQR)	200 (90–275)	200 (100–200)
Methylprednisolone		
No. (%)	48 (59.3)	288 (65.5)
Median Daily Dose, mg (IQR)	120 (60–180)	120 (8–125)
Prednisone		
No. (%)	9 (11.1)	63 (14.3)
Median Daily Dose, mg (IQR)	10 (5–30)	20 (6–40)

eTable 4. Causes of death in tocilizumab-treated and tocilizumab non-treated patients

Cause of Death	Tocilizumab (N=125)	No Tocilizumab (N=1419)
ARDS/respiratory failure – no. (%)	113 (90.4)	1270 (89.5)
Congestive heart failure – no. (%)	14 (11.2)	135 (9.5)
Septic shock – no. (%)	52 (41.6)	553 (39.0)
Acute kidney injury – no. (%)	46 (36.8)	476 (33.5)
Liver failure – no. (%)	5 (4.0)	67 (4.7)
Other causes – no. (%)	16 (12.8)	196 (13.8)

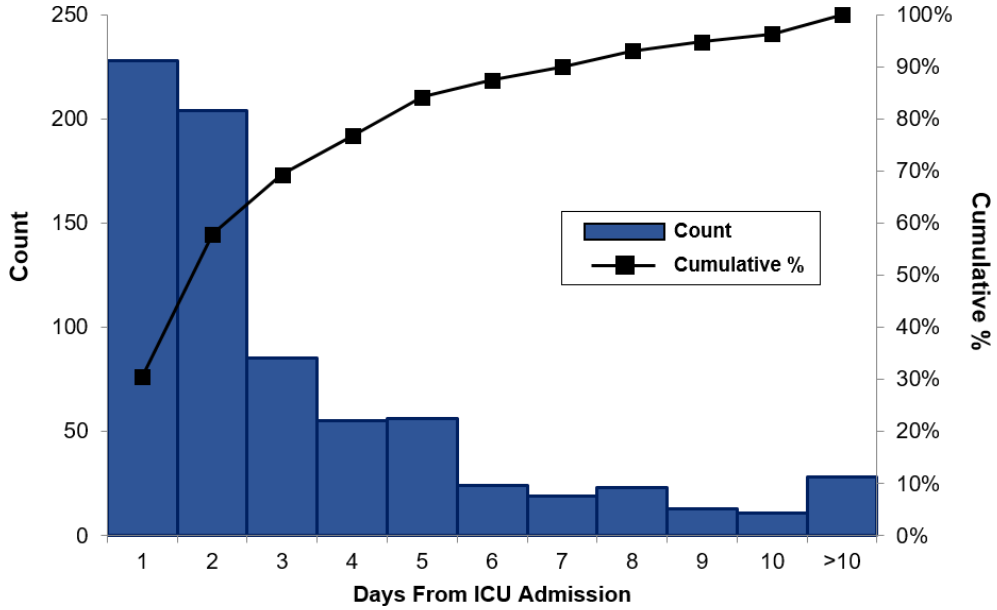
Note: Patients could have had more than 1 cause of death
Abbreviations: ARDS, acute respiratory distress syndrome

eTable 5. Adverse events in tocilizumab-treated and tocilizumab non-treated patients within the first 14 days following ICU admission

