Protocol:
Effect of dexamethasone on inpatient mortality among hospitalized COVID-19 patients

Protocol ID: RQ004-1-20210610

Principal Investigators:

Nicolle Gatto, PhD, MPH
Senior Vice President, Aetion, Inc., New York, NY
nicolle.gatto@aetion.com

Elizabeth M. Garry, PhD, MPH
Senior Principal Scientist, Aetion, Inc., Boston, MA
liz.garry@aetion.com

Aloka Chakravarty, PhD, MStat
US Food and Drug Administration, Silver Spring, MD
Aloka.Chakravarty@fda.hhs.gov

Protocol:
Effect of dexamethasone on inpatient mortality among hospitalized COVID-19 patients
Protocol ID: RQ004-1-20210610
CONTENTS

I. ABBREVIATIONS

ARDS	Acute respiratory distress syndrome
AT	As-treated censoring approach
CED	Cohort entry date
CPT	Current procedural terminology
CSI	Corticosteroids of interest (DEX, hydrocortisone, Methylprednisolone, Prednisone)
DEX	Dexamethasone
ECMO	Extracorporeal membrane oxygenation
EHR	Electronic health records
FDA	US Food and Drug Administration
HCPCS	Healthcare common procedure coding system
HCRU	Healthcare resource utilization
ICD	International classification of disease
IMV	Invasive mechanical ventilation
IT	Initial treatment censoring approach
KRT	Kidney replacement therapy
MPRED	Methylprednisolone
mWHO	Modified World Health Organization severity score
NAAT	Nucleic Acid Amplification Test
NDC	National drug code
NIV	Non-invasive ventilation
O2	Oxygen
PS	Propensity score
RCA	Research collaboration agreement
RWD	Real-world data
RSS	Risk set sampling
SA	Sensitivity analysis
SES	Socioeconomic status
WHO	World Health Organization

II. BACKGROUND

A. Emergence of COVID-19 and treatment

As of May 25, 2021, there have been over *32.7 million cases* of COVID-19 in the United States (US) since the emergence of SARS-CoV-2 in late 2019 ***[WHO 2020(a)]***. COVID-related hospitalizations are at an estimated cumulative rate of *553.8 per 100,000* in the US population, with COVID-19 related deaths in the US reaching over *587,000* ***[CDC 2020]***. Despite the availability of vaccines across the US, the burden of disease remains high and determining the effectiveness of potential COVID-19 therapeutics continues to be an urgent concern.

On June 22, 2020, the RECOVERY Collaborative Group published preliminary findings demonstrating a significant reduction in 28-day mortality among hospitalized COVID-19 patients in the UK treated with the corticosteroid dexamethasone (DEX) ***[RECOVERY 2020]***. As a result of this trial, the NIH Treatment Guidelines Panel updated their guidelines to recommend use of DEX in hospitalized patients receiving supplemental oxygen (O2) therapy or mechanical ventilation and the World Health Organization (WHO) updated their guidelines to recommend DEX in severe and critically ill hospitalized patients ***[NIH 2020(a); WHO 2020(b)]***. In early February 2020, results from the DEXA-ARDS trial among patients in Spain in intensive

care units (ICUs) with established moderate-to-severe acute respiratory distress syndrome (ARDS) and invasive mechanical ventilation (IMV), which occurs among patients with severe COVID-19, reported a decrease in 60-day mortality for patients randomized to intravenous DEX compared to continued routine intensive care [21% (29/139) vs. 36% (50 /138)] ***[Villar 2020]***. However, since the trial was stopped by the data safety monitoring board due to low enrollment rate, we focus here on the RECOVERY trial, but have incorporated a sensitivity analysis to align with the mortality endpoint used in the DEXA-ARDS trial [see *section VII. Sensitivity Analyses, SA5*].

B. Randomized Evaluation of COVID-19 Therapy (RECOVERY): Efficacy of dexamethasone

The RECOVERY trial, which first reported preliminary results on June 22, 2020, was an open-label, 1:2 randomized controlled trial of 6,425 patients randomized to either 6 mg of DEX daily for 10 days (n = 2,104) or usual care (n = 4,321). Overall, 22.9% of patients assigned to the DEX group died within 28 days of randomization vs. 25.7% of the control arm (rate ratio [RR] of 0.83, 95% CI [0.75, 0.93]). The difference in survival between the two groups varied noticeably when stratified by baseline COVID-19 severity level. While the incidence of death was lower in the DEX group, compared to the usual care group, among patients receiving IMV (RR: 0.64, 95% CI [0.51, 0.81]) and among those receiving supplemental O2 therapy but not IMV (RR: 0.82, 95% CI [0.72, 0.94]), there was no discernable difference among patients who were not receiving any supplemental O2 (RR: 1.19, 95% CI [0.91 to 1.55]) ***[RECOVERY 2020]***.

The effectiveness of DEX in preventing death among moderate to severe COVID-19 patients has not yet been confirmed or established in routine care, yet DEX utilization has substantially increased in the US since June 2020 ***[Aetion 2021]***. Thus, to evaluate the effectiveness of DEX in preventing death under routine use in a timely manner, we sought to conduct an observational cohort study using US-based real-world data (RWD).

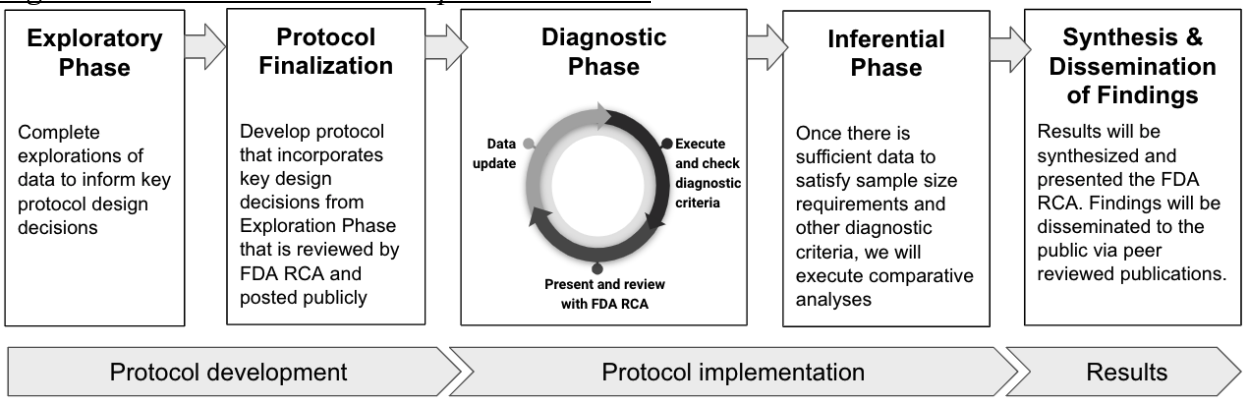
In one of the explorations conducted prior to the finalization of this protocol [see *Appendix A.2. Exploration 2*], we compared the distribution of key patient characteristics between a version of our hospitalized US-based RWD study cohort to those reported for the hospitalized UK-based patients randomized in the RECOVERY trial. The RECOVERY trial was a large, pragmatic randomized trial that simplified the data collection process in an effort to minimize the already heavy burden of the pandemic on frontline workers. Exploration of baseline clinical differences between the real-world study and trial populations was therefore limited by the degree of available information for the trial patients. However, among the available RECOVERY trial data, we identified important clinical and demographic differences in the trial population versus this study's populations at baseline.

C. Aetion Research Collaboration Agreement (RCA) with the US Food and Drug Administration (FDA)

Aetion entered a research collaboration agreement (RCA) with the US Food and Drug Administration (FDA) to use RWD to advance the understanding and the natural history of COVID-19 in specific patient populations, as well as treatment and diagnostic patterns during the coronavirus disease (COVID-19) pandemic. Aetion is collaborating with the FDA to identify and analyze fit-for-purpose data sources (i.e., near-real-time information, meaningful population size, and sufficient capture of key study parameters and longitudinal patient experience ***[Gatto 2019; FDA 2018]***) to characterize COVID-19 patient populations, identify risk factors for COVID-19-related complications, and contribute to the scientific evaluation of potential interventions.

We are using a stepwise approach to conduct this research, separating the research process into three distinct sequential phases: *Exploratory* (data explorations done in advance to facilitate key design decisions, as noted above), *Diagnostic* (requirements that must be satisfied prior to viewing treatment-specific outcomes, e.g., sample size and covariate balance), and *Inferential* (final comparative analyses) *[see Figure 1 below]*.

Figure 1. RWD Research Process phase schematic



In the *Exploratory Phase*, the Aetion research team engaged with the FDA RCA team to better understand the data and make decisions about the study design to inform the development of this protocol. The following explorations, which are further detailed in *Appendix A*, were completed prior to protocol finalization:

1. **Exploration 1**: Defining a COVID-19 study population
2. **Exploration 2**: Comparison of our study cohort and design to the RECOVERY trial
3. **Exploration 3**: Defining the time period for measuring a modified version of the COVID-19 severity score published by the World Health Organization (mWHO score) at treatment index
4. **Exploration 4**: Determining the definitions for mWHO COVID-19 severity components
5. **Exploration 5**: Determining the cohort entry date truncation

Given the periodic data updates during the *Exploratory Phase*, explorations were carried out on earlier cuts of the data. Therefore, although the cohort design remained relatively consistent across explorations, the cohort samples used for each exploration differed from each other, and from the final cohort that will be used for the diagnostic and inferential phases of the analysis [See *IV. Data Source*].

For the *Diagnostic Phase*, we developed a checklist of diagnostic criteria [see *section VI.B*, *Table 4*] that must be satisfactorily completed prior to beginning the implementation of the *Inferential Phase*. As such, the relationship between the treatment and outcome of interest will not be described or evaluated in the analytic dataset until we reach consensus that the diagnostic criteria are satisfied. Upon each data update, the Diagnostic results will be presented to the FDA RCA team until there is consensus that we have satisfied the necessary criteria to progress to the *Inferential Phase* [see *section VI.C*].

III. RESEARCH AIM AND OBJECTIVES

As part of the RCA with the FDA, this study informs the overall project to establish a system for rapid-cycle evaluation, using best practice scientific and operational methods, as applied to relevant and meaningful clinical questions. In this study, we seek to assess the effectiveness of DEX to reduce the risk of inpatient mortality within 28 days among US patients hospitalized with COVID-19 diagnosis or SARS-CoV-2 infection, overall and stratified by COVID-19 severity subgroups [defined according to the presence of encounters indicating supplemental oxygen or ventilation; see section V.D for additional subgroup detail], using the HealthVerity database.

The principles outlined by Hernan et al. emphasize the importance of explicitly emulating a hypothetical target trial when planning a RWD analysis to yield RWE [Hernan, Robins 2016; Hernan, Sauer 2016]. Although we reference the UK RECOVERY trial as a basis for comparison [RECOVERY 2020], we sought to emulate a hypothetical target trial with varied design elements, and research questions tailored for US practice and US patients. In a sensitivity analysis, we further sought to replicate the RECOVERY trial design elements (to the extent possible) [see section *VII. Sensitivity Analyses, SA8*].

Specifically, among US patients hospitalized with COVID-19 diagnosis or SARS-CoV-2 infection without prior use of any corticosteroid of interest (CSI; DEX, methylprednisolone, prednisone, hydrocortisone) estimate the hazard ratio and corresponding 95% confidence interval for the risk of inpatient mortality within 28 days after treatment initiation, comparing *new use of DEX* with:

- *non-users of corticosteroids (***Primary objective I;** *DEX+ vs. CSI-),* to evaluate the potential benefit of DEX above 'standard of care' inpatient treatment for COVID-19, where 'standard of care' excludes all CSIs (DEX, MPRED, PRED, HC).
- *non-users of DEX* (**Secondary objective I;** *DEX+ vs. DEX-*), to evaluate the potential benefit of DEX above 'standard of care' inpatient treatment for COVID-19, where 'standard of care' excludes DEX but may include those newly treated with a non-DEX CSI (MPRED, PRED, HC). This comparison is most similar to the RECOVERY trial that compared DEX to 'standard of care' ***[RECOVERY 2020]***.
- *methylprednisolone active comparator (***Secondary objective II;** *DEX+ vs. MPRED+),* to evaluate the potential benefit of DEX compared to a clinically meaningful treatment alternative (here, another corticosteroid), among patients without prior use of DEX or MPRED (traditional new user design). This comparison excludes those treated with a CSI other than DEX and MPRED.

Under the assumption that all of the CSIs are protective ***[NIH 2020(c)]***, we would expect a stronger effect for CSI- comparison **(Primary objective)**, some protective effect for the DEX- comparison **(Secondary objective I)**, and little or no effect for the MPRED+ comparison **(Secondary objective II).**

Key subgroups:
Findings will be presented for *DEX versus* each of the 3 referent groups (*CSI-, DEX-, MPRED+*) among each of the following subgroups of patients to evaluate the potential benefit of DEX compared to 'standard of care' or a clinically meaningful treatment alternative stratified by COVID-19 disease severity at the time of treatment index [defined according to the presence of encounters indicating oxygen or ventilation; see *section V.D* for additional subgroup detail]:

(1) no O2;
(2) any O2 or non-invasive ventilation (O2/NIV);
(3) any IMV; and
(4) composite of O2/NIV/IMV.

Table 1. Summary of comparisons and populations of interest for each objective

Objective	Exposure	Referent Group	Population(s) of interest
Primary	DEX+	CSI-	Overall
Secondary-I		DEX-	
Secondary-II		MPRED+	
Key subgroups		CSI-	Each of the following subsets of patients: (1) no O2; (2) O2/NIV; (3) IMV; and (4) composite O2/NIV/IMV
		DEX-	
		MPRED+	

IV. DATA SOURCE

This study uses de-identified RWD from HealthVerity comprising open and closed Medical claims and includes major payer types (commercial and federal Medicaid and Medicare programs) from all US states and territories, open and closed Medical and Outpatient Pharmacy claims, Laboratory Data including results, and Hospital Chargemaster transactional records that include diagnoses, procedures and treatments for billable inpatient and outpatient hospital encounters. The data elements derived from chargemaster data are listed in Appendix C along with their percent missingness. The database also includes clinical observations as recorded in electronic health records (EHR), where available. The open claims are sourced in near-real-time from practice management systems, billing systems and claims clearinghouses, and the closed claims are sourced from insurance providers and payers. All ages are included in the full database with ages 65 or older over-sampled to account for the under-representation of Medicare aged patients with commercial coverage. The HealthVerity Medical Claims, Pharmacy, Laboratory and Hospital Chargemaster dataset has been used for scientific publications of COVID-19 research ***[Burn 2020; Gordon 2020; Harvey 2021; Murk 2020]***.

As of ***12-March-2021***, the data spans the time period *01-December-2018* to ***22-February-2021*** (the last date of available data) and included ***58,233,249*** patients. The data will be updated periodically as new data become available until we are able to achieve sufficient sample size, as defined in the diagnostic checklist of the statistical analysis [see *section VI.B*].

The primary endpoint of 28-day mortality is defined using Inpatient Chargemaster claims with a discharge status of 'expired' indicating death. Given that the HealthVerity data source is relatively new, there are no validation studies of this data source yet for mortality or other endpoints. However, use of discharge status to determine mortality endpoints among hospitalized

COVID-19 patients has been used in both in a similar chargemaster data source ***[Rosenthal 2020]*** and for national surveillance reporting ***[CDC-NCHS 2021]***.

We further explored the data given the importance of this variable to our study. When we compared use of discharge status of death to the separate data files from the vendor providing month of death (date of death was redacted for privacy concerns), we found that all patients with a death month indicated also met our endpoint definition. Further, when we compared weekly deaths over time in our overall cohort of patients hospitalized with COVID-19 (reminder that we will not look at death stratified by treatment until after the diagnostic phase is completed), we identified similar trends to that of two external national benchmarks, the Centers for Disease Control ***{CDC-NCHS 2021}*** and data sourced from State and local health agencies {***NYT 2021***}. providing further assurance of the accuracy of the capture of our endpoint.

