Assessment of the Natural History of Coagulopathy in COVID-19

Statistical Analysis Plan

Vincent Lo Re III, MD, MSCE,[1] Sarah Dutcher PhD, MS,[3] Silvia Perez-Vilar, PhD,[3] Brian Kitt, MD, MPH,[3] John Connolly, ScM, ScD,[2] Noelle Cocoros, PhD, MPH[2]

Affiliations: 1. Division of Infectious Diseases, Department of Medicine and Center for Clinical Epidemiology and Biostatistics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA. 2. Department of Population Medicine, Harvard Pilgrim Health Care Institute and Harvard Medical School, Boston, MA; 3. Office of Surveillance and Epidemiology, Center for Drug Evaluation and Research, US Food and Drug Administration, Silver Spring, MD

Version 1.0

March 15, 2021

The Sentinel System is sponsored by the U.S. Food and Drug Administration (FDA) to proactively monitor the safety of FDA-regulated medical products and complements other existing FDA safety surveillance capabilities. The Sentinel System is one piece of FDA's Sentinel Initiative, a long-term, multi-faceted effort to develop a national electronic system. Sentinel Collaborators include Data and Academic Partners that provide access to healthcare data and ongoing scientific, technical, methodological, and organizational expertise. The Sentinel Operations Center is funded by the FDA through the Department of Health and Human Services (HHS) Contract number HHSF223201400030I.

Coagulopathy Statistical Analysis Plan

Table of Contents

History of Modifications

Version	Date	Modification	Author
1.0	03/15/2021	Original Version	Sentinel Operations Center

1 Statistical Analysis Plan

1.1 Descriptive Statistics

We will describe and compare the baseline characteristics of health plan members with COVID-19 to those with influenza virus infection. Group differences will be assessed by using standardized mean differences for continuous variables and standardized difference in proportion for categorical variables, using a threshold of ≥0.10 to suggest meaningful imbalance.

The primary descriptive table will consist of the demographic characteristics, comorbidities, laboratory tests, and outpatient medications of health plan members from the participating Data Partners with COVID-19 or influenza that have been identified either by: 1) outpatient, ambulatory visit, emergency department, institutional stay, or inpatient International Classification of Diseases, Tenth Revision, Clinical Modification [ICD-10-CM] diagnosis code (in any position) of these infections, or 2) a positive nucleic acid test for these infections recorded within any care setting.

Secondary descriptive tables will also be created to characterize health plan members identified using modified cohort definitions. First, we will stratify the primary cohort according to setting at diagnosis (i.e., hospitalization [inpatient or institutional stay encounter] versus ambulatory [outpatient, ambulatory visit, or emergency department encounter]). Second, a separate table will compare characteristics between members with COVID-19 or influenza, restricted to those who were diagnosed by a positive nucleic acid test.

At least one of the integrated delivery system Data Partners will evaluate select inpatient medications used for the treatment of COVID-19 (e.g., remdesivir, dexamethasone) among hospitalized patients and separately examine treatments for arterial or venous thrombotic events (e.g., anticoagulants, thrombolytics), stratified by receipt of anticoagulant/anti-platelet therapy during the baseline period.

1.2 Incidence and Consequences of Thrombotic Events in COVID-19 and Influenza

To address Aim 1, we will calculate the absolute risk and unadjusted incidence rates of **primary** arterial and venous thrombotic endpoints over the 90 days following and including the diagnosis date in the COVID-19 and influenza cohorts. The primary endpoints of arterial thrombosis (acute myocardial infarction [AMI], acute stroke) and venous thromboembolism (acute deep venous thrombosis [DVT], acute pulmonary embolism [PE]) will be identified in the **inpatient** setting via ICD-10 diagnosis codes. Among health plan members with COVID-19 or influenza that have been identified either by outpatient, ambulatory visit, emergency department, institutional stay, or inpatient ICD-10-CM diagnosis code (in any position) or a positive nucleic acid test for these infections recorded within any care setting, results will be determined overall and stratified by:

- Age group (18-44; 45-54; 55-64; 65-74; 75-84; ≥85 years, which were age groups evaluated by the US Centers for Disease Control and Prevention for evaluation of rates of severe COVID-19)
- Sex
- Race, Hispanic ethnicity
- Severity of infection at diagnosis (not hospitalized; hospitalized but never received intensive care or mechanical ventilation; hospitalized and required either intensive care or mechanical ventilation)
- Highest severity of infection up to 14 days after diagnosis