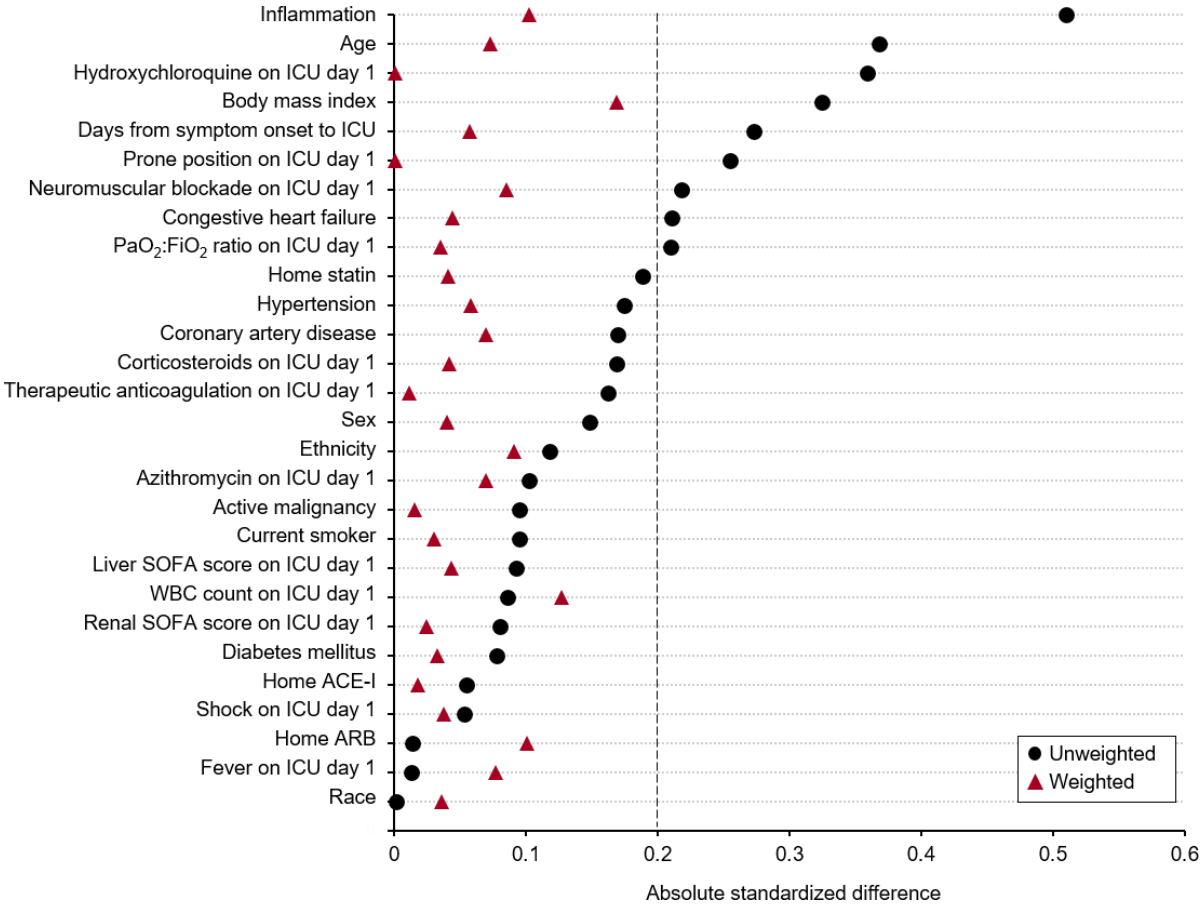
Adverse Events	Tocilizumab (N=433)	No Tocilizumab (N=3491)
Secondary Infections – no. (%)[a]		
Any	140 (32.3)	1085 (31.1)
Pneumonia	112 (25.9)	732 (21.0)
Urosepsis	8 (1.9)	111 (3.2)
Biliary sepsis	0 (0)	5 (0.1)
Bacteremia	29 (6.7)	285 (8.2)
Other	17 (3.9)	135 (3.9)
Liver Dysfunction – no. (%)		
AST or ALT >250 U/L	72 (16.6)	452 (12.9)
AST or ALT >500 U/L	37 (8.5)	196 (5.6)
Arrhythmias – no. (%)		
Any	63 (14.5)	602 (17.2)
Atrial fibrillation or atrial flutter	50 (9.0)	507 (14.5)
Ventricular tachycardia	14 (3.2)	102 (2.9)
Ventricular fibrillation	1 (0.2)	20 (0.6)
Thrombotic Complications – no. (%)		
Any	46 (10.6)	342 (9.8)
Deep vein thrombosis	25 (5.8)	190 (5.4)
Pulmonary embolism	22 (5.1)	83 (2.4)
Stroke	1 (0.2)	36 (1.0)
Other	4 (0.9)	52 (1.5)

