Convalescent Plasma for Treatment of COVID-19 EHR Data Studies

Hospitals' Technical Guide for EHR Data Retrieval and Submission

Background

Dear Colleagues:

Thank you for assisting in the data collection for this important study during the COVID-19 pandemic. You are part of an unprecedented collaboration of physicians, clinical investigators, informaticists and data scientists working across health systems, academia, and the EHR vendor community to bring the best evidence forward on the use of convalescent plasma for care of patients suffering with COVID-19. This EHR data study is a complement to the Expanded Access to Convalescent Plasma for the Treatment of Patients With COVID-19 study (https://clinicaltrials.gov/ct2/show/NCT04338360), managed by Mayo Clinic (referred herein as the "registry study"), which enrolls only patients receiving convalescent plasma, so there are no patients for a comparator group of patients not exposed to convalescent plasma. The EHR data will allow us to retrospectively approximate a control group using case matching. Using EHR data from a large number of hospitals will allow us to rapidly assemble evidence related to the efficacy and safety of convalescent plasma.

This guide should be used in conjunction with the "Convalescent Plasma for treatment of COVID-19 EHR Data Studies Master Protocol.", which has been included as an amendment to the registry study protocol.

Contact Information

For general information regarding the registry study please contact uscovidplasma@mayo.edu.

For questions on this technical implementation guide please contact convplasma@mitre.org.

Contents

Change History

Version	Date	Author/Owner	Section	Description of Change
1.0	6/ 11/2020	The MITRE Corporation	n/a	Initial version for publication
1.1	6/16/2020	The MITRE Corporation	Background	Added links to registry study protocol and amendment.
			1 Frequently Asked Questions (FAQ)	Updated Cerner contact information on FAQ 1.5 How do I contact my Epic or Cerner EHR vendor for this study? Added FAQ 1.7 What facilities should I include in the reported data? Added FAQ 1.8 What patients should I include in the reported data? Added FAQ 1.9 The details and data definitions included in the Study Protocol and the technical guide are not always aligned. Which definitions should I use?
			2.5 Step 5. De-identify data including disguising dates	Updated date disguising procedure to make use of each individual patient admission date instead of the first CP patient admission to the facility.
			Appendix A – Data definitions	Added admission epoch data element definition. Corrected typo in definition logic for severity of illness.
			Appendix C – Reporting File Structure	Corrected typo in allowable values for age in patient-level data file structure.
1.2	6/17/2020	The MITRE Corporation	1 Frequently Asked Questions (FAQ)	Updated contact and instructions for connecting with Cerner and Epic on FAQ 1.5 How do I contact my Epic or Cerner EHR vendor for this study?

Version	Date	Author/Owner	Section	Description of Change
			2.5 Step 5. De-identify data including disguising dates	Simplified date disguising procedure to define admission date as day 0, since the researchers will no longer need to rely on the date of admission of the first CP patient as an indicator of epoch.
			2.4 Step 4. Extract data for and calculate phase 1 variables	Added instructions to create a deidentified study patient ID.
			Appendix A – Data definitions	Added clarifications on the definition and use of admission epoch data element. Added epoch as a matching variable. Updated Table 1 title to clarify that it includes all phase 1 variables, and not just matching variables.
			Appendix C – Reporting File Structure	Added study patient ID as additional variable in the patient-level data file. Added notes for the transmission of admission date (only required if using old date disguising method) Added positive COVID-19 laboratory test variable to distinguish individuals who were included in the study population by testing vs. clinical diagnosis.
1.3	6/25/2020	The MITRE Corporation	1 Frequently Asked Questions (FAQ)	Clarified that the same pool of patients should be used for Phase 1 and Phase 2 submission in FAQ 1.2 What is meant by “first phase” and “second phase” of data collection? Added FAQ 1.10 What study entry date should be used to ascertain whether a patient should be included in the data set?