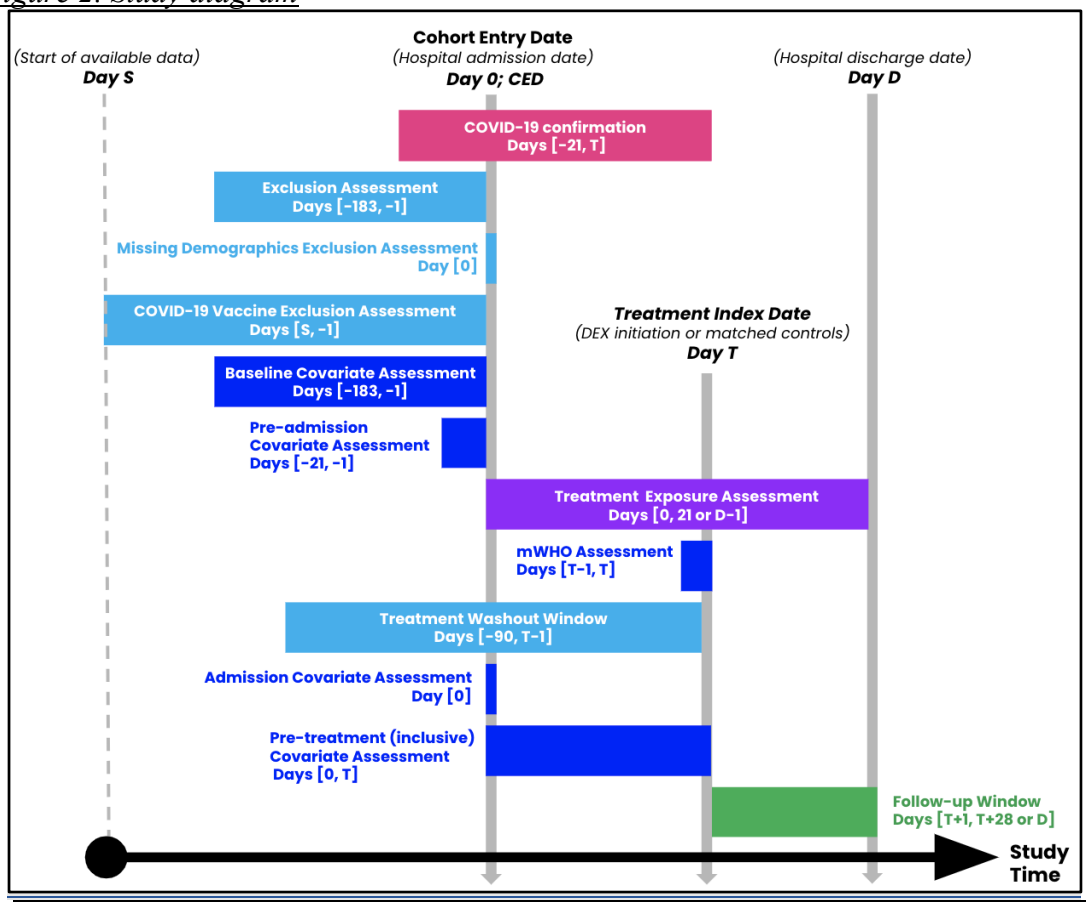
V. STUDY DESIGN

Key elements of the study design are described in the text below and additional measure definition detail is available upon request.

A. Study design

This study is an observational cohort study using secondary healthcare data. The study diagram and its components are illustrated and further detailed below *[see Figure 2]*.

Figure 2. Study diagram



Inclusion/Exclusion Criteria	
COVID-19 confirmation	Hospitalized patients with confirmed COVID-19 (ICD 10 diagnosis: U07.1 or positive or presumptive positive SARS-CoV-2 diagnostic laboratory test result; **Day -21 to Day T**).
Baseline Exclusion	Exclude patients without activity in the 183 day baseline period **(Day -183 to Day -1)**.
Missing Demo- graphics Exclusion	Exclude patients with missing age, sex, or region on the hospital admission date **(Day 0; CED).**
COVID-19 Vaccine Exclusion	Exclude patients with any record of a COVID-19 vaccine on or any time prior to the treatment index date given that these patients are assumed to be selectively different **(Day S to Day -1)**
Treatment Exposure	Hospitalized COVID-19 patients with new use of systemic dexamethasone (DEX+) or referent groups (CSI-, DEX-, MPRED+). Treatment index date **(Day T)** must occur between hospital admission date **(Day 0; CED)** and 1 day prior to hospital discharge date **(Day D-1)**.
Treatment Washout	Exclude patients with prior systemic CSI use from all comparisons **(Day -90 to Day T-1)**.
Covariate Assessment	
Baseline	Baseline (pre-admission) comorbidities, co-medications, and healthcare resource utilization (HCRU; **Day -183 to Day -1**).
Pre-admission	COVID-19-related symptoms, complications, and HCRU leading up to the hospital admission date (**Day -21 to Day -1**).
On admission	Demographics, facility characteristics, and covariates related to admitting status **(Day 0; CED)**.
Pre-treatment (Inclusive)	Treatments administered prior to or concurrent with treatment index date, pre-treatment clinical status, and time from admission to treatment **(Day 0 to Day T)**.
mWHO Severity	COVID-19 severity per modified version of WHO ordinal scale (mWHO), used for matching and subgrouping of DEX+ and referent patients **(Day T-1 to Day T)**.
Outcome Assessment	
Follow-up	Follow-up window for primary outcome of 28-day inpatient mortality (**Day T+1 to the earlier of Day T+28 or D**).

CED, *Cohort Entry Date****; D,*** *Hospital discharge date****; S****, Start of available data; **T**, Treatment index date*

B. Exposure definition

The primary exposure of interest is new (incident) initiation of systemic DEX. DEX exposure is defined with procedure codes (current procedural terminology, CPT-4/healthcare common procedure coding system, HCPCS ***[CMS 2020(b)]***), hospital charge codes for the corresponding text strings (e.g., *'dexamethasone', 'decadron')*, and drug names referenced from national drug codes (NDC), where there is an absence of evidence of a non-systemic route of administration (e.g., ciprofloxacin 0.3%/dexamethasone 0.1% ear drops are excluded). All available data types with treatment-specific information are used for assessment of DEX exposures, including medical claims, pharmacy claims, and chargemaster records. DEX use is considered "new" if the patient had no record of systemic DEX dispensing or remaining outpatient pharmacy supply during the 90-day *Treatment Washout Window* prior to the cohort entry and up until 1 day prior to treatment index (Day 0, hospital admission date).

C. Study population

1. Cohort entry

Hospitalized Patients with COVID-19:
Patients will enter the cohort upon the admission date of the first qualifying inpatient hospitalization observed in the hospital chargemaster data. The first eligible cohort entry date is **01-April-2020**, the date when the ICD-10 code for COVID-19 diagnosis was added. The last possible cohort entry date will be 60 days prior to the last date of available data [see *Appendix A.5 Exploration 5* for additional detail justifying this design decision].

COVID-19 confirmation was defined as either of the following occurring between 21 days prior to the hospital admission date and the treatment index date, which is described in section C.2. below (inclusive; **Day -21** to **Day T**):

- Outpatient medical encounter prior to or on the admission date or diagnosis 'present at admission' with ICD-10-CM diagnosis of U07.1 for COVID-19, virus identified; OR
- positive or presumptive positive SARS-CoV-2 diagnostic laboratory test result

The decision to include either diagnosis or positive laboratory test results was based on an initial exploration [see justification detail in *Appendix A.1. Exploration 1*].

For the DEX+ versus MPRED+ comparison (**Secondary Objective II**), given the increased use of MPRED earlier on in the COVID-19 pandemic, we may consider delaying the cohort entry to start in May or June pending data exploration.

2. Treatment index

Primary Objective -- New users of DEX (DEX+) versus matched non-users of CSI (CSI-):
From the patients who enter the cohort, new users of DEX between the admission date (*Day 0*) and 1 day prior to discharge date (*Day D-1*) or a maximum period of 21 days after admission (*Day 21*) will be selected. Greater than 80% of the patients who initiated DEX did so within 1 or 2 days of their hospitalization. A sensitivity analysis (SA2) will be conducted to explore initiation dates further. For each new user of DEX, at least 1 patient who has not (yet) used any CSI as of the DEX patient's index date or within the 90-day washout period prior matched on the

criteria defined in Table 2 below will be randomly risk-set sampled (RSS). DEX+ versus CSI- matching ratios of 1:1, 1:2, 1:3, and 1:4 will be considered to select up to 4 non-user referent patients for each exposed, and the ideal ratio will be selected based on the diagnostic criteria [defined below in section *VI.B. Diagnostic phase and checklist*].

RSS matching allows us to proxy the 'standard of care' by randomly selecting 1 or more patients who have not initiated DEX from the risk set of patients who are eligible for initiating DEX at the same time ***[Li 2001]***. Since the COVID-19 pandemic and treatment of patients with COVID-19 is rapidly changing over time, there is much potential for calendar-time bias. Furthermore, patients prescribed DEX may appreciably differ from those not prescribed DEX. Employing an RSS that matches DEX users and non-users on the date of DEX treatment (in the users) and time since admission will minimize the potential for this bias. Cohort entry month will also be included in the propensity score (PS) calculation to further adjust for time-related biases. Geographic region was considered for the RSS matching criteria. However, during our data explorations, the comparison of baseline balance showed that matching without region yielded better baseline balance than matching with region. Therefore, we removed region from the RSS matching. However, given known regional differences, we still account for region in our propensity score model.

The *treatment index date* (Day T) for patients who are users and for the matched non-user referent patients will be defined by the index date of the matched users.

Secondary Objective I -- New users of DEX (DEX+) versus matched non-users of DEX (DEX-):
From the patients who enter the cohort, new users of DEX between the admission date (*Day 0*) and 1 day prior to discharge date (*Day D-1*) or a maximum period of 21 days after admission (*Day 21*) will be selected. For each new user of DEX, at least 1 patient who has not (yet) used DEX as of the DEX patient's index date or within the 90-day washout period prior matched on the criteria defined in Table 2 below will be randomly RSS. DEX+ versus DEX- matching ratios of 1:1, 1:2, 1:3, and 1:4 will be considered to select up to 4 non-user referent patients for each exposed, and the ideal ratio will be selected based on the diagnostic criteria [defined below in section *VI.B. Diagnostic phase and checklist*].

Secondary Objective II -- New users of DEX (DEX+) versus new users of MPRED (MPRED):
From the patients who enter the cohort, all patients who begin use of DEX or MPRED between the admission date (*Day 0*) and 1 day prior to discharge date (*Day D-1*) or a maximum period of 21 days after admission (*Day 21*). New use will be determined using a 90-day washout period to exclude patients with prior use of either DEX or MPRED during the 90 days prior. If a patient uses both treatments over the course of the admission, they will be assigned to their first qualifying treatment arm. If a patient receives both DEX and MPRED on the treatment index, they will be excluded since the effects or the drugs cannot be disentangled among these patients.

Table 2: Risk set sampling (RSS) matching criteria

RSS matching criteria	Variable type	Categories (for categorical) / calipers (for continuous)
Calendar date	continuous	±3 days
Age	continuous	±3 years
Sex	categorical	male, female
Combined comorbidity score over 183d baseline period ***[Gagne 2011]***	categorical	0-1, 2-3, 4-5, 6+
Time since hospital admission	categorical	0-1, 2-3, 4-5, 6 - 9, 10 - 14, 15 - 19, 20+ days
COVID-19 severity at treatment index date, per modified WHO (mWHO) score	categorical	No O2, O2/NIV, IMV (highest recorded from days T-1 to T) [see additional subgroup detail below in section *V.D. Subgroups*]

3. Exclusion criteria

Patients who meet any of the following criteria will be excluded from the study cohort:

- No interaction with the healthcare system (medical encounter) in the 183 days prior to and including hospital admission (*Day -183* to *Day 0*) to minimize the potential for misclassification of baseline covariates and/or new use
- No age or sex recorded in the 183 days prior to and including hospital admission (*Day -183* to *Day 0*) as required for matching
- No geographic region recorded in 183 days prior to and including hospital admission (*Day -183* to *Day 0*) as necessary to adjust for regional variation
- Any recorded use of any CSI (dispensing or remaining supply) in the 90-day washout prior to the treatment index date (*Day T-90* to *Day T-1*) to satisfy new use definition
- Any record of a COVID-19 vaccine recorded on or any time prior to the treatment index date given that these patients are assumed to be selectively different (*Day S* to *Day T*)

The following will be evaluated (one at a time) in sensitivity analyses in order to tighten or loosen some of the exclusion criteria:

- Two sensitivity analyses will restrict the study population by adding the following further exclusions: (1) patients who entered the cohort prior to the release of the initial RECOVERY trial results; and (2) patients who initiated DEX use more than 1 day after the hospital admission date [see section *VII. SENSITIVITY ANALYSES*, SA1, SA2].
- Two sensitivity analyses will broaden the study population by loosening exclusions: (1) extend the lookback period to confirm a prior healthcare encounter to 365 days to allow for better capture of covariates among patients who do not have as frequent medical encounters (office visit or otherwise); and (2) allow patients with confirmed COVID-19 only after treatment index during the index hospitalization (no confirmation on/prior to treatment index or within 21 days prior to admission) [see section *VII. SENSITIVITY ANALYSES*, SA3, SA4].

D. Subgroups: COVID-19 severity at treatment index

The World Health Organization (WHO) Clinical Characterisation and Management Working Group proposed the WHO Clinical Progression Scale, which was created to classify COVID-19 patient outcomes according to disease progression and severity among ambulatory and hospitalized patients ***[WHO Working Group 2020]***. We developed an adaptation of the scale that is a modified version of the WHO scale, referred to here as the mWHO score, to categorize

COVID-19 severity at treatment index among hospitalized patients. This mWHO score restricts to the WHO Scale categories specific to hospitalized COVID-19 patients (categories 4-9 in the original scale), which are then collapsed into three mutually exclusive categories based on supplemental oxygen and other ventilation support requirements (no oxygen, oxygen or NIV, IMV). The scale was operationalized by incorporating data from Day T-1 to Day T to account for the lag time in recording ventilation and oxygen supplementation data in the medical claims and chargemaster data. Patients will be assigned to one of the following mWHO categories based on the highest severity feature recorded from 1 day prior to treatment index to the treatment index date (*Day T-1* to *Day T*):

- No oxygen at treatment index or within 1 day prior (No O2)
- Any oxygen or non-invasive ventilation with no record of invasive ventilation at treatment index or within 1 day prior (O2/NIV)
- Any use of invasive mechanical ventilation at treatment index or within 1 day prior (IMV)
- Composite of any O2/NIV or IMV at treatment index or within 1 day prior

Explorations were conducted to (1) determine the assessment period and justify the additional lookback to day T-1 for capture of the mWHO categories [see *Appendix A.3 Exploration 3]* and (2) justify the components of the mWHO algorithm definition [*Appendix A.4. Exploration 4*]. respectively. The definitions of these severity categories included procedure codes (O2 and NIV procedure codes for O2/NIV; intubation, IMV, and extracorporeal membrane oxygenation [ECMO] procedure codes for IMV) and diagnoses indicating a clinical need for O2 (diagnosis of hypoxia or hypoxemia) or IMV (diagnosis of ARDS). While the primary algorithms used to define O2/NIV and IMV refer to both procedural and diagnostic encounters in the data, a sensitivity analysis was also added to evaluate the robustness of findings among the mWHO categories using an algorithm based on procedures alone [see section *VII. SENSITIVITY ANALYSES*, SA7].

Another COVID-19 severity score was developed by the FDA Sentinel Initiative to classify patients as asymptomatic/very mild, mild, moderate, severe, and critical ***[Yih 2020]***. However, this score requires day-level diagnoses that are not always available via the HealthVerity inpatient chargemaster data (which is generally limited to diagnoses reported on the admission and discharge date). Therefore, use of this FDA Sentinel Initiative COVID-19 severity score is not feasible in this implementation.