Note: Some patients had more than one secondary infection, arrhythmia, and/or thrombotic event.
[a]Data on secondary infection were missing in 19 out of 3924 patients (0.5%).
Abbreviations: ALT, alanine aminotransferase; AST, aspartate aminotransferase.

eFigure 1. Distribution of tocilizumab receipt according to the number of days following ICU admission.



eFigure 2. Standardized differences before and after applying inverse probability weighting. This figure shows the absolute standardized differences between tocilizumab-treated and non-treated patients for each of the 28 baseline covariates in the unweighted sample and after applying the weights derived from the inverse probability weighting. The vertical dashed line denotes a standardized difference of 0.2, as effects sizes below 0.2 are considered to be small,[12] and effects sizes below 0.1 are considered to be very small.[13] The standardized differences in the unweighted sample exceeded 0.2 for 9 of the 28 covariates (32.1%). In contrast, none of the standardized differences in the weighted sample exceeded 0.2, and only 2 of the 28 covariates (7.1%) exceeded 0.1. This indicates that the baseline characteristics were well balanced between tocilizumab-treated and non-treated patients after applying the weighting.



eReferences

1. Xu S, Ross C, Raebel MA, Shetterly S, Blanchette C, Smith D. Use of stabilized inverse propensity scores as weights to directly estimate relative risk and its confidence intervals. Value Health 2010;13:273-7.
2. Hernan MA, Robins JM. Using Big Data to Emulate a Target Trial When a Randomized Trial Is Not Available. Am J Epidemiol 2016;183:758-64.
3. Danaei G, Garcia Rodriguez LA, Cantero OF, Logan RW, Hernan MA. Electronic medical records can be used to emulate target trials of sustained treatment strategies. J Clin Epidemiol 2018;96:12-22.
4. Gupta S, Hayek SS, Wang W, et al. Factors Associated With Death in Critically Ill Patients With Coronavirus Disease 2019 in the US. JAMA Intern Med 2020.
5. Vincent JL, Moreno R, Takala J, et al. The SOFA (Sepsis-related Organ Failure Assessment) score to describe organ dysfunction/failure. On behalf of the Working Group on Sepsis-Related Problems of the European Society of Intensive Care Medicine. Intensive Care Med 1996;22:707-10.
6. Herold T, Jurinovic V, Arnreich C, et al. Elevated levels of IL-6 and CRP predict the need for mechanical ventilation in COVID-19. J Allergy Clin Immunol 2020.
7. Landry A, Docherty P, Ouellette S, Cartier LJ. Causes and outcomes of markedly elevated C-reactive protein levels. Can Fam Physician 2017;63:e316-e23.
8. Moore C, Jr., Ormseth M, Fuchs H. Causes and significance of markedly elevated serum ferritin levels in an academic medical center. J Clin Rheumatol 2013;19:324-8.
9. Raith EP, Udy AA, Bailey M, et al. Prognostic Accuracy of the SOFA Score, SIRS Criteria, and qSOFA Score for In-Hospital Mortality Among Adults With Suspected Infection Admitted to the Intensive Care Unit. JAMA 2017;317:290-300.
10. Jentzer JC, Bennett C, Wiley BM, et al. Predictive Value of the Sequential Organ Failure Assessment Score for Mortality in a Contemporary Cardiac Intensive Care Unit Population. J Am Heart Assoc 2018;7.
11. Aakre C, Franco PM, Ferreyra M, Kitson J, Li M, Herasevich V. Prospective validation of a near real-time EHR-integrated automated SOFA score calculator. Int J Med Inform 2017;103:1-6.
12. Cohen J. Statistical Power Analysis for the Behavioral Sciences. 2nd ed. Hillsdale, NJ; 1988.
13. Austin PC, Stuart EA. Moving towards best practice when using inverse probability of treatment weighting (IPTW) using the propensity score to estimate causal treatment effects in observational studies. Stat Med 2015;34:3661-79.