Version	Date	Author/Owner	Section	Description of Change
			2 Technical Guidance on Data Collection and Transmission	Removed section 2.1 Step 1. Determine and report first COVID-19 convalescent plasma use at the hospital and date of data extraction as this was intended to provide a sense of epoch, and epoch (the month of CP transfusion) is now included in the list of variables to transmit for each patient. Added section 2.3 Step 3: Determine study entry and exit dates, and date of data extraction, providing further details on how to determine whether a patient belongs in the data set.
			Appendix B - Phase 2 variables	Added discharge disposition to the list of Phase 2 variables.
1.4	10/20/2020	The MITRE Corporation	1 Frequently Asked Questions (FAQ)	Updated FAQs 1.2, 1.6, 1.7, 1.8 and 1.9 with additional information on Phase 2 data submission. Updated FAQ 1.10 to include a cutoff date for hospitalizations to be included in the study for the purposes of ensuring IRB coverage. Updated Figure 1 to more clearly illustrate study entry dates. Added FAQ 1.11 Can I send Phase 1 and Phase 2 data in the same file? Added FAQ 1.12 Do I need to submit additional Phase 1 data in Phase 2? Added FAQ 1.13 What naming conventions should I follow

Version	Date	Author/Owner	Section	Description of Change
			2. Technical Guidance on Data Collection and Transmission	Updated study entry criteria in section 2.1 Phase 1, Step 3. Determine study entry and exit dates to ensure compliance with approved IRB protocol. Added section 2.1 Phase 1, Step 5. Validate data extraction output. Added section 2.2 Phase 2 outlining the steps necessary to successfully submit Phase 2 data.
			Appendix A – Phase 1 Data Definitions, Terminology and File Structure	Consolidated Phase 1 data definitions, terminology, and file structure into a single Appendix (Appendix A). Appendix C previously contained Phase 1 file structure and has been eliminated. Updated definition and logic for COVID-19 related hospitalization to clarify that hospitalizations solely for hospice care and transfers to the health system sending data from another hospital should be excluded from the data.
			Appendix B - Phase 2 Data Definitions and File Structure	Added consolidated data definitions and reporting file structure for Phase 2 data submission.
			Appendix C – Reporting File Structure	Consolidated Appendix C into Appendix A – Phase 1 Data Definitions, Terminology and File Structure.

Version	Date	Author/Owner	Section	Description of Change
1.5	11/5/2020	The MITRE Corporation	1 Frequently Asked Questions (FAQ)	Updated FAQs 1.7 What facilities should I include in the reported data? and 1.8 What patients should I include in the reported data? to clarify relationship between and options for Phase 1 and Phase 2 data transmission. Added FAQ 1.14 I did not save a crosswalk to reidentify study IDs from Phase 1 so I cannot reuse them for Phase 2. What should I do? Added FAQ 1.15 What is the deadline for Phase 2 data submission? Fixed unresolved references and typos.
			2 Technical Guidance on Data Collection and Transmission	Corrected error in example provided in 2.1 Phase 1 Step 6 De-identify data including disguising dates. Fixed unresolved references and typos.

Version	Date	Author/Owner	Section	Description of Change
			Appendix B - Phase 2 Data Definitions and File Structure	**Terminology used in Phase 2 Data Definitions**: emphasized and clarified that terminology (i.e. code lists, value sets) is available and critical to Phase 2 data extraction and how the companion terminology file to this guide can be obtained. **Patient Demographics**: added mapping table for smoking status for organizations using the SNOMED CT smoking status value set outlined in the Common Clinical Data Set. **Convalescent Plasma Administration**: added NULL allowable value where missing and added guidance on how to use it. **Comorbidities:** • Clarified the use of the comorbidities timeframes in the context of data sourced from the problem list vs. billing/encounter diagnoses in Figure 2. • Updated allowable values from 0/1 to TRUE/FALSE for consistency with Phase 1 Boolean variables. **Laboratory Tests:** added NULL allowable value definition. **Other Therapies:** added NULL allowable value and definition.
1.6	11/13/2020	The MITRE Corporation	Appendix B - Phase 2 Data Definitions and File Structure	**Patient Demographics:** added study ID variable, which should be submitting with phase 2 data and match study IDs submitted in Phase 1. **Laboratory Tests:** fixed table formatting issues. **Other Therapies:** fixed table formatting issues.
1.7	1/11/2020	The MITRE Corporation	Appendix B - Phase 2 Data Definitions and File Structure	**Convalescent Plasma Administration**: marked "Total volume of convalescent plasma administered" and "Volume of plasma unit" data elements as optional (do not have to be included in the Phase 2 data submission file). Fixed unresolved references and typos.

Contributors

National COVID-19 Convalescent Plasma Expanded Access Program/National COVID-19 Convalescent Plasma Project

Mayo Clinic

Michael Joyner, MD, Mayo Clinic

Rickey Carter, PhD, Mayo Clinic

Vitaly Herasevich, MD, PhD, Mayo Clinic

Johns Hopkins University

Arturo Casadevall, MD, MS, PhD, Johns Hopkins University

Michigan State University

Chenxi Li, PhD, Michigan State University

Mat Reeves, BVSc, PhD, Michigan State University

Nigel Paneth, MD, MPH, Michigan State University

For more information about the National COVID-10 Convalescent Plasma Expanded Access Program visit http://uscovidplasma.org. For more information about the National COVID-19 Convalescent Plasma Project visit http://ccpp19.org.