Additional subgroups were added for patients who were in the intensive care unit upon treatment index and those who were not as a sensitivity analysis [see section *VII. SENSITIVITY ANALYSES*, SA6].

E. Primary outcome definition and follow-up (28-day mortality)

The primary outcome will be inpatient mortality over a 28-day period, sourced from the chargemaster discharge status field. In the primary analysis, follow-up will begin 1 day after treatment index *(Day T+1*) and will continue until the earliest occurrence of the mortality endpoint, 28 days of follow-up reached *(Day T+28)* or discharge from hospital *(Day D)*. This approach, referred to here as the "initial-treatment" (IT) approach, assumes patients continue the initial treatment without censoring upon treatment changes, similar to the intent-to-treat approach used in the RECOVERY trial and other randomized controlled trials.

We will also implement an as-treated (AT) censoring approach that additionally censors upon treatment changes. For the DEX+ versus CSI- comparison (**Primary Objective**), this will censor the follow-up of non-users of CSI upon start of any CSI use. For the DEX+ versus DEX- comparison (**Secondary Objective I**), this will censor the follow-up of non-users of DEX upon start of any DEX use. For the DEX+ vs. MPRED+ comparison (**Secondary Objective II**), this will censor the follow-up of patients who start DEX use upon start of any MPRED use and the follow-up of patients who start MPRED use upon start of any DEX use.

A sensitivity analysis will also broaden the follow-up period for capturing inpatient mortality to consider deaths within 60 days after treatment index (T+60) if they occur during the index hospitalization. [see section *VII. SENSITIVITY ANALYSES*, SA5].

F. Covariates and their corresponding assessment windows

Table 3 provides an overview of the covariates and their corresponding assessment windows considered for inclusion in the PS model. These variables were selected *a priori* based on prior literature, including published studies ***[Izurieta 2020]*** and ***[Yih 2020]***, and discussions with clinicians, and conditions listed in the 'COVID-19 associated hospitalization related to underlying medical conditions' of the National Strategy for the COVID-19 Response and Pandemic Preparedness ***[Biden 2021, p.105]***. Although all covariates will be considered, not all covariates will ultimately be included in the final model as PS model assumptions will also be checked via the diagnostic checklist *[*see *section VI.B Diagnostic phase and checklist]*. Missingness among covariates of interest within our study cohort will be enumerated and reported in our findings.

Table 3: Description of patient characteristics and covariate assessment windows

Covariate Category ***[Assessment Window]***	**Details**	**Patient Characteristics**
Baseline Covariates *(Day -183 to Day -1)*	Baseline (pre-admission) comorbidities, co-medications, and health resource utilization.	- **Lifestyle factors:** History of smoking/tobacco use*, Insurance type (commercial versus Medicare/Medicaid coverage) - **Comorbidities and comedications**: asthma, hematological cancers, solid cancers, chronic lung disease, cardiovascular disease (any), diabetes, immunosuppressive conditions, kidney or liver disease, overweight, obesity (diagnosis in medical claims/chargemaster or body mass index ≥25-<40 in EHR), severe obesity (diagnosis in medical claims/chargemaster or body mass index ≥40 in EHR), dependence on respiratory support, frailty score, neurological/cognitive impairment, 2+ or 3+ high-risk conditions (asthma, obesity, diabetes, chronic kidney disease, severe obesity, coronary artery disease, stroke, and chronic obstructive pulmonary disease), hypertension, statin use, steroid use (any, including inhaled), any immunosuppressant use, anticoagulant and/or antiplatelet use as well as the Combined comorbidity score ***[Gagne 2011]*** - **Health resource utilization**: Number of days hospitalized, number of outpatient visits, number of pharmacy claims, number of distinct medications dispensed, any encounter indicating skilled nursing facility, influenza vaccination status

Pre-admission Covariates *(Day -21 to Day -1)* *AND* Pre-treatment Covariates *(Day 0 to Day T, inclusive)*	COVID-19 characteristics and treatments	- **Pre-treatment COVID-19 status:** dyspnea or hypoxia, pneumonia, ARDS, shock, non-respiratory organ failure, other organ support procedures (ECMO, KRT, Pressors), positive SARS-CoV-2 lab test result - **Pre-treatment medications including those that overlap with treatment index (Day T):** Anticoagulant and/or antiplatelet agents, antibiotics, other experimental COVID-19 therapy listed in NIH guidelines (hydroxychloroquine/chloroquine, remdesivir, lopinavir/ritonavir and other HIV protease inhibitors, ivermectin, IL-6 inhibitors, IL-1 inhibitors, BTK inhibitors, JAK inhibitors, interferons, convalescent plasma, IVIG-SARS-CoV-2, Mesenchymal stem cells, COVID-19 mAbs ***[NIH 2020(b) as of the date of protocol finalization and may be updated as deemed necessary])***
Pre-admission Covariates *(Day -21 to Day -1)*	COVID-19 symptoms and HCRU prior to the admission date.	- **COVID-19 symptoms:** pre-admission COVID-19 symptoms - **HCRU**: Inpatient hospitalization, emergency room encounter
Admission Covariates *(Day 0; CED)*	Demographics, hospital facility characteristics, and COVID-19 severity on the admission date.	- **Hospital and admitting characteristics**: Hospital setting type, hospital teaching status, hospital number of beds, admission source, emergency room encounter, ambulance encounter, admission type - **COVID-19 severity at admission:** mWHO severity categorization on the admission date
Pre-treatment Covariates *(Day 0 to Day T, inclusive)*	Treatments administered prior to or concurrent with treatment index date, pre-treatment clinical status, and time from admission to treatment.	- **Pre-treatment COVID-19 status:** Highest recorded mWHO severity - **Pre-treatment utilization characteristics:** Number of days hospitalized, number of days with critical/ICU care, in ICU at treatment index, number of unique department codes, number of unique procedure codes - **Demographics**: Month of treatment index, age, sex, insurance type, US region

**Smoking/tobacco use may be under-reported [see section VIII. Limitations for additional detail]*

G. Statistical Software

Primary analyses will be conducted using the Aetion Evidence Platform® (2020) software for real-world data analysis, which has been validated for a range of studies ***[Wang 2016]***. Supplemental analyses and additional plots will be created using R 4.0.3 (2020-10-10).

VI. STATISTICAL METHODS

A. Propensity score model fitting

Propensity score (PS) matching will be used to control for confounding by identifying referent patients who are comparable to DEX+ patients with respect to several covariates identified *a*

priori for their potential to introduce confounding. Multivariable logistic regression will be used to estimate the propensity to initiate DEX+ versus each referent group (CSI-; DEX-; MPRED+), based on main effects of the independent covariates without interaction terms. PS model coefficients will be estimated using the overall study cohort. Appropriate referent groups will be identified using nearest neighbor matching with a caliper width of ±1%. Referent groups for subgroup analyses will be identified using the overall PS model ***[Wang 2018, Strategy A]***.

Model fit and performance will be evaluated using the criteria outlined in *Table 4*. The PS model for the primary objective will be deemed acceptable if all the criteria described in *Table 4* are achieved. We will further attempt to satisfy all criteria for secondary objectives, but strategies such as covariate adjustment to control for residual imbalance may be used in the event that some criteria are not satisfied.

As a secondary modeling strategy, we may consider estimating PS model coefficients within each subgroup and pooling subgroup matches to form the overall cohort. This modeling strategy may be used for the final analysis if diagnostic criteria outlined in Table 4 are improved ***[Wang 2018, Strategy C]***.

In our primary variable selection strategy, we will 1) define a list of potential covariates *a priori* as described in *Section V.F*; 2) specify one or several models using all or a manually selected subset of the covariates; 3) assess model diagnostics, including but not limited to diagnostic assessments described in *Table 4*; and 4) refit models in response to diagnostic results providing information about over-fitting, positivity violations, or covariate instability. Possible strategies for improving model fits may include collapsing sparse or collinear variables into composite variables where appropriate, or removing variables from the model if covariate balance is maintained between DEX+ and referent groups. We may also consider automated variable selection strategies such as HDPS or LASSO if they improve model fits relative to manual variable selection.

B. Diagnostic phase and checklist

The diagnostic phase checklist developed for this study is shown in *Table 4* below. Upon each data update, the diagnostic results will be executed and discussed with the FDA RCA team until there is consensus that we have satisfied the necessary criteria for the primary objective in the overall cohort to progress to the *Inferential Phase*. Sample size calculations are part of the iterative cycle of the *Diagnostic Phase* rather than fully completed in advance to allow consideration of new study findings (in addition to the RECOVERY trial ***[RECOVERY 2020]***) that impact the assumptions underlying the sample size calculations. Treatment effects will not be estimated during the *Diagnostic Phase*. All treatment effects will be examined during the *Inferential Phase* after the data is locked and propensity score models are finalized.

Table 4: Diagnostic checklist to complete prior to progression to the inferential phase

#	Diagnostic	Description	✓
1	Confirm adequate sample size	Adequate sample size will be defined as the minimum sample size required to achieve 80% power using a 2-sided 95% confidence interval (0.05-*alpha)* assuming a hazard ratio of 0.8 and the baseline risk of 28-day mortality in the primary referent CSI- group estimated as the risk in the RSS-matched CSI- group prior to PS matching (since the baseline risk is otherwise unknown in the	

		US patients hospitalized for COVID-19). Further confirm which subgroups of interest achieved minimum sample size for formal hypothesis testing.	
2	Confirm positivity of variables	PS distributions will be visually inspected, and overlap in all areas of the PS distributions will be confirmed.	
3	Confirm baseline confounder balance	We will confirm that the distributions of all potential confounders are balanced for DEX+ vs. referent. Covariate balance will be defined as absolute standardized differences (ASDs) ≤ 0.10 ***[Austin 2009]***. Although variables with balance prior to PS matching may be removed from the PS model, balance of these variables will still be confirmed after PS matching. If there is a variable with a small residual imbalance ($0.10 \leq \text{ASD} \leq 0.20$), balance may be deemed acceptable if the variable does not predict the outcome among the referent group (defined as an ASD <0.10 when comparing the risk of the outcome in those with the variable vs. those without it). Evaluation of imbalance in outcome prediction will only be conducted once at the end of the diagnostic phase if all other diagnostic criteria are met.	
4	Confirm models are not overfit	We will confirm that all models contain approximately 12 exposed subjects or more per covariate	

C. Inferential phase

The *Inferential Phase* begins after all diagnostic criteria are satisfied or the protocol is publicly posted, whichever is later. To prevent changes to the underlying patient or event-level data from occurring when new data are available, the data cut will be locked at the start of the *Inferential Phase*.

For the primary outcome of 28-day mortality, the following will be reported in the overall cohort and within each subgroup of interest before PS matching and after PS matching:

- Rates reported per 1,000 person-days, overall and for each subgroup.
- Hazard ratios (HR) and corresponding 95% CIs for the effect of DEX+ compared to each referent (CSI-; DEX-; MPRED+) overall and for each subgroup using PS-matched cohorts. If residual imbalance ($0.10 \leq \text{ASD} \leq 0.15$) remains after PS matching, Cox proportional hazard models may be specified to adjust for any imbalanced confounders. Otherwise, HRs will be unadjusted.

In subgroups that do not achieve the minimum sample size requirements, results will be presented in the context of an estimation framework rather than a formal hypothesis testing framework, and findings should be interpreted with caution.

VII. SENSITIVITY ANALYSES

The following eight sensitivity analyses (SA1-SA8) in *Table 5* below were determined *a priori*. Additional sensitivity analyses may be considered *post hoc* upon review of preliminary results with the FDA RCA team. Unless otherwise specified, all sensitivity analyses will be applied in response to the primary objective but not the secondary objectives.

Table 5: Summary of all sensitivity analyses planned a priori

Category	**Brief Description**	**Detail / Purpose**
Inclusion/ exclusion criteria (more specific)	SA1: Restrict study population to hospitalizations occurring after release of initial RECOVERY findings*	Detail: Restrict cohort entry to on or after June 16, 2020, given the release of the RECOVERY results in early June 2020. Purpose: Evaluate whether there is (qualitative) heterogeneity of the HR (which may be due to treatment pattern changes) after the substantial increase in DEX use in the US.
	SA2: Restrict study population to patients with immediate treatment initiation*	Detail: Restrict DEX treatment index to DEX initiation on, or within 1d after, admission date. Purpose: Evaluate whether there is (qualitative) heterogeneity of the HR (which may be due to covariate differences in patients treated earlier vs. later in the hospital stay) when restricting to DEX use upon admission, noting that the 1 day after admission was added to account for patients who may be admitted late in the evening who do not initiate treatment until the next morning.
Inclusion/ exclusion criteria (more sensitive)	SA3: Broaden population to include patients with less frequent medical encounters*	Detail: Lengthen baseline window to lookback 365 days to allow for better capture of covariates among patients who do not have as frequent medical encounters (office visit or otherwise). Evaluate whether covariate status using 183 day versus 365 day lookback is differential by treatment. Purpose: Evaluate whether HR is robust within a broader cohort that may include people with less healthcare utilization (patients without a healthcare encounter within 183 days that do have one within 365 days prior to admission) prior to hospitalization for COVID-19. Determine if the lookback period impacts the distribution of baseline characteristics.
	SA4: Broaden patient population to include patients with confirmed COVID after treatment index*	Detail: Additionally include patients with positive or presumed positive lab values or COVID-19 diagnosis occurring anytime during hospitalization, including discharge diagnoses. Purpose: Evaluate whether HR is robust within a broader cohort including people without a COVID-19 diagnosis prior to or at admission. Will increase the sample size by approximately 10%, but may have the potential for misclassification of patients who were hospitalized for reasons other than COVID-19 who had a secondary diagnosis of COVID-19 or contracted COVID-19 during hospitalization as patients hospitalized for COVID-19.

Category	Brief Description	Detail / Purpose
Follow up censoring	SA5: Broaden the follow-up period for capturing inpatient mortality endpoints*	Detail: Determine whether treatment with DEX reduces the risk of inpatient mortality within 60 days during the initial hospitalization. Purpose: Evaluate effect of DEX on risk of inpatient death over a longer time frame. Aligns with mortality endpoint definition used in dexamethasone-ARDS trial ***[Villar 2020]***.
Additional subgroup	SA6: Stratify by ICU at Treatment index	Detail: Estimate HR within ICU subgroup, defined based on the presence of a medical encounter with a revenue or department code indicating ICU at treatment index. Purpose: Evaluate whether patients who initiate treatment while in the ICU differ from those who do not (within the overall cohort), an alternative proxy for COVID-19 severity.
Alternative subgroup definition	SA7: Define COVID-19 severity based on procedures alone	Detail: Estimate HR within each of the mWHO subgroups using an algorithm that relies on procedural encounters only for defining O2/NIV and IMV (without the additional diagnostic encounters included in the primary definitions). Purpose: Evaluate the impact of the robustness of findings among patients identified as O2/NIV and IMV and those identified as having neither (no O2) and understand the impact of the shift from lower to higher COVID-19 severity on the effect of DEX+ on 28-day mortality.
Inclusion/ exclusion criteria related to RECOVERY	SA8: One comprehensive sensitivity analysis that attempts to more closely resemble RECOVERY population	Detail: 1. Remove new user washout 2. Exclude patients aged 80 years or older 3. Exclude patients with systemic fungal infections Purpose: To more closely align with RECOVERY design and baseline population 1. RECOVERY included any use of DEX. 2. Our study population has half the number of patients 80+ as RECOVERY due to under-report of older patients in our commercial data source, and these patients may therefore be selectively different. 3. RECOVERY trial excluded patients contraindicated “in the opinion of the attending clinician.” Systemic fungal infection is the only contraindication listed in the package insert for DEX, so excluding these patients is an attempt to satisfy this criterion.

