

# Proposal: Convalescent Plasma for treatment of COVID-19 EHR Data Studies Master Protocol

## Investigators

Michael Joyner, Rickey Carter, Nigel Paneth, Arturo Casadevall, Vitaly Herasevich, Chenxi Li, Mat Reeves, Francis X. Champion, Brian Anderson, Ryan Luginbuhl, Rute Martins, Peter Smart, Rehan Waheed, Andrea Noel, Janet Campbell

## Goal

Create a flexible master protocol for the federated analysis of electronic health record (EHR) data from health systems across the United States for rapid cycle evaluation of the experience of hospitalized adult patients treated with convalescent plasma (CP) for COVID-19 infection. This method will identify early signals of efficacy and harm by evaluating key outcomes including ventilator use, death rates, and discharge disposition for COVID-19 patients receiving convalescent plasma. The first phase of the study will focus on the unadjusted outcomes. To fully validate those findings, a second phase of the study requires a second, more detailed data pull, with additional data elements such as co-morbidities and laboratory test results, enabling deeper covariate analysis.

This EHR data study is intended to run in the same hospitals participating in the Convalescent Plasma Expanded Access Program coordinated through the Mayo Clinic (also referred to as “the registry study”). Details can be found on the website <https://www.uscovidplasma.org/>

## Options for Hospital Participation

The master protocol will allow for a variety of “sub-protocols” which vary only by specific data definitions in use by varying hospitals. In this manner, hospitals using different EHR system configurations can contribute to the research effort. Hospitals have a choice to perform their own data analysis, contribute data to a participating research consortium, or simply send their de-identified data to the Mayo Clinic data coordinating center. Research groups opting to perform their own analysis are encouraged to share the same data analysis plan.

Fundamental to all treatment-control protocols will be the creation of two cohorts – one cohort comprised of the treated population and a second of untreated controls, who will be as similar as possible in risk factors for the key outcomes using available data. All protocols will attempt to create the best approximation possible, from observational data, of a randomized controlled trial of CP.

The several sub-protocols may differ in:

1. The electronic record database used to select treated and untreated controls
2. The precise definition of each variable, which is subject to the nature of available data
3. The sample size
4. The outcomes of interest
5. The stage or stages of disease focused upon
6. The calendar time period of treatment

## Design

This study is a historical observational design in which exposed patients (recipients of CP) are matched to an unexposed control population (non-recipients of CP) and followed from day of treatment in the exposed and a matched day in controls (referred to as day zero, see below) until the end of the period of observation which is designed to be at least 28 days in all subjects. This kind of design has been referred to as exposure-control study in the epidemiologic literature and differs from a traditional historical cohort design in the matching of unexposed to exposed, and from the usual case-control design by directionality, which here is from exposure to outcome and not from outcome to exposure.

### **1. Source population for the study.**

The source population for the study will be all patients, 18 years and older, with COVID-19 admitted to hospitals in the Mayo registry study of COVID-19 treated with CP. The source population will be restricted to admissions occurring between the date of admission of the first recipient of CP in each hospital until 28 days before the date of extraction of data. Patients with lesser periods of follow-up will not be included in the present study but may be studied later.

### **2. Creation of a study variable set**

All patients eligible to be included in the source population will have the data listed in the two tables below. Patients without data elements to be used in matching or to study outcomes will be excluded. Incompleteness in other variables will not be a reason to exclude the participant. Unless noted elsewhere, all variables will be extracted for the entire study population of COVID-19 patients who were admitted at least 28 days prior to the date the study is undertaken.

Both Cerner and Epic are already working on detailed queries and data definitions in support of this study. For health systems who wish to leverage this work should contact:

- Cerner EHR: contact your account executive.
- Epic EHR: contact your Epic Best Friend Forever (BFF), inquire about the “Insights Project.”

### **3. Definition of COVID-19 patient and COVID-19 related hospitalization**

The definition of COVID-19 case will be the diagnosis made in the hospital of care. It is recognized that there are several possible case definitions, but this study will not at this time attempt to impose a specific case definition unique to the study. The COVID-19 Healthcare Coalition has worked to develop common data definitions that *can* be leveraged for the study. The current definitions are included in Appendix A.