COVID-19 Health Care Coalition

Cerner Corporation

Peter Smart, MS, Cerner Corporation

Rehan Waheed, MD, Cerner Corporation

Epic

Andrea Noel, MD, Epic

Janet Campbell, BA, Epic

The MITRE Corporation

Brian Anderson, MD, The MITRE Corporation

Francis X. Campion, MD, The MITRE Corporation

Rute Martins, MS, The MITRE Corporation

For more information about the COVID-19 Healthcare Coalition visit http://c19hcc.org.

1 Frequently Asked Questions (FAQ)

1.1 Which hospitals should provide data?

Every hospital participating in the Convalescent Plasma Expanded Access Program coordinated through the Mayo Clinic may consider participating. Those hospitals with a high volume of patients (>30 convalescent plasma treated patients) and that are on either the Epic EHR or the Cerner EHR are preferred sites for contributing data in rapid fashion.

Details regarding the registry study can be found on the website https://www.uscovidplasma.org/.

1.2 What is meant by "first phase" and "second phase" of data collection?

The protocol mentions that time is of the essence to detect an initial signal regarding the safety of convalescent plasma in COVID-19 patients. For this reason, "phase 1" will focus on the outcomes of progression to mechanical ventilation or death adjusted only in the design, via matching. To more fully validate those findings, the second phase of the study will include additional data elements such as co-morbidities and laboratory test results, to permit further adjustment for potential confounders and biasing factors. The phase 2 data transmission will encompass the same patients included in the phase 1 data set.

1.3 How often should our hospital data set be updated and sent to Mayo Clinic?

Hospitals are asked to send their Phase 1 data as soon possible during June 2020. Hospitals are encouraged to submit Phase 2 data by the end of November 2020. If your organization is unable to submit Phase 2 data by November 30 2020, please contact Breanna Kornatowski at kornatow@msu.edu to discuss your options.

Health systems who have sent data to Mayo Clinic previously are encouraged to send additional Phase 1 data according to the study dates outlined in Section 2.1 Phase 1, Step 3. Determine study entry and exit dates

1.4 Who will perform the matching to controls?

The researchers at Mayo Clinic will perform the matching. We will seek to match at least one and up to four control patients for each patient who has received CP.

1.5 How do I contact my Epic or Cerner EHR vendor for this study?

Both Cerner and Epic have developed detailed queries and data definitions specific for their EHRs in support of this study that can be leveraged by participating hospitals. Hospitals seeking to leverage this assistance should contact their respective vendors per the instructions below.

Instructions for Cerner EHR customers

Please contact convplasma@mitre.org for assistance in reaching the appropriate Cerner contact.

Instructions for Epic customers

You should work with your local IT team to extract data from your EHR (Epic) and transmit it to Mayo/MSU.

- **Step One:** Find a contact in the health IT department. In the same way that you might receive other IT support (for example, a support ticket or a help line), reach out to find "someone who works on the EHR." Remember that your organization may call your EHR something other than "Epic."
    - If you know the CMIO at your health system, that person may be able to help you.
    - It will be important for you to make this connection with IT directly. Your IT team may need to tweak the reports to match how you document in your EHR. They will also be responsible for actually running the reports.
- **Step Two:** When you reach someone on the IT team, tell them "I was told you could help me extract data that Mayo/MSU needs for reporting. They said I should contact you and ask you to contact Epic. They said to mention SLG 5197774."

If you have are having trouble contacting the right individuals within your EHR vendor organization, please contact us at convplasma@mitre.org and we will facilitate an introduction.

1.6 What format should be used for reporting data?

For Phase 1, data will be reported using two .csv files. The first will include hospital-level information, including the names and addresses for each hospital for which data is being collectively reported, as well as high-level information about data extraction. The second file will include patient level data. If your organization has already submitted Phase 1 data, see FAQ 1.12 for additional information on what data you should to submit in Phase 2. Phase 1 Reporting File Structure provides detailed information about the structure and content of these files.

Phase 2 patient-level data will be reported in a separate .csv file and include data for a subset of Phase 1 patients only. Appendix B - Phase 2 Data Definitions and File Structure provides detailed information about the structure and content to include for phase 2 data transmission.