** Alternative versions of the study diagrams for these sensitivity analyses are illustrated in Appendix B.*

VIII. POTENTIAL LIMITATIONS

A. Addressed in Design and Analyses

The following limitations have been addressed via study design or analyses, but will still be taken into consideration when interpreting and disseminating the findings from the implementation of this protocol:

- **Difficulty determining COVID-19 severity for the modified WHO categorization:** Oxygen supplementation is critical to the mWHO COVID-19 severity categorization (IMV, O2/NIV, no O2) upon admission and upon treatment initiation. Although it seems IMV is generally well captured via procedural claims data alone, it is unclear if oxygen support is fully captured via procedural claims. We have developed algorithms to determine IMV or O2/NIV support based on the presence of relevant procedures or diagnoses, but determination of no O2 will rely on the absence of recorded procedures or diagnoses for IMV or O2/NIV, which may increase the potential for misclassification in the no O2 group. Furthermore, given the lack of day-level diagnostic information (i.e., only admitting and discharge diagnoses are available), there is potential for under-reporting of severity on days during the hospital stay. Lastly, reporting of ARDS used to define IMV might not have been as well captured early in the pandemic, so there is increased potential for misclassification among patients hospitalized earlier in the study period. The inclusion of ECMO in the IMV definition may minimize the potential for under-report of ARDS since ECMO has been given as a short-term rescue therapy in patients with ARDS. In SA1, we restrict cohort entry to on or after June 16, 2020, given the release of the RECOVERY results in early June 2020 [see *section VII. Sensitivity Analyses, SA1*]. Through this analysis, we will be able to see how ARDS is reported among patients after the early months to better understand this concern. Further, SA7 allows us to evaluate the impact of the shift from lower to higher COVID-19 severity on the effect of DEX+ on 28-day mortality using an algorithm that relies on procedural encounters only for defining O2/NIV and IMV [see *section VII. Sensitivity Analyses, SA7*].

- **COVID-19 diagnosis:** The unique ICD-10 code for COVID-19 diagnosis (U07.1: COVID-19, virus identified) was created by the World Health Organization (WHO) in April 2020. Prior to the widespread use of the new code, other respiratory-related codes were used for billing purposes. We chose to delay the start of our cohort selection to April when the code was available. There is a possibility that early on, some facilities were not yet using the new code. However, given our findings from exploration 1 [see *Appendix A.1. Exploration 1]*, concerns about the specificity of the ICD-10 codes other than the designated U07.1 code, we agreed to proceed with the definition of confirmed COVID-19 via diagnosis of U07.1 or positive/presumptive positive SARS-CoV-2 diagnostic laboratory test result.
- **Unstructured laboratory encounter data:** In addition to using the unique diagnosis code for COVID-19 (ICD-10: U07.1), we will include patients without a diagnosis claim if they have laboratory results identifying positive or presumed positive SARS-CoV-2 diagnostic test via nucleic acid amplification test (NAAT). The laboratory data has unstructured free-text for test type and results, which present a challenge. Although we anticipate that we will not be able to categorize all COVID-19 laboratory results, we rely on an algorithm developed by the data vendor to discern the majority of test types and results.

- **Open claims:** Data from open claims can be captured within days of a healthcare encounter to provide near-real-time data. However, the composition and completeness may be less stable for the most recent calendar dates. Based on our findings from exploration 5, we have truncated cohort entry to remove the last 60 days, which should minimize this concern [see *Appendix A.5 Exploration 5*].
- **Lack of enrollment information in open claims to determine denominator of patients at risk with the same potential to contribute baseline data:** Without beneficiary enrollment information like that included with closed claims data, use of "open" datasets limits the ability to clearly determine when a patient is observable or "at risk." However, we are confident that our cohort will include all patients hospitalized available in the data and will additionally require the presence of any encounter during the 183-day baseline period for inclusion as a proxy for enrollment and extend this period to 365 days in a sensitivity analysis.
- **Unstructured inpatient medication data:** Unlike outpatient pharmacy data, inpatient medication data in the HealthVerity data is largely captured via vendor charge code fields (e.g., 'DEXAMETHASONE 4 MG OR TABS') in addition to medical procedural claims such as HCPCS for medication administration (e.g., J1095 - INJECTION, DEXAMETHASONE ACETATE, PER 8 MG) that have non-standardized free-text descriptions rather than standardized fields (e.g., NDC codes). Therefore, text search strings will be used to query the data when procedure codes are not available. For this reason, it is possible that some of the patients included using our primary definition for systemic DEX may have non-systematic use that was not excluded due to a lack of indication in the search string. However, we assume this to be rare. Our team has developed a series of search string algorithms to minimize the potential for error and will incorporate additional manual review for key measures.
- **Potential for residual confounding**: Although best practices will be applied to control for confounding, the potential for residual confounding still exists in observational cohort analyses. We will apply RSS to match on time-related biases, which are assumed to be strong given the evolving healthcare landscape of the COVID-19 pandemic that will include other matching factors of importance (i.e., time since admission, age, sex, and comorbidity score). Furthermore, we will use PS matching on top of RSS matching and will not conduct inferential comparative analyses until we have satisfied the balancing criteria specified in the diagnostic checklist. Finally, a number of *a priori* specified sensitivity analyses will evaluate the robustness of our findings. If warranted, post hoc sensitivity analyses may also be conducted to evaluate the potential impact of unmeasured confounding. Although we recognize that factors causing residual confounding may manifest after index, this study does not account for changes in disease severity and other COVID-19 treatment received after index date to avoid additional potential bias that may be induced by conditioning on mediating variables in the causal pathway, or on colliders.
- **Completeness of tobacco/smoking covariate:** Smoking/tobacco use will be defined according to the presence of the indicator for current or past user in the electronic health record (EHR), a diagnosis indicating use or abuse of tobacco, or presence of a procedural consultation or medication dispensed for cessation. The addition of EHR data makes this definition more complete than algorithms based on claims alone, but misclassification via under-reporting of smoking behavior is likely.

B. To Be Considered in Interpretation Only

The following limitations will be taken into consideration when interpreting and disseminating the findings from the implementation of this protocol:

- **Lack of indicators of socio-economic status (SES):** The HealthVerity data leveraged for this study does not include SES. We will include insurance type in the PS model for covariate adjustment as may be indicative of SES (e.g., patients with a PPO generally have higher SES than those with HMO), but no other proxies are available to adjust for SES.
- **Lack of race/ethnicity information:** The HealthVerity data leveraged for this study does not include race or ethnicity demographic data. Therefore, we are unable to evaluate whether the effect of DEX on 28-day mortality varies by race or ethnicity, explore potential racial and ethnic disparities related to COVID-19 severity and treatment, or control for confounding by race or ethnicity.
- **Incomplete dose information:** Inpatient medication dose is not explicitly captured in the HealthVerity data. Dose can be abstracted from the free-text descriptions of the vendor codes, but requires manual review and does not allow for all medications to be assigned a dose. Thus, since a subgroup analysis by dose for DEX (or other CSIs) is not possible for all patients

IX. STRENGTHS

As noted throughout this protocol and incorporated into the limitation section above, there are many strengths of this study. As recommended to determine fit-for-purpose ***[Gatto 2019; FDA 2018]***, we selected a RWD source with near-real-time information, a reasonable population size, and sufficient capture of key study parameters and longitudinal patient experience to evaluate our research question. Further, our stepwise approach enforces best practices for application of principled RWD epidemiology. This is done first by exploring the data to reinforce data fit-for-purpose and inform key study design choices, such as the COVID-19 severity category definitions and the assessment period used to assign them, and then by using a diagnostic checklist to ensure that all necessary criteria, such as sample size and confounding control, have been fully satisfied prior to estimating and interpreting treatment effects and finalizing the inferential analysis plan. Lastly, sensitivity analyses have been developed a priori with documented justifications to evaluate the robustness of findings.

X. HUMAN SUBJECTS

The use of this de-identified data source was approved for exemption by the New England Independent Review Board.

XI. REFERENCES

Aetion. 'Usage of medications in-hospital' chart generated in Real-Time Insights and Evidence, a solution powered by Aetion and HealthVerity illustrating U.S. hospital visits over time, with corresponding use of common medicines used to treat COVID-19. https://aetion.com/covid-19/, Accessed 15-January-2021.

Austin PC. Balance diagnostics for comparing the distribution of baseline covariates between treatment groups in propensity-score matched samples. Stat Med. 2009 Nov 10;28(25):3083-107. doi: 10.1002/sim.3697. PMID: 19757444; PMCID: PMC3472075.

Biden JR. National Strategy for the COVID-19 Response and Pandemic Preparedness. 2021 Jan 21. https://www.whitehouse.gov/wp-content/uploads/2021/01/National-Strategy-for-the-COVID-19-Response-and-Pandemic-Preparedness.pdf

Burn E, Sena AG, Prats-Uribe A, et al. Use of dialysis, tracheostomy, and extracorporeal membrane oxygenation among 240,392 patients hospitalized with COVID-19 in the United States. medRxiv [Preprint]. 2020 Nov 27:2020.11.25.20229088. Doi: 10.1101/2020.11.25.20229088. PMID: 33269356; PMCID: PMC7709172.

Centers for Disease Control (CDC). COVIDView: A weekly surveillance summary of U.S. Covid-19 activity. Cdc.gov. Published December 11, 2020. Accessed March 23, 2021. https://www.cdc.gov/coronavirus/2019-ncov/covid-data/covidview/index.html

Centers for Medicare and Medicaid Services (CMS) (a). 2020 - ICD-10-CM Official Guidelines for Coding and Reporting FY 2021 (October 1, 2020 - September 30, 2021). Accessed March 23, 2021. https://www.cms.gov/Medicare/Coding/MedHCPCSGenInfo/HCPCS_Coding_Questions https://www.cms.gov/files/document/2021-coding-guidelines-updated-12162020.pdf

Centers for Medicare and Medicaid Services (CMS) (b). HCPCS Coding Questions. Page Last Modified: February 11, 2020. Accessed March 23, 2021. https://www.cms.gov/Medicare/Coding/MedHCPCSGenInfo/HCPCS_Coding_Questions

Centers for Disease Control National Center for Health Statistics (CDC-NCHS). In-hospital mMortality among hospital confirmed COVID-19 encounters by week from selected hospitals. Page last reviewed: May 17, 2021. Accessed May 20, 2021. https://www.cdc.gov/nchs/covid19/nhcs/hospital-mortality-by-week.htm

Food and Drug Administration (FDA). Framework for FDA's Real-World Evidence Program, Food and Drug Administration, 2018. https://www.fda.gov/media/120060/download

Gagne JJ, Glynn RJ, Avorn J, Levin R, Schneeweiss S. A combined comorbidity score predicted mortality in elderly patients better than existing scores. J Clin Epidemiol. 2011 Jul;64(7):749-59. doi: 10.1016/j.jclinepi.2010.10.004. Epub 2011 Jan 5. PMID: 21208778; PMCID: PMC3100405.

Gatto NM, Reynolds RF, Campbell UB. A Structured Preapproval and Postapproval Comparative Study Design Framework to Generate Valid and Transparent Real-World Evidence for Regulatory Decisions. Clin Pharmacol Ther. 2019;106(1):103-115. doi:10.1002/cpt.1480

Gordon DE, Hiatt J, Bouhaddou M, et al. Comparative host-coronavirus protein interaction networks reveal pan-viral disease mechanisms. Science. 2020 Dec 4;370(6521):eabe9403. doi:10.1126/science.abe9403. Epub 2020 Oct 15. PMID: 33060197.

Harvey RA, Rassen JA, Kabelac CA, Turenne W, Leonard S, Klesh R, Meyer WA 3rd, Kaufman HW, Anderson S, Cohen O, Petkov VI, Cronin KA, Van Dyke AL, Lowy DR, Sharpless NE, Penberthy LT. Association of SARS-CoV-2 Seropositive Antibody Test With Risk of Future Infection. JAMA Intern Med. 2021 Feb 24. doi: 10.1001/jamainternmed.2021.0366. Epub ahead of print. PMID: 33625463.

Hernán MA, Robins JM. Using Big Data to Emulate a Target Trial When a Randomized Trial Is Not Available. Am J Epidemiol. 2016 Apr 15;183(8):758-64. doi: 10.1093/aje/kwv254. Epub 2016 Mar 18. PMID: 26994063; PMCID: PMC4832051. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4832051/

Hernán MA, Sauer BC, Hernández-Díaz S, Platt R, Shrier I. Specifying a target trial prevents immortal time bias and other self-inflicted injuries in observational analyses. J Clin Epidemiol. 2016 Nov;79:70-75. doi: 10.1016/j.jclinepi.2016.04.014. Epub 2016 May 27. PMID: 27237061; PMCID:PMC5124536. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5124536/

Izurieta HS, Graham DJ, Jiao Y, Hu M, Lu Y, Wu Y, Chillarige Y, Wernecke M, Menis M, Pratt D, Kelman J, Forshee R. Natural history of COVID-19: Risk factors for hospitalizations and deaths among >26 million U.S. Medicare beneficiaries. J Infect Dis. 2020 Dec 16:jiaa767. doi: 10.1093/infdis/jiaa767. Epub ahead of print. PMID: 33325510; PMCID: PMC7799044. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7799044/

Li YP, Propert KJ, Rosenbaum PR. Balanced Risk Set Matching. J Am Stat Assoc. 2001; 96 (455): 870-8.

Murk W, Gierada M, Fralick M, et al. Diagnosis-wide analysis of COVID-19 complications: an exposure-crossover study. CMAJ. 2020 Dec 7;193(1):E10–8. doi: 10.1503/cmaj.201686. Epub ahead of print. PMID: 33293424; PMCID: PMC7774475.

National Institute of Health (NIH) (a). Therapeutic Management of Patients with COVID-19. NIH.gov. Accessed December 15, 2020. https://www.covid19treatmentguidelines.nih.gov/therapeutic-management/

National Institutes of Health (NIH) (b.). Blood-Derived Products Under Evaluation for the Treatment of COVID-19. Last Updated: November 3, 2020. Accessed February 3, 2021. https://www.covid19treatmentguidelines.nih.gov/immune-based-therapy/blood-derived-products/

National Institutes of Health (NIH) (c.). Coronavirus Disease 2019 (COVID-19) Treatment Guidelines: Corticosteroids. Last Updated: February 23, 2021. Accessed March 3, 2021. https://www.covid19treatmentguidelines.nih.gov/immunomodulators/corticosteroids/

New York Times (NYT). Coronavirus in the U.S.: Latest Map and Case Count. Death data sourced from State and local health agencies. Updated May 24, 2021. Accessed May 20, 2021. https://www.nytimes.com/interactive/2021/us/covid-cases.html

RECOVERY Collaborative Group, Horby P, Lim WS et al. Dexamethasone in Hospitalized Patients with Covid-19 - Preliminary Report. N Engl J Med. 2020 Jul 17:NEJMoa2021436. doi: 10.1056/NEJMoa2021436. Epub ahead of print. PMID: 32678530; PMCID: PMC7383595. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7383595/

Rosenthal N, Cao Z, Gundrum J, Sianis J, Safo S. Risk Factors Associated With In-Hospital Mortality in a US National Sample of Patients With COVID-19 [published correction appears in JAMA Netw Open. 2021 Jan 4;4(1):e2036103]. JAMA Netw Open. 2020;3(12):e2029058. Published 2020 Dec 1. doi:10.1001/jamanetworkopen.2020.29058

Villar J, Ferrando C, Martínez D, Ambrós A, Muñoz T, Soler JA, Aguilar G, Alba F, González-Higueras E, Conesa LA, Martín-Rodríguez C, Díaz-Domínguez FJ, Serna-Grande P, Rivas R, Ferreres J, Belda J, Capilla L, Tallet A, Añón JM, Fernández RL, González-Martín JM; dexamethasone in ARDS network. Dexamethasone treatment for the acute respiratory distress syndrome: a multicentre, randomised controlled trial. Lancet Respir Med. 2020 Mar;8(3):267-276. doi: 10.1016/S2213-2600(19)30417-5. Epub 2020 Feb 7. PMID: 32043986. https://www.thelancet.com/journals/lanres/article/PIIS2213-2600(19)30417-5/fulltext

Wang SV, Jin Y, Fireman B, Gruber S, He M, Wyss R, Shin H, Ma Y, Keeton S, Karami S, Major JM, Schneeweiss S, Gagne JJ. Relative Performance of Propensity Score Matching Strategies for Subgroup Analyses. Am J Epidemiol. 2018 Aug 1;187(8):1799-1807. doi: 10.1093/aje/kwy049. PMID: 29554199.