#### **4. Identification of the treated cohort.**

An attempt will be made to identify all patients in the EPIC file with the diagnosis of COVID-19 disease who were admitted to the treating hospital until 28 days before the date of extraction of data, and who received, at some point during their hospitalization, one or more transfusions of convalescent plasma, regardless of the number of units or spike antibody titer. No exclusions will be made.

#### **5. Identification of the untreated control cohort.**

The pool of potential controls is all other COVID-19 patients not treated with convalescent plasma at any time during their hospitalization, admitted between the date of admission of the first COVID-19 patient treated with convalescent plasma at that hospital and a date 28 days or more before the date of extraction of data.

#### **6. Disguising of dates.**

To maintain data confidentiality, a random number will be assigned, separately for each hospital, to the earliest date in the cohort, which is the date of admission of the first CP treated patient in each hospital. All dates subsequent to that date will be computed from that date. Thus if the first date is March 25<sup>th</sup> in hospital A, and the random number assigned to that date is 876, if a treated patient was intubated on March 29<sup>th</sup>, received CP on April 2<sup>nd</sup>, was extubated on April 5<sup>th</sup>, and discharged on April 10<sup>th</sup>, those dates would be recoded, respectively, as 880 (876 + 4), 884, 887 and 892. Calculation of duration of events (see outcome variables below) will require subtraction of the random number from the disguised date number.

#### **7. Matching process and matching variables**

The six variables used in matching are listed in Table 1 below. Hospital (to be identified only by a code), sex are dichotomous; age will be matched in ten-year intervals ( $\leq 20$  and  $>80$  are additional intervals); oxygenation requirement is defined as a proxy for severity of illness. Duration of stay prior to CP treatment will be used to determine date of cohort inception (referred to here as time zero) for the controls. The day zero for potential matched controls will be determined, and a control sought who had the same illness severity on day zero as the treated patient on the day of treatment.

If possible, that date will be the exact number of days between admission and treatment/time zero for the treated and control cohorts, but if that is not possible the match will be within 3 days duration (3 days pre or 3 days post day zero). Treatment/time zero will be the dates used to calculate all duration outcomes. Up to four controls for treated patients will be sought.

## 8. Outcomes of interest (Table 2)

**a. Primary outcome** – death at any time after admission recorded in the EMR. Deaths will be measured at 7 days, 14 days, 21 days and 28 days, and total deaths.

**b. Secondary outcomes** (all times measured from day of treatment in the treatment arm and from the corresponding day zero in controls).

- i. Number of days to first extubation (in mechanically ventilated patients)
- ii. Number of days to first intubation (in non-mechanically ventilated patients)
- iii. Number of days to death
- iv. Number of days to discharge

## 9. Data analysis

### a. Initial analyses

The initial analysis will examine death rates at the times specified above separately using mixed-effects logistic regression where the random effect account for the matched sampling. Mortality differences will be expressed as Odds Ratios for the treatment arm compared to the control arm.

For secondary outcomes, we will examine time from inception of treatment/time zero to each of the secondary outcomes specified above, including the appropriate sub-populations (mechanically ventilated at time zero/not mechanically ventilated at time zero) for analyses of time to intubation/extubation).

Time to death will be analyzed at first by the Kaplan-Meier curves and a shared-frailty (accounting for matching) model, treating discharged patients as being alive until the time of analysis. Then times to death and discharge will be analyzed jointly by a Cox-type, cause-specific hazard model with a bivariate log-normal frailty (accounting for matching), treating these two events as competing risks.

Times to extubation (intubation) and death will be analyzed jointly in the same way. In the analysis of times to extubation and death, treated patients will be further matched with untreated patients on duration of ventilation prior to CP to define time zero for the controls. In all the analyses, survival differences will be expressed as (cause-specific) hazard ratios. The

Kaplan-Meier curves, the shared-frailty model and the cause-specific hazard frailty models all account for observation time of study subjects still in hospital at the time of data analysis.

#### **b. Multivariate analysis of mortality**

In the second phase of analysis, variables in Appendix B will be examined for their effects on mortality in the entire database. Any variable associated with mortality (change in OR of 20% or more, regardless of p value) will be entered into a separate mixed-effects logistic regression model for each of the four mortality time periods and into the regression models for survival analysis. These models will account for variable matching ratios of controls to cases.