- Care setting of index diagnosis (hospital or institutional stay versus outpatient/ambulatory/emergency department versus unknown)
- Recent institutional stay encounter (days -90 through -1),
- Baseline history of cardiovascular disease or venous thromboembolism (further stratified by use of anticoagulant drug exposure within -183 to -3 days prior to index diagnosis)
- Baseline dispensed anticoagulant drug (-183 to -3 days prior to index diagnosis)
- Baseline dispensed antiplatelet drug (-183 to -3 days prior to index diagnosis)
- Polycythemia, as determined by hemoglobin >16 g/dL (based on the cut-off recommended by the World Health Organization [Arber et al., 2016]) or a recorded polycythemia diagnosis
- Thrombocytosis, as determined by platelet count >450 x 10^9/L (based on the cut-off recommended by the World Health Organization [Arber et al., 2016]) or a recorded thrombocytosis diagnosis
- Neurologic disease that promotes stasis/immobility
- Month of index diagnosis (April/May; June/July; August/September; October/November [COVID-19 cohort only])

We will then repeat the above analyses for the cohorts stratified by care setting (inpatient only; ambulatory/outpatient/emergency department only) and restricted to members diagnosed by a positive nucleic acid test.

Among health plan members diagnosed with COVID-19 or influenza by either ICD-10-CM diagnosis code or positive nucleic acid test, we will calculate the absolute risk and unadjusted incidence rates of the following secondary endpoints in each cohort:

1. Emergency department, institutional stay, or hospital discharge ICD-10-CM diagnosis for an **expanded** arterial thrombosis outcome (including diagnoses of: AMI; acute stroke; angina; transient ischemic attack; peripheral arterial disease; amputation) and an **expanded** venous thromboembolism outcome (including diagnoses of: acute DVT; acute PE; venous thrombosis of devices, implants, or grafts)
2. Among health plan members identified as having a primary arterial or venous thrombotic endpoint, we will determine the absolute risk and rate of death within the following 30 days (including the date of the endpoint), stratified by event type. This will be based on hospital discharge disposition for those who died in the hospital; some Data Partners will have more broad, timely death data available, allowing us to assess out of hospital deaths, though the data may be incomplete.

Among at least one of the integrated delivery system Data Partners, we will calculate the absolute risk and unadjusted incidence rates of **primary** arterial and venous thrombotic events over the 90 days of follow-up stratified by exposure to **inpatient** anticoagulation treatment (further stratified according to dispensed anticoagulant drug during the baseline period), thrombolytic therapy, anti-platelet drugs, and receipt of any COVID-19 treatment (e.g., remdesivir, dexamethasone, monoclonal antibody, convalescent plasma), if possible, based on prescription fills in the hospital setting. We expect under-capture of inpatient medication administrations.

1.3 Relative Hazards of Thrombotic Events in Persons with COVID-19 Versus Influenza

There may be differences in the prevalence of demographic characteristics, comorbidities, laboratory abnormalities, severity of infection, usage of medications, and healthcare utilization during the baseline period between the COVID-19 and influenza cohorts. Because of the many

potential confounders relative to the number of arterial or venous thrombotic events in this analysis, we will develop propensity scores to control for confounding. We will estimate propensity scores for COVID-19 diagnosis, within each Data Partner, by using logistic regression, with COVID-19 (versus influenza) status as the dependent variable. A single propensity score will be calculated per member based on their first qualifying COVID-19 index or influenza event. We will exclude health plan members from the COVID-19 cohort whose propensity score exceeds the maximum or minimum values in the comparator cohort and vice-versa (i.e., trim the tails). We propose to use propensity score stratification (rather than matching) in order to efficiently retain health plan members in the analysis because it may be difficult to identify propensity score matches for some patients, which would reduce sample size and, consequently, power to detect associations. We will assign eligible patients to one of 50 propensity score strata, which will be formed based on the propensity score distribution among COVID-19 patients only. We will calculate average treatment effect (ATE) stratum weights for each patient based on the distribution of their exposure (COVID-19 or influenza) within their propensity score stratum. For COVID-19 patients, the ATE stratum weights will be calculated as (proportion of total patients in stratum i) / (proportion of COVID-19 patients in stratum i). For influenza patients, the ATE stratum weights will be calculated as (proportion of patients in stratum i) / (proportion of influenza patients in stratum i). Documentation related to the Sentinel propensity score analytic tool can be found online.

To address Aim 3, we will then use weighted Cox regression, accounting for propensity score, to calculate hazard ratios (HRs) with robust 95% confidence intervals (CIs; Shu et al., 2020) of the **primary** arterial and venous thrombotic endpoints between health plan members with COVID-19 compared to those with influenza.