Wang S, Verpillat P, Rassen J, Patrick A, Garry E, Bartels D. Transparency and reproducibility of observational cohort studies using large healthcare databases. *Clinical Pharmacology & Therapeutics.* 2016;99(3):325-332.

World Health Organization (WHO)(a) Coronavirus Disease (COVID-19) Dashboard. Geneva: World Health Organization, 2020. Available online: https://covid19.who.int/. Accessed March 24, 2021.

World Health Organization (WHO)(b). Corticosteroids for COVID-19: Living Guidance. Updated 2 September 2020. Available online: https://www.who.int/publications/i/item/WHO-2019-nCoV-Corticosteroids-2020.1. Accessed December 29, 2020.

World Health Organization (WHO) Working Group on the Clinical Characterisation and Management of COVID-19 infection. A minimal common outcome measure set for COVID-19 clinical research. Lancet Infect Dis. 2020 Aug;20(8):e192-e197. doi: 10.1016/S1473-3099(20)30483-7. Epub 2020 Jun 12. Erratum in: Lancet Infect Dis. 2020 Oct;20(10):e250. PMID: 32539990; PMCID: PMC7292605. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7292605/

Yih WK, Hua W, Draper C, Dutcher S, Fuller C, Kempner M, Kit B, Lyons J, Driscoll MR, Toh SD, Lo Re V. COVID-19 Natural History Master Protocol. Sentinel Initiative, version 3.0, October 9, 2020. https://www.sentinelinitiative.org/sites/default/files/Methods/COVID-19_Natural_History_Protocol_V3.0.pdf

XII. APPENDICES

A. APPENDIX A: Explorations to justify design decisions

In the *Exploratory Phase*, researchers engaged in investigations to better understand the data and make decisions about the study design. The following explorations were completed in advance of the development of this protocol:

1. **Exploration 1**: Defining the COVID-19 study population (cohort)
2. **Exploration 2**: Comparison of our study cohort and study design to the RECOVERY trial
4. **Exploration 3**: Defining the time period for measuring mWHO COVID-19 severity category at treatment index
6. **Exploration 4**: Determining the definitions for mWHO COVID-19 severity components
7. **Exploration 5**: Determining the cohort entry date truncation

The explorations were carried out in advance of subsequent *Diagnostic Phase* and *Inferential Phase*. Given the periodic data updates during the *Exploratory Phase*, explorations were carried out on earlier versions of the study cohort prior to the most recent data update and finalized design decisions. Therefore, the cohorts used for each exploration may differ from each other and the final cohort used for the *Diagnostic Phase* and *Inferential Phase*.

1. Exploration 1: Defining the COVID-19 study population (cohort)

Challenge/goal: Describe ways in which the available data elements could be used for COVID-19 cohort identification (lab results and diagnoses).

Description of exploration: After identifying relevant definition components (i.e., positive/presumed positive SARS-CoV-2 laboratory results and ICD-10 U07.1 – COVID-19, virus identified), we explored the distribution and overlap of each component/feature within our COVID-19 population, specifically:

- Distribution/overlap of components: SARS-CoV-2 positive/presumed positive lab results and ICD-10 U07.1
- Distribution/overlap of assessment windows considered for cohort inclusion: pre-admission, at admission, at discharge
- Trends over time in SARS-CoV-2 positive/presumed positive lab results, ICD-10 U07.1, other coronavirus NOS (‘not otherwise specified’) codes

In addition, we also explored the distribution and overlap of the narrow definition of COVID-19 (U07.1 code or positive/presumed positive SARS-CoV-2 laboratory results) with a broader definition of COVID-19 that additionally included the following interim ICD-10 codes that may have been used for COVID-19 diagnosis prior to release and widespread adoption of the specific U07.1 code:

- B97.21 – SARS-associated coronavirus as the cause of diseases classified elsewhere
- B97.29 – Other coronavirus as the cause of diseases classified elsewhere

- J12.81 - Pneumonia due to SARS-associated coronavirus
- B34.2 - Coronavirus infection, unspecified

Summary of key findings:

Figure A.1 illustrates the overlap and non-overlap of COVID-19 confirmation via diagnosis and laboratory results. Almost all patients (99.5%) had the ICD-10 diagnosis code (U07.1), 8.8% had both the diagnosis code and the positive or presumed positive lab result, and 0.5% had only the lab result.

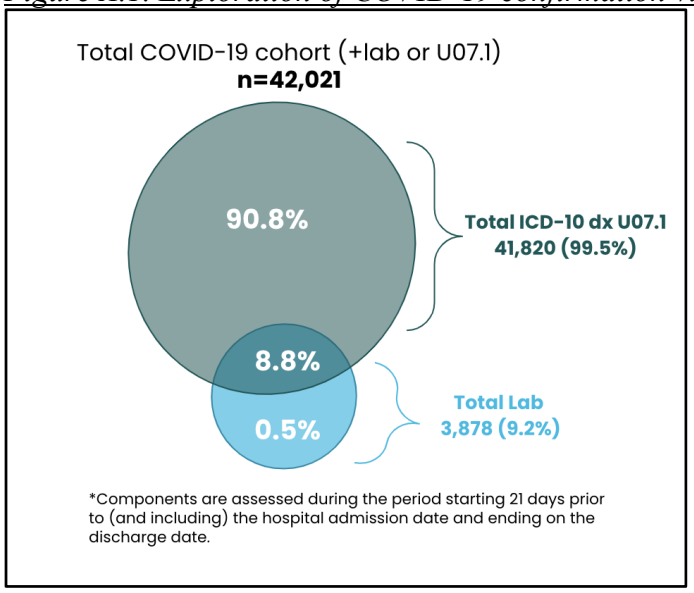
Using the narrow and broad definitions yielded cohorts of patients with similar demographic characteristics, comorbidities, and COVID-19 symptoms, but some differences in baseline medication utilization were seen; use of medications such as antihypertensives, antibiotics, and antifibrinolytics decreased over time in the broadly defined cohort, whereas medication use appeared to be more stable over time using the narrowly-defined cohort. The narrow definition alone identified approximately 84% of patients identified using the broad cohort.

Design decision:

Although findings confirmed that the majority of patients have the ICD-10 diagnosis code (U07.1), given the supporting coding guidance from CMS ***[CMS 2020(a)]***, we opted to define COVID-19 confirmation as either an ICD-10 diagnosis or a positive or presumptive diagnostic lab result.

Given concerns about the validity of the ICD-10 codes other than the designated U07.1 code, we agreed to proceed with the narrower cohort definition to increase specificity of our case definition (decrease the percentage of false positives).

Figure A.1. Exploration of COVID-19 confirmation via diagnosis and laboratory results



2. Exploration 2: Comparison of our study cohort and design to the RECOVERY trial

Challenge/goal:	Understand how our US study population and real-world study design compares to the UK RECOVERY trial on DEX.
Description of exploration:	Compare the distribution of key patient characteristics between our study cohort at the time of exploration to those reported for the DEX RECOVERY trial.
Summary of key findings:	*Table A.1* summarizes the overall comparison of our study to the DEX RECOVERY trial. *Table A.2* and *Table A.3* present the distribution of key patient characteristics, overall and stratified by COVID-19 severity, respectively. Similar to the RECOVERY trial ***[RECOVERY 2020]***, the primary objective for the current study aims to determine if treatment with DEX among patients hospitalized in the US with COVID-19 reduces the risk of inpatient mortality within 28 days among patients with supplemental oxygen and IMV respiratory support. Compared with our US RWD patients, the UK RECOVERY trial patients were older and more likely to be male, and appeared to have a lower frequency of diabetes and a higher frequency of heart disease, pulmonary disease, severe liver and kidney disease, noting that our real-world study relies on coding to capture comorbidities and that it was not possible to compare the study populations with respect to baseline characteristics not measured in RECOVERY (e.g., medication history, comorbidity scores). RECOVERY trial patients were also much more likely to start treatment on O2/NIV or IMV.
Design decision:	Given important differences identified via this exploration for the source populations, data types, and study designs used for our RWD study as compared to the RECOVERY trial, we prioritized application robust methodology for use of RWD and sought to replicate RECOVERY trial design elements where possible (e.g., use of an "initial-treatment" (IT) censoring approach that assumes patients continue the initial treatment without censor upon treatment changes, similar to the intent-to-treat approach used in the RECOVERY trial) in a sensitivity analysis [see *Section VI1, SA8*]. Differences will be taken into consideration when interpreting findings as they compare to those from the trial.

Table A.1: Summary of comparison of our real-world study to the RECOVERY trial

Design parameter	RECOVERY trial	HealthVerity study population
Study population	Compared to our study population, the RECOVERY trial population was more likely to be male, of older age, on higher doses of DEX, and more likely to be receiving O2/NIV or IMV at randomization (treatment index). Geographical differences and distinctions in US versus UK care systems may also influence the populations under study. There were study population aspects of the trial that we were not able to account for: 1. We assume that 'usual care' in the unexposed of the trial may differ from our matched non-users of DEX. 2. We were unable to account for the following trial exclusion criteria given its subjective nature "Does the patient have any medical history that might, in the opinion of the attending clinician, put the patient at significant risk if they were to participate in the trial?" 3. Since the RECOVERY trial was a large, practical trial that, while randomized, simplified the data collection process in an effort to minimize the already heavy burden of the pandemic on frontline workers, exploration of baseline clinical differences between the real-world study and trial populations to determine patient selection was limited by the degree of available information for the trial patients.	
Data collection	Collected baseline and follow-up data through web-based forms filled out by attending physicians and/or research staff at each participating hospital. Follow-up information was identified through communications with clinical staff and review of medical notes, routine healthcare system data, and registry data.	Our study used secondary data from hospital chargemaster transactional records, open and closed medical and pharmacy claims, laboratory data, and EHR records to measure baseline, exposure, and follow-up variables. We cannot measure all variables reported in the RECOVERY trial in the HealthVerity dataset (e.g., cannot reliably measure dose, viral load, eGFR) due to differences in data collection methods and available data types. Additionally, we cannot compare many baseline characteristics that are measurable in the HealthVerity dataset but were not measured in RECOVERY.
Primary exposure	Randomized to oral/IV DEX (6 mg once daily for up to 10 days) or 'usual care'. Over the course of the study follow-up, 8% of referent patients used DEX in routine care. All use was considered without requiring no prior use of DEX, although we assume no use prior to the trial.	New use of oral/IV DEX (any dose for any duration) or risk-set matched patients without DEX (as of the day the DEX patient initiated use). New use required no prior use of DEX during 90d washout.
COVID-19 inclusion criteria	Lab-confirmed Nucleic Acid Amplification Test (NAAT) SARS-CoV-2 infection.	Either of the following: ● positive or presumptive positive COVID-19 diagnostic laboratory test result; OR

		● medical encounter or diagnosis 'present at admission' with ICD-10-CM diagnosis of U07.1 – COVID-19, virus identified
Medical history exclusion	"No medical history that might, in the opinion of the attending clinician, put the patient at significant risk."	Unable to proxy this subjective criteria.
Follow-up used to capture death endpoint	Patients were followed from randomization until death or discharge or at 28 days after first randomization (whichever is sooner).	In the primary analysis, follow-up will begin 1 day after treatment index (**Day T+1**) and continue until the earliest of 28 days of follow-up reached (**Day T+28**) or discharge from hospital (**Day D**).

Table A.2: Comparison of our real-world study population to the RECOVERY trial, Overall

	Overall (all patients)			
	RSS+PS population		RECOVERY trial	
	DEX+ (N=1840)	DEX- (N=1840)	DEX+ (N=2104)	Usual Care (N=4321)
Percent of Total	**100%**	**100%**	**100%**	**100%**
Age				
mean (sd)	62.48 (17.44)	62.71 (17.05)	66.9 (15.4)	65.8 (15.4)
Age distribution				
<70 yr	1,193 (64.8%)	1,184 (64.3%)	1,141 (54.2%)	2,504 (57.9%)
70 to 79 yr	359 (19.5%)	360 (19.6%)	469 (22.3%)	859 (19.9%)
≥80 yr	288 (15.7%)	296 (16.1%)	494 (23.5%)	958 (22.2%)
Gender				
Male	866 (47.1%)	858 (46.6%)	1,338 (63.6%)	2,749 (63.6%)
Female	974 (52.9%)	982 (53.4%)	766 (36.4%)	1,572 (36.4%)
Median no. of days since hospitalization	2 [2, 3]	2 [2, 3]	2 [1, 5]	2 [1, 5]
Previous coexisting disease				
Any diabetes	601 (32.7%)	599 (32.6%)	521 (24.8%)	1,025 (23.7%)
Heart disease (CAD)	244 (13.3%)	254 (13.8%)	586 (27.9%)	1171 (27.1%)
Chronic pulmonary disease	301 (16.4%)	306 (16.6%)	415 (19.7%)	931 (21.5%)
Tuberculosis	0 (0.0%)	0 (0.0%)	6 (0.3%)	19 (0.4%)
HIV infection	7 (0.4%)	10 (0.5%)	12 (0.6%)	20 (0.5%)
Any liver disease	82 (4.5%)	84 (4.6%)	-	-
Severe liver disease	14 (0.8%)	17 (0.9%)	37 (1.8%)	82 (1.9%)
Any kidney disease	347 (18.9%)	341 (18.5%)	-	-
End-stage/severe kidney disease	62 (3.4%)	54 (2.9%)	166 (7.9%)	358 (8.3%)
SARS-CoV-2 test result				
Positive result	71 (3.9%)	64 (3.5%)	1,850 (87.9%)	3,848 (89.1%)
Negative result	39 (2.1%)	41 (2.2%)	247 (11.7%)	453 (10.5%)
No lab test (or none recorded)	1,730 (94.0%)	1,735 (94.3%)	7 (0.3%)	20 (0.5%)

Table A.3: Comparison of our real-world study population to the RECOVERY trial, stratified by COVID-19 severity