#### **10. Patient confidentiality**

As this study is restricted to the analysis of data, the only human subjects concern is confidentiality. The study does not abstract any identifying information on the participant, disguises all dates linked to individuals, and uses a large sample of participants with unidentified hospitals, thus minimizing the risk that non-identifying variables could be used to identify participants. Nonetheless, all Mayo procedures for protecting computerized data will be followed, including the use of password protection and the presence of firewalls separating the study data from other forms of data in storage.

<b>TABLE 1 – MATCHING VARIABLES</b>	
<b>Data Element Category</b>	<b>Data Element</b>
Matching	Hospital (deidentified code for each facility)
	Age <sup>1</sup>
	Administrative gender <sup>2</sup>
	Severity of respiratory illness (see definition below) on day of admission (day 0)
	Severity of respiratory illness (day 1)
	Severity of respiratory illness (day 2)
	Severity of respiratory illness (day 3)
	Severity of respiratory illness (day 4)
	Severity of respiratory illness (day 5)
	Severity of respiratory illness (day 6)
	Severity of respiratory illness (day 7)
	Severity of respiratory illness (day 8)
	Severity of respiratory illness (day 9)
	Severity of respiratory illness (day 10)
Outcome/time variable <sup>3,4</sup>	Admission date
	Date of convalescent plasma administration
	Date of first intubation (start of mechanical ventilation)
	Date of first extubation (end of mechanical ventilation)
	Date of death
	Date of discharge
Other variables	Still in hospital on date of data extraction (Y/N)

<sup>1</sup> Should be “>90 years” instead of actual age for patients over 90 for deidentification purposes

<sup>2</sup> Male, female or unknown/other

<sup>3</sup> All that apply to each case

<sup>4</sup> Please reference disguising of dates section for details on how to deidentify the dates before submitting any data

Severity of illness refers to the severity of the patient’s respiratory symptoms. The goal is to measure severity on the day of treatment with CP infusion. It is recommended to calculate and store one daily value (midnight to midnight) of severity for the first 10 days of hospitalization. “Case matching” will require use of this value, therefore we need this value to be also calculated on “control” patients.

<b>TABLE 2: SEVERITY OF ILLNESS CATEGORIES FOR MATCHING</b>		
<b>Severity of Illness Allowable Value</b>	<b>Description</b>	<b>Definition</b>
5	Not on supplemental oxygen	Not on supplemental oxygen, on high-flow supplemental oxygen or mechanically ventilated, per definitions below.
4	On conventional supplemental oxygen therapy	On nasal cannula or oxygen facial mask < 30L/min
3	On high-flow supplemental oxygen	- On high-flow nasal cannula (HFNC) or oxygen facial mask $\geq$ 30L/min - Non-invasive positive pressure ventilation (NIPPV), including BiPAP, or CPAP between 8am and 9pm (8am to 9pm requirement on CPAP to rule-out regular home CPAP use)
2	Invasive mechanical ventilation	Mechanical ventilation (as evidenced by PEEP, vent mode change, FiO <sub>2</sub> flowsheet documentation) or ECMO

## Appendix A

These definitions were arrived upon through the work of the COVID-19 Healthcare Coalition partners, including a multidisciplinary group of clinical, informatics and EHR data experts. If you have any questions on these definitions, please contact Rute Martins ([rute@mitre.org](mailto:rute@mitre.org)).

Data Element	Description	Logic
COVID-19 positive	A patient who has been clinically diagnosed with COVID-19 or who tested positive for COVID-19.	COVID-19 confirmed diagnosis <b>OR</b> COVID-19 confirmatory laboratory test
COVID-19-positive laboratory test	A laboratory test indicating that the patient has a COVID-19 infection	Laboratory test in COVID-19 Qualitative Laboratory Test value set AND (laboratory test result ~ detected <b>OR</b> ~positive)
COVID-19 confirmed diagnosis	A clinical diagnosis (any encounter diagnosis, billing diagnosis or problem list entry) of Confirmed COVID-19 infection	Condition in Confirmed COVID-19 Infection value set AND Condition.type ~ (encounter diagnosis, discharge diagnosis, final diagnosis, primary diagnosis, billing diagnosis, problem list entry) AND Condition.type NOT ~ admitting diagnosis
COVID-19-positive date	The earliest date associated with the confirmation of the COVID-19 infection.	Earliest of (COVID-19 confirmatory laboratory test specimen collection date <sup>5</sup> , first COVID-19 confirmed diagnosis)
Age at COVID-19 positive date	The age (in years) of the patient on the date they were diagnosed with COVID-19.	COVID-19 positive date minus date of birth