If your organization has not yet submitted Phase 1 data, you can opt to submit Phase 1 and Phase 2 together in the same file (see FAQ 1.11 If I have not submitted Phase 1 data yet, can I send Phase 1 and Phase 2 data in the same file?).

1.7 What facilities should I include in the reported data?

If your organization has transfused plasma at more than one facility, please report the data for each individual facility. The researchers will be performing matching at the individual facility level, as opposed to the health system level. Appendix A – Phase 1 Data Definitions, Terminology and File Structure has information about how to identify each facility and tie it to individual patient data.

Phase 2 patient-level data will be reported in a separate .csv file and include data for all or a subset of Phase 1 patients only. Appendix D – Phase 2 Reporting File Structure provides detailed information about the structure and content to include for phase 2 data transmission. FAQ 1.11 If I have not submitted Phase 1 data yet, can I send Phase 1 and Phase 2 data in the same file? And FAQs 1.12 If I have already submitted Phase 1 data, do I need to submit additional Phase 1 data in Phase 2? Include additional information on Phase 2 data transmission.

1.8 What patients should I include in the reported data?

This study is a case-control study. Unlike the registry study, we are looking for data for both plasma-transfused patients and patients who did not receive convalescent plasma. Section 2 describes the steps to identify the relevant patients. Appendix A includes data definitions that can be useful to ensure consistent data extraction across hospitals. Specific variables to be reported in Phase 1 are included in Appendix A – Phase 1 Data Definitions, Terminology and File Structure. Variables to be reported in Phase 2 are included in Appendix B - Phase 2 Data Definitions and File Structure.

FAQ 1.11 If I have not submitted Phase 1 data yet, can I send Phase 1 and Phase 2 data in the same file? And FAQs 1.12 If I have already submitted Phase 1 data, do I need to submit additional Phase 1 data in Phase 2? Include additional information on Phase 2 data transmission.

1.9 The details and data definitions included in the Study Protocol and the technical guide are not always aligned. Which definitions should I use?

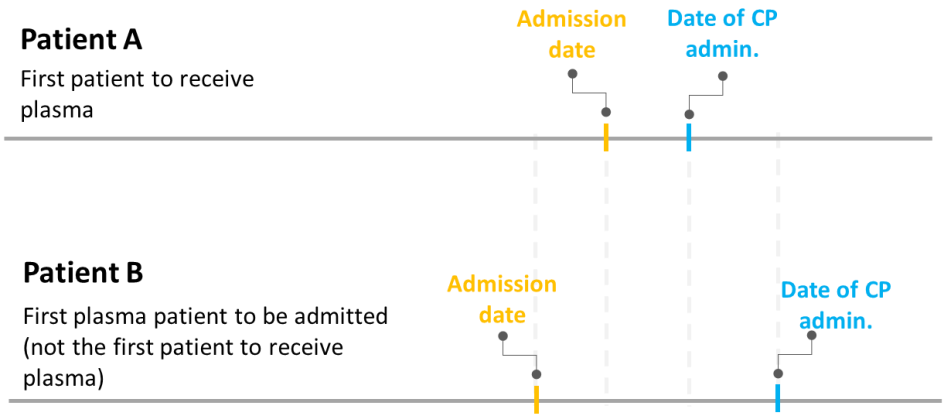
The technical guide includes the most up-to-date definitions that should be used for data extraction, based on iterative discussions with researchers, EHR vendors and participating health systems. However, the variables in the technical guide are the same as those outlined in the protocol for Phase 1 of the study.

For Phase 2, the list of variables that have been included in the original master protocol was notional, and has been revised based on federated data available to EHR vendors, as well an effort/benefit analysis to minimize effort from health systems while retaining data elements critical for analysis. A complete list of Phase 2 variables is provided in Table 3. List of Phase 2 Variables, and detailed definitions for each variable are included in Appendix B - Phase 2 Data Definitions and File Structure.

1.10 What study entry date should be used to ascertain whether a patient should be included in the data set?

There was some ambiguity in the definition of the study start date in the protocol and earlier versions of this technical guide. The date that should be used to ascertain inclusion of *any* patient in the data set is the earliest admission of any patient who received convalescent plasma, regardless of when they received it during the hospital stay. As it turns out, that patient might not be the first to receive convalescent plasma in your hospital. You might have a patient admitted later than the patient described above, but who received convalescent plasma earlier. For example, a patient might be admitted on April 1 and receive plasma on April 7. Another patient might have been admitted on April 3 and received plasma on April 5. The first admission of someone who received plasma was April 1, even though that patient was not the first in your hospital to get plasma, and we would like that admission date – April 1 - to be the point from which you start identifying COVID-19 patients for this study. The diagram below illustrates this distinction.