	No Oxygen				Oxygen only				Invasive Mechanical Ventilation			
	RSS+PS population		RECOVERY trial		RSS+PS population		RECOVERY trial		RSS+PS population		RECOVERY trial	
	DEX+ (N=1098)	DEX- (N=1098)	DEX+ (N=501)	Usual Care (N=1034)	DEX+ (N=592)	DEX- (N=592)	DEX+ (N=1,279)	Usual Care (N=2,604)	DEX+ (62)	DEX- (62)	DEX+ (N=324)	Usual Care (N=683)
Percent of Total	**59.7%**	**59.7%**	**23.8%**	**23.9%**	**32.2%**	**32.2%**	**60.8%**	**60.3%**	**3.4%**	**3.4%**	**15.4%**	**15.8%**
Age												
mean (sd)	61.09 (18.58)	62.03 (17.91)	71.1 (16.3)	68.5 (18.0)	64.22 (16.10)	63.55 (15.95)	67.2 (15.2)	66.4 (15.3)	63.35 (10.14)	62.63 (10.89)	58.8 (11.3)	59.2 (11.5)
Age distribution												
<70 yr	731 (66.6%)	707 (64.4%)	197 (39.3%)	462 (44.7%)	361 (61.0%)	379 (64.0%)	675 (52.8%)	1473 (56.6%)	47 (75.8%)	50 (80.6%)	269 (83.0%)	569 (83.3%)
70 to 79 yr	185 (16.8%)	210 (19.1%)	114 (22.8%)	224 (21.7%)	139 (23.5%)	121 (20.4%)	306 (23.9%)	531 (20.4%)	12 (19.4%)	7 (11.3%)	49 (15.1%)	104 (15.2%)
≥80 yr	182 (16.6%)	181 (16.5%)	190 (37.9%)	348 (33.7%)	92 (15.5%)	92 (15.5%)	298 (23.3%)	600 (23.0%)	3 (4.8%)	5 (8.1%)	6 (1.9%)	10 (1.5%)
Gender												
Male	480 (43.7%)	504 (45.9%)	286 (57.1%)	605 (58.5%)	299 (50.5%)	288 (48.6%)	819 (64.0%)	1643 (63.1%)	38 (61.3%)	32 (51.6%)	233 (71.9%)	501 (73.4%)
Female	618 (56.3%)	594 (54.1%)	215 (42.9%)	429 (41.5%)	293 (49.5%)	304 (51.4%)	460 (36.0%)	961 (36.9%)	24 (38.7%)	30 (48.4%)	91 (28.1%)	182 (26.6%)
Median no. of days since hospitalization	2 [2, 3]	2 [2, 3]	2 [1, 6]	2 [1, 5]	2 [2, 4]	3 [2, 3]	2 [1, 4]	2 [1, 4]	9 [4, 18]	7 [3, 15]	5 [3, 10]	5 [3, 9]
Previous coexisting disease	0 (0.0%)	0 (0.0%)										
Any diabetes	340 (31.0%)	330 (30.1%)	119 (23.8%)	223 (21.6%)	213 (36.0%)	210 (35.5%)	320 (25.0%)	630 (24.2%)	27 (43.5%)	29 (46.8%)	82 (25.3%)	172 (25.2%)
Heart disease (CAD)	140 (12.8%)	144 (13.1%)	180 (35.9%)	339 (32.8%)	90 (15.2%)	103 (17.4%)	357 (27.9%)	717 (27.5%)	9 (14.5%)	10 (16.1%)	49 (15.1%)	115 (16.8%)
Chronic pulmonary disease	139 (12.7%)	139 (12.7%)	121 (24.2%)	230 (22.2%)	146 (24.7%)	146 (24.7%)	259 (20.3%)	624 (24.0%)	12 (19.4%)	14 (22.6%)	35 (10.8%)	77 (11.3%)
Tuberculosis	0 (0.0%)	0 (0.0%)	2 (0.4%)	6 (0.6%)	0 (0.0%)	0 (0.0%)	1 (0.1%)	10 (0.4%)	0 (0.0%)	0 (0.0%)	3 (0.9%)	3 (0.4%)
HIV infection	2 (0.2%)	8 (0.7%)	2 (0.4%)	3 (0.3%)	3 (0.5%)	4 (0.7%)	9 (0.7%)	12 (0.5%)	1 (1.6%)	0 (0.0%)	1 (0.3%)	5 (0.7%)
Any liver disease	47 (4.3%)	55 (5.0%)	-	-	29 (4.9%)	28 (4.7%)	-	-	2 (3.2%)	3 (4.8%)	-	-
Severe liver disease	8 (0.7%)	11 (1.0%)	13 (2.6%)	19 (1.8%)	5 (0.8%)	6 (1.0%)	20 (1.6%)	52 (2.0%)	1 (1.6%)	0 (0.0%)	4 (1.2%)	11 (1.6%)
Any kidney disease	176 (16.0%)	176 (16.0%)	-	-	140 (23.6%)	153 (25.8%)	-	-	14 (22.6%)	12 (19.4%)	-	-
End-stage/severe kidney disease	31 (2.8%)	26 (2.4%)	28 (5.6%)	91 (8.8%)	20 (3.4%)	27 (4.6%)	85 (6.6%)	168 (6.5%)	2 (3.2%)	0 (0.0%)	53 (16.4%)	99 (14.5%)
SARS-CoV-2 test result												
Positive result	37 (3.4%)	37 (3.4%)	425 (84.8%)	908 (87.8%)	18 (3.0%)	22 (3.7%)	1123 (87.8%)	2293 (88.1%)	3 (4.8%)	3 (4.8%)	302 (93.2%)	647 (94.7%)
Negative result	20 (1.8%)	21 (1.9%)	74 (14.8%)	119 (11.5%)	14 (2.4%)	16 (2.7%)	152 (11.9%)	300 (11.5%)	2 (3.2%)	1 (1.6%)	21 (6.5%)	34 (5.0%)
No lab test (or none recorded)	1,041 (94.8%)	1,040 (94.7%)	2 (0.4%)	7 (0.7%)	560 (94.6%)	554 (93.6%)	4 (0.3%)	11 (0.4%)	57 (91.9%)	58 (93.5%)	1 (0.3%)	2 (0.3%)

3. Exploration 3: Defining the time period for measuring a modified version of the COVID-19 severity score published by the World Health Organization (mWHO score) at treatment index

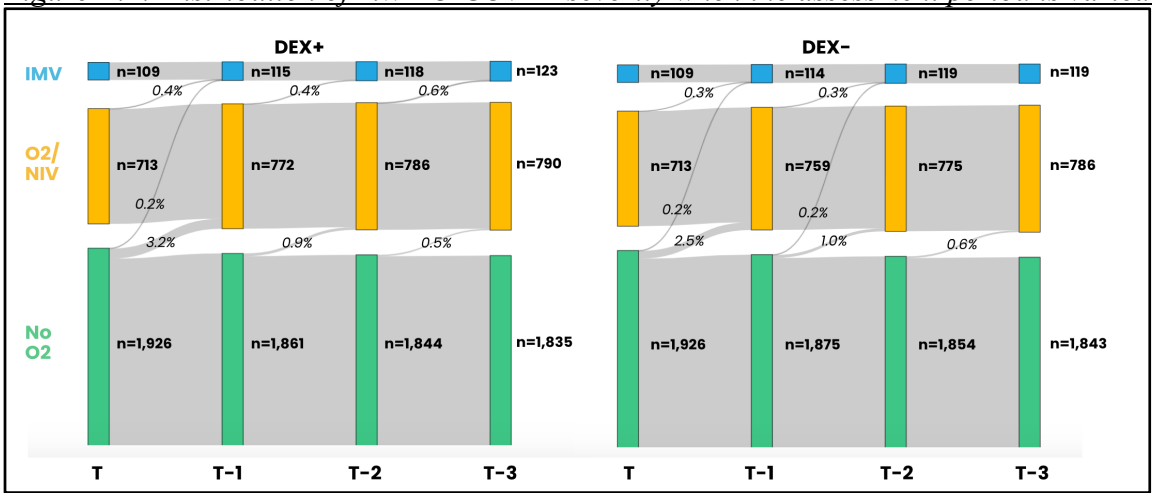
Challenge/ goal: Describe how changing the assessment period for determining mWHO severity categorization used for RSS matching and subgroups (no O2, O2/NIV, IMV) to additionally include days prior to treatment index date (**Day T**) changes the distribution of severity categorization.

Description of exploration: We aimed to account for situations where patients received respiratory support in other medical settings (e.g., emergency room), in transit prior to admission, or cases where the billing date for a procedure was captured at a later date than the procedure was performed. We began without an additional lookback evaluating only on treatment index (**Day T**), and additionally including a lookback of 1 day (**Day T-1** to **Day T**), 2 days (**Day T-2** to **Day T**) or 3 days (**Day T-3** to **Day T**) prior to treatment index. We also considered looking forward in time to account for instances where procedure dates were coded on a day after the procedure had been performed, but did not wish to introduce concern for immortal person-time bias.

Summary of key findings: *Figure A.2* illustrates how the distribution of mWHO COVID severity categorization shifts when the assessment period looks further back from treatment index date (**Day T**). When the assessment period additionally included a 1-day lookback, a small distribution of patients categorized as no oxygen when the assessment window was only on the day of treatment index (Day T) would shift to O2/NIV (2.5-3.2%) and IMV (0.2%) when the assessment period additionally included a 1-day lookback, and 0.3-0.4% of patients categorized as oxygen would shift to IMV.

Design decision: Although shifts identified via this exploration were small, they were considered meaningful. Given the trade-off for reducing misclassification without introducing bias, it was decided that we would extend the lookback to 1 day but not further, assigning the mWHO categorization based on max score between T-1 to T.

Figure A.2: Distribution of mWHO COVID severity when the assessment period is varied



All plots follow the same cohort of patients RSS-matched according to the mWHO score on day T, changing the assessment window to T-1, T-2, and T-3 for the analysis subgroups to see how patients would have been categorized differently if the window changed.

4. Exploration 4: Determining the definitions for mWHO COVID-19 severity components

Challenge/ goal:	The WHO Clinical Characterisation and Management Working Group proposed the WHO Clinical Progression Scale, which was created to classify COVID-19 patients according to disease progression and severity ***[WHO working group 2020]***. Given that the WHO Clinical Progression Scale was devised for use in studies using primary data sources, such as prospective cohort studies and randomized clinical trials, we devised a modified WHO Clinical Progression Scale (mWHO scale) to capture similar constructs of COVID-19 severity using only information available in RWD (no O2, O2/NIV, IMV) that also aligns with the COVID-19 severity categories used for the RECOVERY trial ***[RECOVERY 2020]***.
Description of exploration:	We compared procedure-based definitions alone to those that utilized diagnoses and quantified the percent of patients in each category before and after the addition of the diagnoses.
Summary of key findings:	Additional detail related to the mWHO definition components are shown in *Table A.4*. When mWHO categorization with additional clinical diagnoses [see *Figure A.3, v3*] was compared to mWHO categorization without additional clinical observations [see *Figure A.3, v1*], there were substantial shifts from the least severe category (no O2) to more severe categories (25-48% to O2/NIV and 1-2% to IMV) and from the O2/NIV to IMV category (4-6%). The shifts were differential according to treatment, as patients who initiated DEX had greater shifts and therefore differentially more misclassification/under-reporting of severity. Figure A.4 and Figure A.5 are Venn diagrams illustrating the overlap and non-overlap of patients with corresponding encounters for the procedures, chargemaster diagnoses, and medical claim diagnoses for O2/NIV and IMV, respectively, to better visualize the percent of patients that would or would not be included if any of the encounter components were to be left out of the composite definitions used to define these subgroups. From these figures, we can conclude that the shifts from no O2 to O2/NIV were primarily driven by the inclusion of clinical diagnoses in the Chargemaster data and Medical Claims, e.g., the addition of hypoxia or hypoxemia diagnosis, as 47.6% of O2/NIV patients had Chargemaster or Medical Claims but did not have other procedure-based components [see *Figure A.4*]. Similarly, 4.9% of IMV patients had a diagnosis of ARDS via Chargemaster data or Medical Claims but did not have other procedure-based components [see *Figure A.5*].
Design decision:	Given the substantial shifts, especially from no O2 to O2/NIV, we incorporated both the procedure-based encounters and the diagnoses into our final algorithm.

Table A.4: mWHO definition components

mWHO Subgroup	Components used to define mWHO subgroups	Additional Details on Component definitions in HealthVerity data
No O2	Absence of evidence of a higher category	✓
O2/NIV	Standardized Procedure-related Diagnosis Codes	✓- **ICD-10 dx for O2/ventilation use**
	Standardized Clinical Diagnosis Codes	✓- **ICD-10 dx hypoxia or hypoxemia**
	Standardized Procedure Codes	✓ - HCPCS, CPT, **Revenue Codes**
	Free-text	✓ - Charge Codes
IMV	Standardized Procedure-related Diagnosis Codes	✓- **ICD-10 dx for intubation/IMV**
	Standardized Clinical Diagnosis Codes	✓- **ICD-10 dx for ARDS**
	Standardized Procedure Codes	✓- HCPCS, CPT, ICD-10
	Free-text	✓- Charge Codes

Figure A.3. mWHO categorization before and after additional clinical observations were included (cohort RSS-matched according to v1)

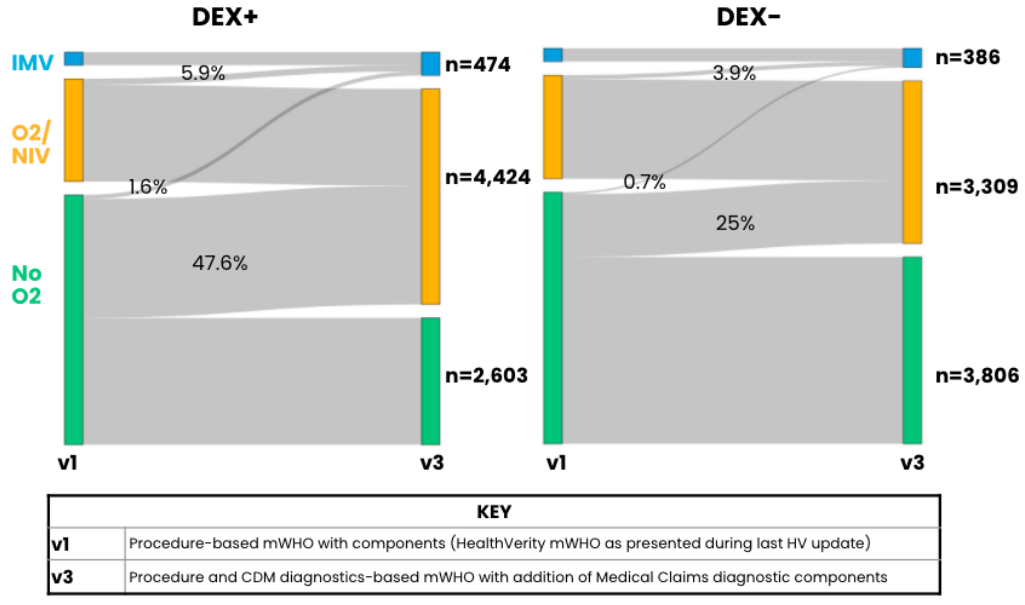


Figure A.4. O2/NIV drivers in the shift in mWHO categorization (cohort RSS-matched according to v3, n=8306)