<sup>5</sup> The result date can be used when specimen collection date is not available



Data Element	Description	Logic
COVID-19-related hospitalization <sup>6</sup>	An encounter for inpatient care that is associated with COVID-19.	Encounter class ~ INP <b>OR</b> ~ inpatient <b>OR</b> ~acute inpatient <b>AND</b> ( COVID-19-positive date during hospitalization <b>OR</b> COVID-19-positive date <= 14 days <u>prior</u> to hospitalization <b>AND</b> respiratory diagnosis associated with hospitalization <sup>7</sup> )
Respiratory diagnosis associated with hospitalization	Any diagnosis for a respiratory condition associated with a hospitalization	Condition code ICD-10-CM J00-J99 <b>AND</b> Condition.type ~ (encounter diagnosis, discharge diagnosis, final diagnosis, chief complaint, primary diagnosis, billing diagnosis, problem list entry)
Convalescent Plasma administration	Convalescent plasma administration regardless of the number of units or antibody titer.	Blood product order LIKE %COVID-19% <b>OR</b> Blood product order administration product code in ISBT 128 E codes for COVID-19 convalescent plasma

<sup>6</sup> It is recognized that this will include COVID-related hospitalizations but also hospitalizations of COVID-19 patients who may be asymptomatic and nosocomial COVID-19 infections.

<sup>7</sup> Proxy for symptomatic COVID-19 infection when patient is admitted for reasons unrelated to COVID-19

## Terminology

**Confirmed COVID-19 Infection Value Set**

Includes	<p>Conditions associated with confirmed COVID-19 infection, including laboratory-confirmed COVID-19 (symptomatic or asymptomatic).</p> <p>ICD-10-CM:  U07.1 only available since 4/1/2020, will be used for lab-confirmed cases regardless of symptom presentation  <del>B97.29 used largely before 4/1/2020.</del></p> <p>SNOMED-CT:  840539006 Disease caused by severe acute respiratory syndrome coronavirus 2 (disorder)</p>
Excludes	ICD-10-CM and SNOMED-CT codes indicative of suspicion or exposure only

**SARS-CoV-2 Laboratory Tests**

Includes	<p>SARS-CoV-2-specific or SARS-like  PCR or NAAT  SARS-CoV-2 RNA in serum/plasma  SARS-CoV-2 panels (not recommended for results by Regenstrief but used in the field)  Qualitative results</p>
Excludes	<p>Human coronavirus tests (non-SARS/SARS-like tests) and MERS tests  Antibody and antigen tests  Quantitative results (e.g. cycle threshold #)</p>

## Appendix B

The following table provide a list of data elements currently identified as potential covariates to support logistic regression and Cox models. These elements will be considered in a second phase of the study.

<b>Data Element Category</b>	<b>Data Element</b>
Patient characteristics	Race
	Ethnicity
	Body Mass Index
	Smoking status
Comorbidities	Pre-hospital diabetes
	Pre-hospital chronic lung disease
	Pre-hospital asthma
	Pre-hospital hypertension
	Pre-hospital cardiovascular disease
	Pre-hospital cerebrovascular disease
	Pre-hospital End Stage Renal Disease (ESRD)
	Pre-hospital chronic kidney disease (CKD) other than ESRD
	Pre-hospital malignancy
Hospitalization	ICU admission
	In-hospital renal impairment
	In-hospital sepsis
	In-hospital multi-organ failure
Laboratory test results <sup>8</sup>	D-dimer
	Full WBC
	pO <sub>2</sub> /FIO <sub>2</sub> and SpO <sub>2</sub> /FIO <sub>2</sub> ratios
	SPO <sub>2</sub>
	CRP
	LDH
	Cardiac troponin
Other therapies administered during hospitalization	Antivirals (Remdesivir)
	Interleukin-6 agents (tocilizumab, sarilumab)
	H <sub>2</sub> -blockers (famotidine)
	Corticosteroids (prednisone, methylprednisolone, dexamethasone)

---

<sup>8</sup> For laboratory values, we will have to determine whether we want all values downloaded or select highs or lows or other fractions of the distributions, since there are likely to be multiple tests, especially with oxygenation.