Additionally, for the evaluation of arterial thrombotic events, we will also conduct a secondary analysis stratifying results according to baseline history of cardiovascular disease (defined by a diagnosis of AMI, stroke, coronary artery disease, cerebrovascular disease, or peripheral arterial disease [including limb ischemia]). For the evaluation of venous thromboembolism events, we will conduct a secondary analysis stratifying results according to baseline history of venous thromboembolism.

In secondary analyses, we will repeat the above analyses, restricting to members who were: 1) diagnosed in the inpatient setting at the index date, 2) diagnosed in the ambulatory setting at the index date, and 3) diagnosed by a positive nucleic acid test.

Among health plan members diagnosed with COVID-19 or influenza by either ICD-10-CM or positive nucleic acid test, we will also determine the HRs of the following secondary endpoints in each cohort:

1. Emergency department, institutional stay, or hospital discharge ICD-10-CM diagnosis for an **expanded** arterial thrombosis outcome (including diagnoses of: AMI; acute stroke; angina; transient ischemic attack; peripheral arterial disease; amputation) and an **expanded** venous thromboembolism outcome (including diagnoses of: acute DVT; acute PE; venous thrombosis of devices, implants, or grafts)
2. All-cause mortality within 30 days after an arterial or venous thrombotic event, defined by in-hospital or out-of-hospital death (the latter will be among a subset of Data Partners with available relevant data)

We will assess the effect of unmeasured confounders on the HRs of arterial thrombotic and venous thromboembolic events in the primary analysis using the E-Value.

1.4 Risk Factors for Arterial and Venous Thrombotic Events in COVID-19

To address Aim 2, among COVID-19 members diagnosed by either ICD-10-CM or positive nucleic acid test, we will use multivariable Cox regression to calculate adjusted HRs with 95% CIs of the primary arterial and venous thrombotic endpoints associated with each hypothesized risk factor of interest (i.e., older age, male sex, obesity, alcohol dependence/abuse, current tobacco use, current pregnancy, chronic kidney disease, cancer, chronic obstructive pulmonary disease [COPD], diabetes mellitus, hyperlipidemia, hypertension, neurologic diseases that promote immobility [e.g., dementia, Parkinson's disease], rheumatic disease, history of cardiovascular disease [for arterial thrombotic outcome analysis], history of venous thromboembolism [for venous thrombotic outcome analysis], atrial fibrillation, antiphospholipid antibody syndrome, inherited thrombophilia, polycythemia, dispensed anticoagulant [-183 through -3 days prior to index diagnosis], dispensed anti-platelet drug [-183 through -3 days prior to index diagnosis], dispensed statin [-183 through -3 days prior to index diagnosis], heart failure, and thrombocytosis), adjusting for all others in the model. We will not include severity of COVID-19 at diagnosis within multivariable models, since this is in the causal pathway between the risk factors of interest and arterial or venous thrombotic events; inclusion of this variable could adjust away associations between the hypothesized risk factors and primary endpoints of interest. The analyses will be repeated, restricting to members who were: 1) hospitalized at the index date, and 2) diagnosed by a positive nucleic acid test.

As a secondary analysis, the above analysis will be repeated evaluating as outcomes emergency department, institutional stay, or hospital discharge ICD-10-CM diagnosis for an **expanded** arterial thrombosis outcome (including diagnoses of: AMI; acute stroke; angina; transient ischemic attack; peripheral arterial disease; amputation) and an **expanded** venous thromboembolism outcome (including diagnoses of: acute DVT; acute PE; venous thrombosis of devices, implants, or grafts. We will also repeat the analysis restricting the COVID-19 cohort to those diagnosed in the inpatient setting, restricting to those diagnosed in the ambulatory setting, and separately to those diagnosed via a positive nucleic acid test from any care setting. Because small numbers of events are expected within some of the Data Partners for some of the analyses, which raises concerns about model convergence, this set of analyses will be conducted using combined individual-level, de-identified datasets from all participating Data Partners.

We will be performing a complete case analysis for Aim 2 and are not including any variables we know will have missing data – e.g., race and ethnicity are not included in the risk factor analyses because they are expected to be highly missing. For all analyses, we are utilizing electronic healthcare data with diagnosis and procedure codes and medication dispensings. The lack of a code or dispensing of interest is considered evidence of not having the condition or drug of interest. Data on some conditions, such as current tobacco use and obesity, and use of some medications, such as non-steroidal anti-inflammatory drugs, may be incomplete.

1.5 References

1. Arber DA, Orazi A, Hasserjian R, et al. The 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukemia. *Blood* 2016;127(20):2391-2405.
2. Shu D, Yoshida K, Fireman BH, Toh S. Inverse probability weighted Cox model in multi-site studies without sharing individual-level data. *Statistical Methods in Medical Research* 2020;29(6):1668-1681