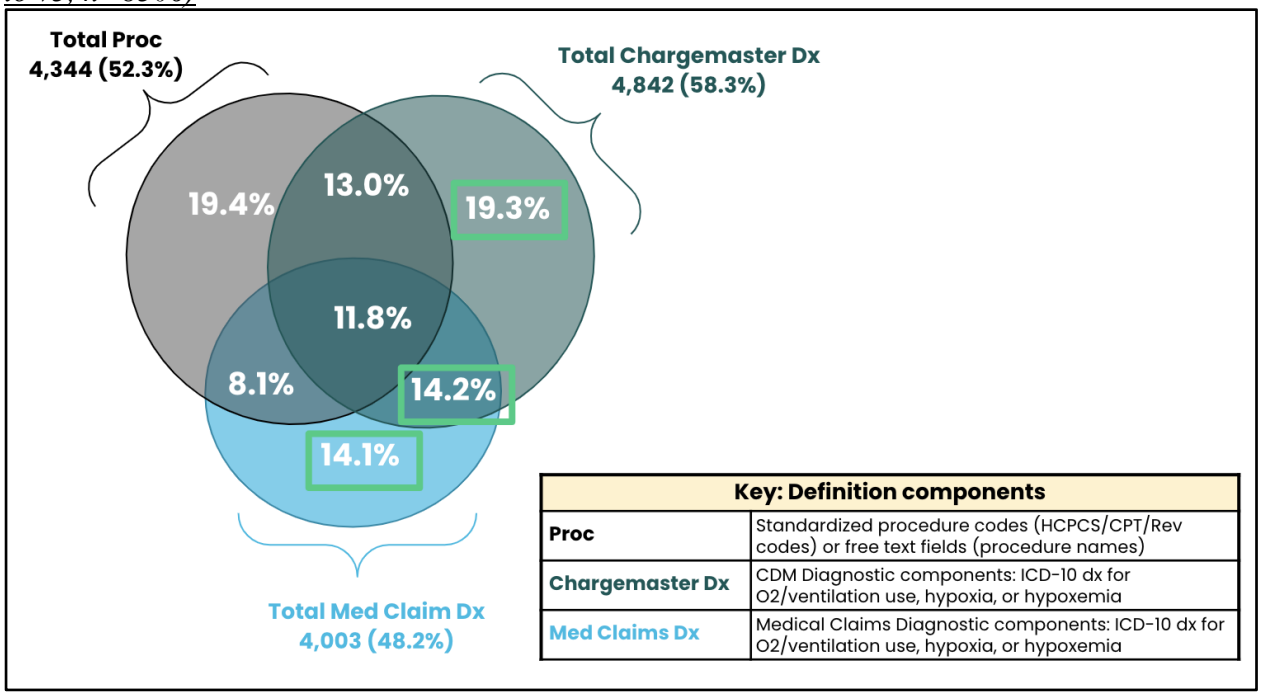
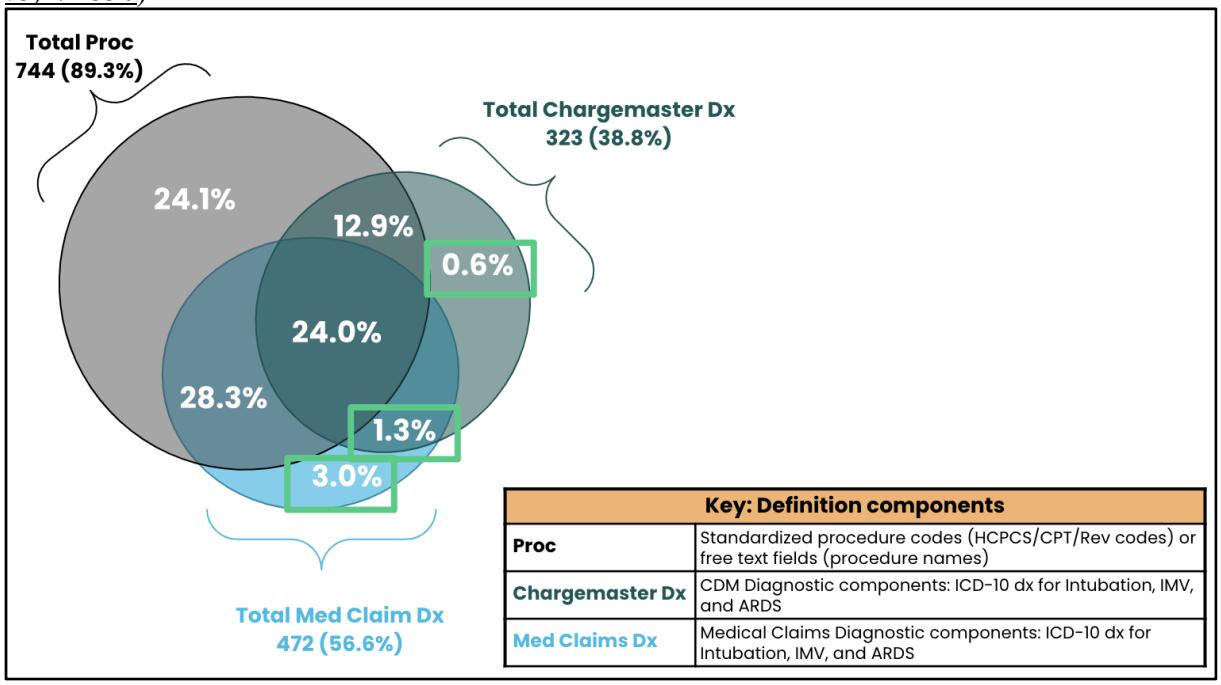


Figure A.5. IMV drivers in the shift in mWHO categorization (cohort RSS-matched according to v3, n=834)



5. Exploration 5: Determining the cohort entry date truncation

Challenge/ goal:

Although the lag (time between the latest data encounter and now) of this HealthVerity "open" claims RWD source is much shorter to provide near-real-time data, the composition and completeness may be less stable for the most recent calendar dates, creating what we refer to here as an 'artificial truncation.' This is due to two issues:

1. Variation of the update schedule cadence of each medical provider source, as some are daily while others are weekly or monthly; and
2. Hospital stay information is not included in the data until at least 10 days after a patient is discharged, so patients who are currently still hospitalized (i.e., patients with longer hospital stays) or recently discharged are not yet included in the data most proximal to the date of the data cut. This attribute is likely to skew the most recent period towards patients with shorter hospital stays.

The goal of this exploration is to understand the impact of "open" claims on the most recent data and determine the optimal latest cohort entry date in which data are reliable for comparative studies for the hospitalized COVID-19 population.

Description of exploration:

We selected our hospitalized cohort of patients with confirmed COVID-19 from 5 separate data cuts received on the following dates: 14-September-2020, 05-October-2020, 02-November-2020, 16-November-2020, 7-December-2020, 11-January-2021, 08-February-2021, and 22-February-2021.

First, we plotted and evaluated the weekly median length of hospital stays (percentiles) in the newest data cut received 7-December.

Next, we plotted and compared weekly average length of stay (LOS; i.e., duration of hospital stays) trends in an older data cut from 16-November-2020 and the newest data cut from 22-February-2021 to determine the appropriate number of days prior to the end of data to consider omitting from cohort selection due to data completeness concerns.

Lastly, we plotted and compared monthly calendar time trends for the number of admissions and deaths among each of the data cuts.

Summary of key findings:

Within our hospitalized COVID-19 cohorts, for patients in the data cut received 22-February-2021, the latest hospital admission was 07-February-2021 and the median LOS was 6 days. We see a data lag of ~15-20 days between the latest available admission date and the day of data receipt, which was also seen via other data cuts. Considering a data lag of 20 days and the 95th percentile of 27 days LOS, we assume the end of the study period should be a minimum of 47 days prior to the end of the data to account for at least 95% of the observed hospital stays [see *Figure A.6*].

In the weekly LOS plots comparing an older data cut from 16-November-2020 to a newer data cut from 22-February-2021, we began to see the lines diverge the week of 17-September-2020, approximately 60 days before the latest

admission date observed within the data cut received 16-November-2020. [see *Figure A.7*]

For each data cut, the number of monthly admissions drops quickly, and likely artificially, as the cohort entry date approaches the data cutoff [see *Figure A.8*]. Death events also drop quickly, and likely artificially, as the cohort entry date approaches the data cutoff [see *Figure A.9*]. For both admissions and deaths, data from the more recent data cuts converge with data from earlier data cuts approximately 60 days prior to the data cutoff.

Design decision: After evaluating time trends in LOS, and numbers of admissions and deaths across data cuts, we determined that the cohort entry date selection period should end 60 days earlier than the end of the data to ensure complete data for reporting the 28-day primary mortality endpoint. The truncation of the cohort entry period will further delay the *Inferential Phase* as we wait for sufficient data to accumulate in order to satisfy the diagnostic criteria.

Figure A.6. Length of hospital stay (percentiles) in the newest data cut received 22-February-2021

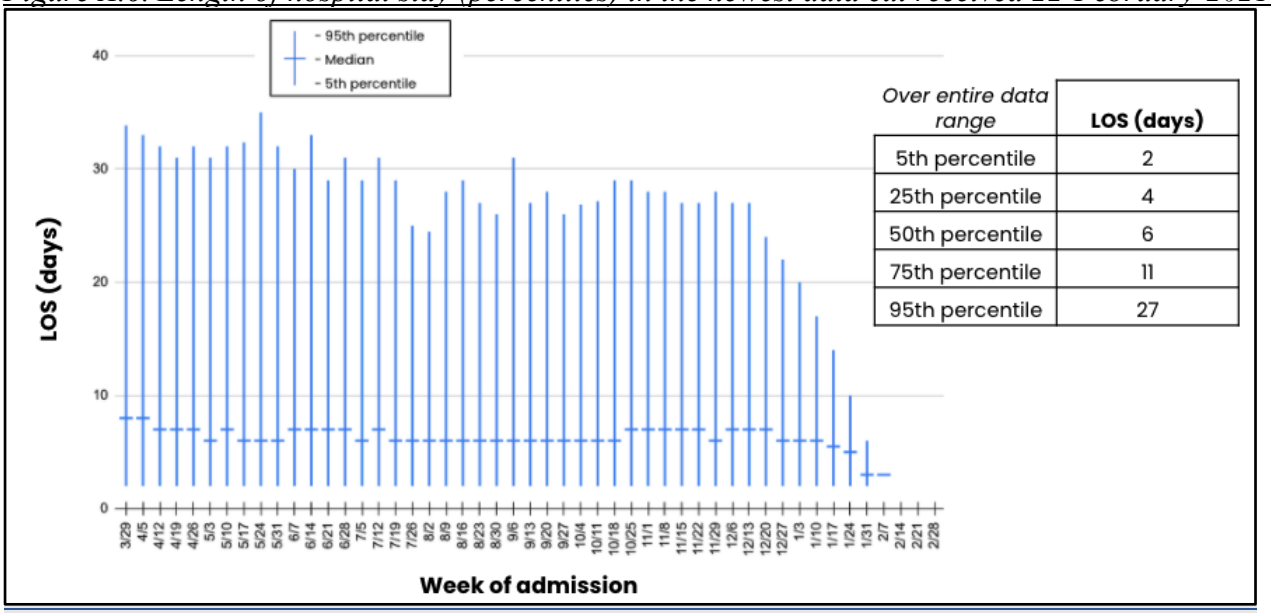


Figure A.7. Trends in chargemaster hospitalizations and average length of stay comparing older and newer data cuts

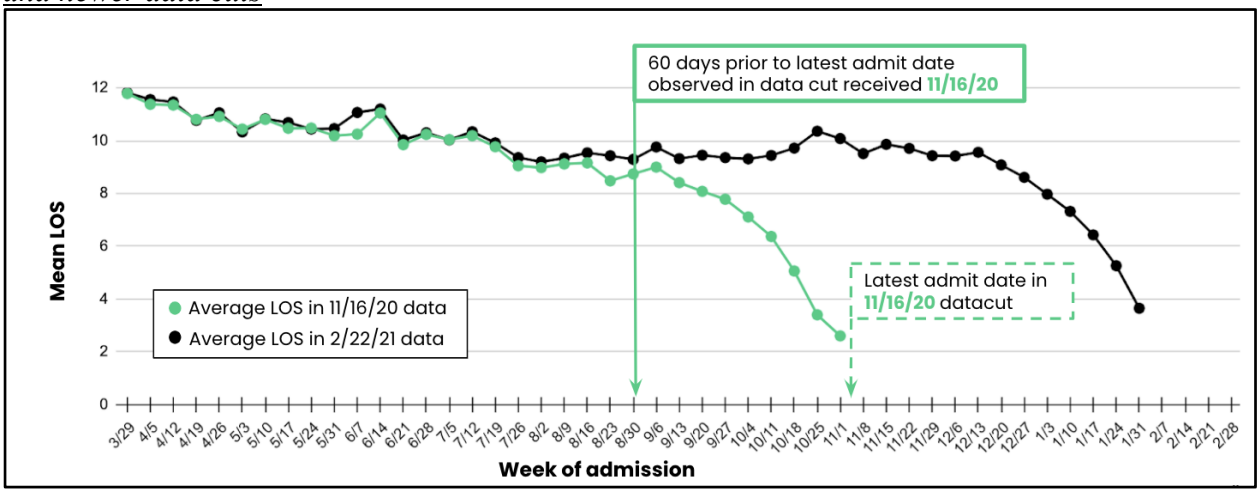


Figure A.8. Number of admissions by week in a hospitalized COVID-19 cohort

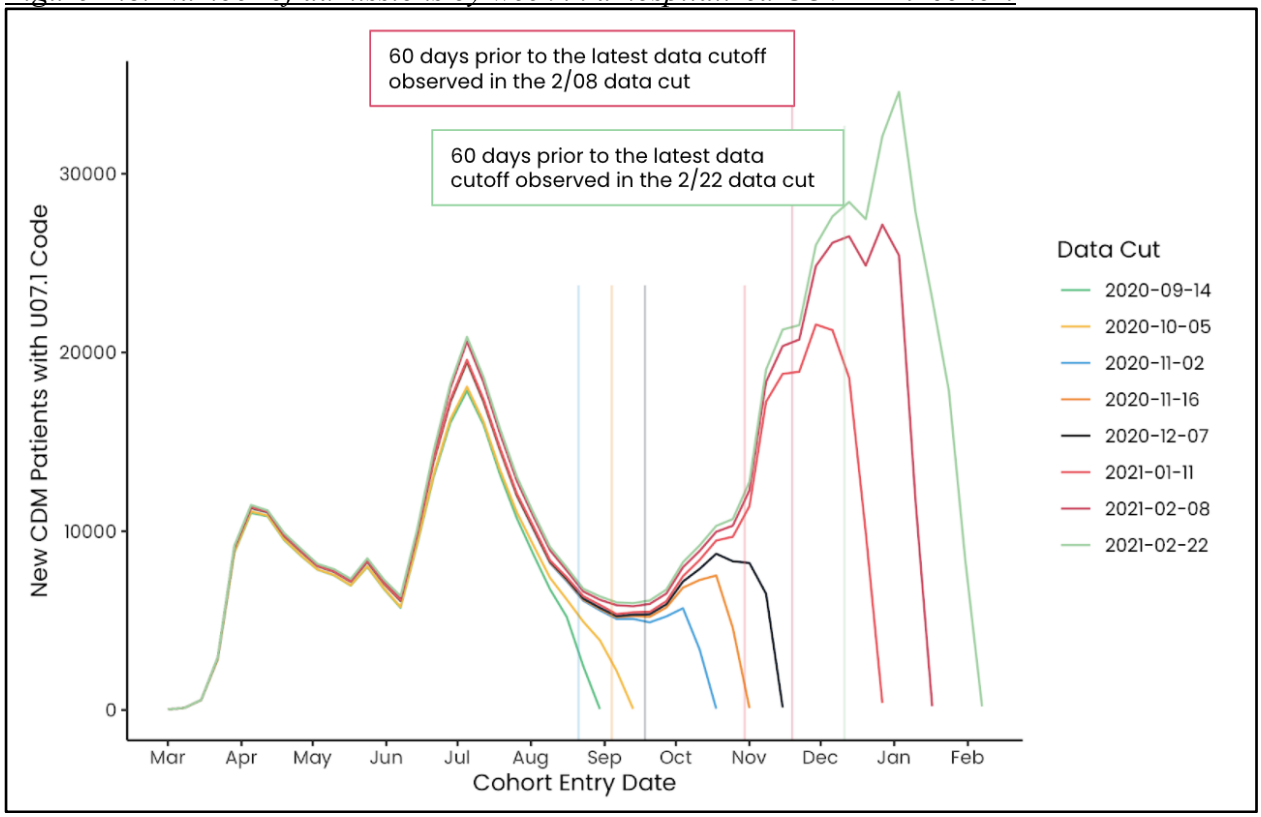
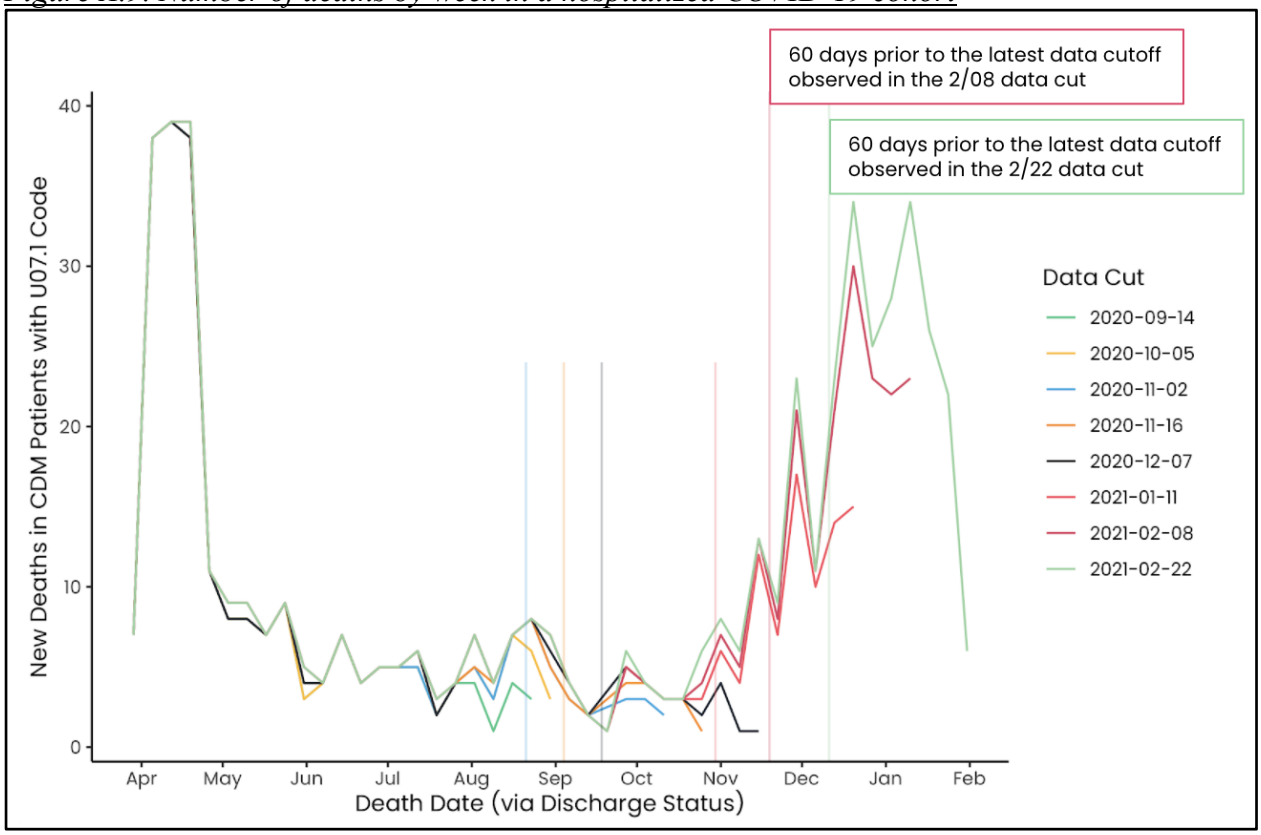


Figure A.9. Number of deaths by week in a hospitalized COVID-19 cohort



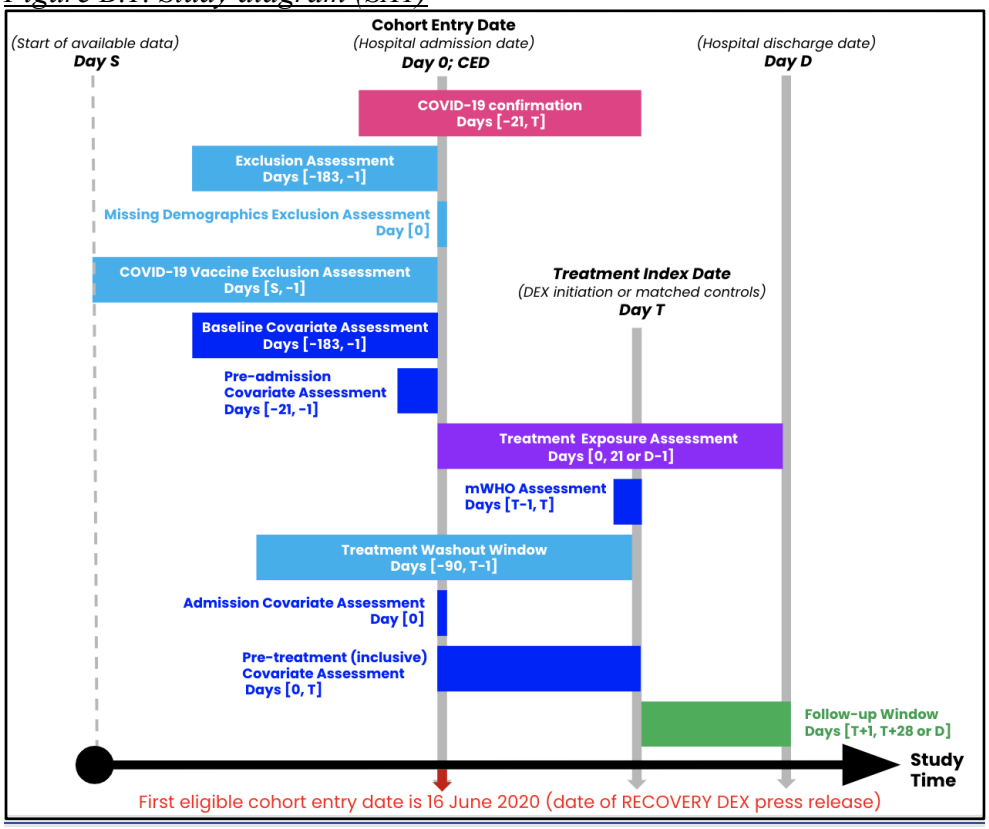
B. APPENDIX B: Study design diagrams for each sensitivity analysis

SA1: Restrict study population to hospitalizations occurring after release of initial RECOVERY findings

Detail: Restrict cohort entry to on or after June 16 given the release of the RECOVERY results in early June.

Purpose: Evaluate whether there is (qualitative) heterogeneity of the HR, which may be due to treatment pattern changes, after the substantial increase in DEX use in the US.

Figure B.1. Study diagram (SA1)

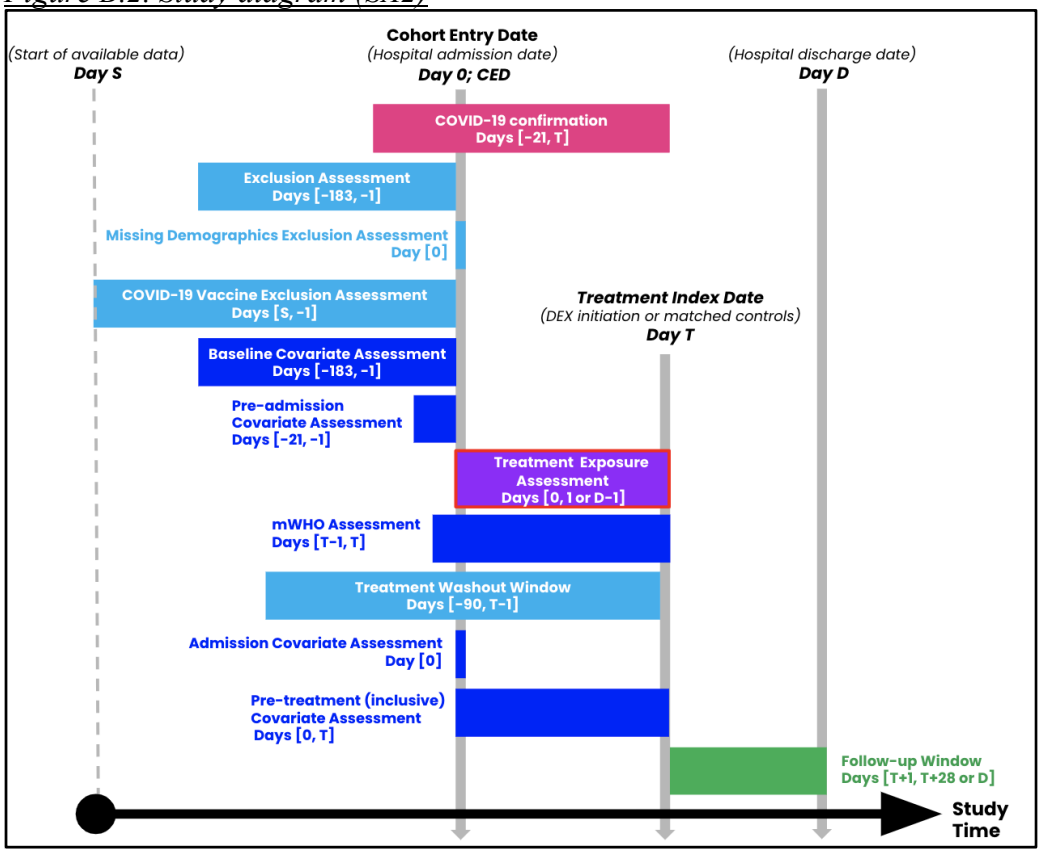


SA2: Restrict study population to patients with immediate treatment initiation

Detail: Restrict DEX treatment index to DEX initiation on, or within 1d after, admission date.

Purpose: Evaluate whether there is (qualitative) heterogeneity of the HR (which may be due to covariate differences in patients treated earlier vs. later in the hospital stay) when restricting to DEX use early in the hospital stay.

Figure B.2. Study diagram (SA2)

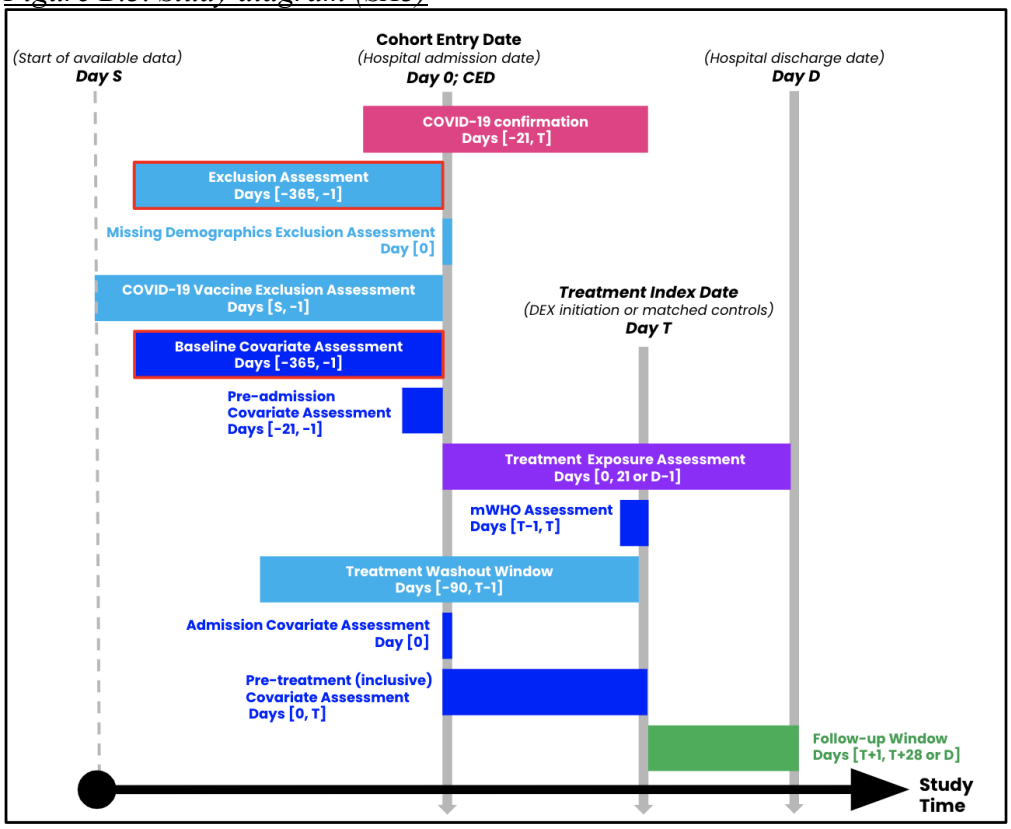


SA3: Broaden population to include patients with less frequent medical encounters

Detail: Lengthen baseline window to lookback 365 days to allow for better capture of covariates among patients who do not have as frequent medical encounters (office visit or otherwise). Evaluate whether covariate status using 183 day versus 365 day lookback is differential by treatment.

Purpose: Evaluate whether HR is robust within a broader cohort that may include people with less healthcare utilization (patients without a healthcare encounter within 183 days that do have one within 365 days prior to admission) prior to hospitalization for COVID-19. Determine if the lookback period impacts the distribution of baseline characteristics.

Figure B.3. Study diagram (SA3)

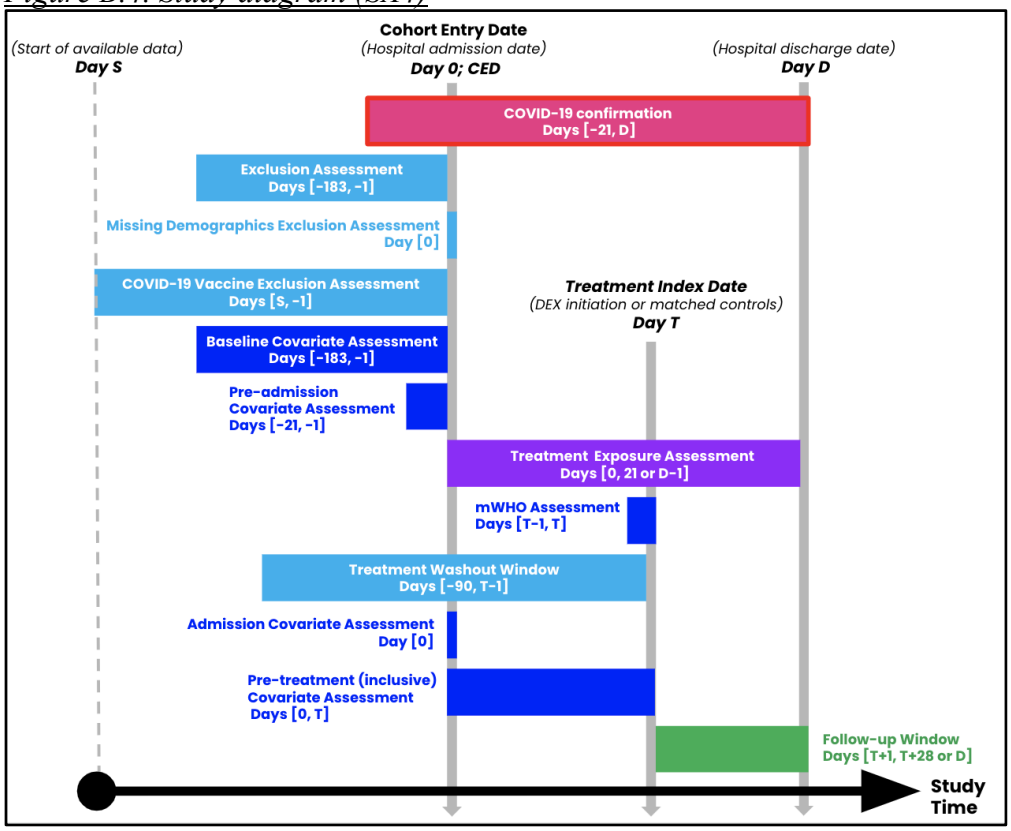


SA4: Broaden patient population to include patients with confirmed COVID after treatment index

Detail: Additionally include patients with positive or presumed positive lab values or COVID-19 diagnosis occurring anytime during hospitalization, including discharge diagnoses.

Purpose: Evaluate whether HR is robust within a broader cohort including people without a COVID-19 diagnosis prior to or at admission. Will increase the sample size by approximately 10%, but may have the potential for misclassification of patients who were hospitalized for reasons other than COVID-19 who had a secondary diagnosis of COVID-19 or contracted COVID-19 during hospitalization as patients hospitalized for COVID-19.

Figure B.4. Study diagram (SA4)



SA5: Broaden the follow-up period for capturing inpatient mortality endpoints

Detail: Determine whether treatment with DEX reduces the risk of inpatient mortality within 60 days during the initial hospitalization.

Purpose: Evaluate effect of DEX on risk of inpatient death over a longer time frame. Aligns with mortality endpoint definition used in dexamethasone-ARDS trial ***[Villar 2020]***.

Figure B.5. Study diagram (SA5)