Figure 1. How to identify the study entry date at your organization. Although patient A received plasma prior to patient B, patient B was actually admitted prior to patient A. Patient B's admission date should be used as the study entry data for all patients (treated and non-treated).



Section 2.1 Phase 1 Step 3 includes more information about study entry and exit dates.

1.11 If I have not submitted Phase 1 data yet, can I send Phase 1 and Phase 2 data in the same file?

If you have not yet submitted Phase 1 data, you're welcome to submit Phase 1 and Phase 2 data together in the same file. Just be sure to follow the file structures provided in Appendix A – Phase 1 Data Definitions, Terminology and File Structure and Appendix B - Phase 2 Data Definitions and File Structure.

If sending Phase 1 and Phase 2 data in a single file, all Phase 1 columns should precede Phase 2 columns.

1.12 If I have already submitted Phase 1 data, do I need to submit additional Phase 1 data in Phase 2?

If you have already submitted Phase 1 data, you are welcome to – but not required – to submit additional patients not included in your initial Phase 1 submission. Additional Phase 1 data should strictly follow the updated guidance provided in section 2.1 Phase 1, Step 3. Determine study entry and exit dates. Please note that no data for patients admitted after August 20, 2020 should be submitted as part of the study to ensure IRB coverage for all the data used in the study.

1.13 What naming conventions should I follow when submitting data?

Your files should indicate 1) the name of organization you are submitting data for (no acronyms), 2) the phase(s) the data refers to, and 3) the data the file was submitted. A few examples:

- HospitalX_Phase1_20201020
- HealthSystemY_Phase1_Phase2_20201019

1.14 I did not save a crosswalk to reidentify study IDs from Phase 1 so I cannot reuse them for Phase 2. What should I do?

If you have submitted Phase 1 data and are unable to reuse the study IDs you submitted for Phase 1 data, we recommend you rerun the Phase 1 query and resubmit Phase 1 data along with Phase 2 data. If you resubmit Phase 1 data, please strictly follow the updated guidance provided in section 2.1 Phase 1, Step 3. Determine study entry and exit dates. **Please note that no data for patients admitted after August 20, 2020 should be submitted as part of the study** to ensure IRB coverage for all the data used in the study.

See FAQ 1.11 for details on submitting Phase 1 and Phase 2 data in the same file.

1.15 What is the deadline for Phase 2 data submission?

Hospitals are encouraged to submit Phase 2 data by the end of November 2020. If your organization is unable to submit Phase 2 data by November 30 2020, please contact Breanna Kornatowski at kornatow@msu.edu to discuss your options.

2 Technical Guidance on Data Collection and Transmission

This section provides information on data extraction, validation and transmission.

2.1 Phase 1

For complete details on Phase 1 variable definitions, please consult Appendix A – Phase 1 Data Definitions.

Step 1. Identify all COVID-19 hospitalized patients

The source population is adult patients (18 years and older) hospitalized for inpatient care associated with COVID-19 infection. This is done to ensure that we have an adequate period of follow up for all subjects. Patients with lesser periods of follow-up now will not be included in the present study but may be studied later. A COVID-19-related hospitalization includes patients who may have been diagnosed up to 14 days prior to the hospitalization, or during the hospitalization, as well as patients who may only have a clinical diagnosis of COVID-19 but no confirmatory laboratory test. We realize that lab testing may have occurred prior to admission and the lab result itself might not reside in the EHR data set from the hospital episode of care. See Appendix A – Phase 1 Data Definitions, Terminology and File Structure for more details on how to identify COVID-19 hospitalized patients.

Step 2. Identify COVID-19 patients who have received convalescent plasma (CP) infusion

The specific implementation of blood bank software and the EHR will determine the most efficient and complete manner to detect COVID-19-convalescent plasma transfusion. Some EHRs use the International Society of Blood Transfusion “E-codes” and others use specific order-entry routines for determining CP transfusion. We are interested in those patients who received at least one unit of COVID-19 CP. For Phase 1, you do not need to report the number of transfused units. You do not need to report antibody titers or content in the transfused unit(s).

Step 3. Determine study entry and exit dates

The population will be restricted to admissions starting between the earliest date of admission for CP transfused patients in each hospital (study entry date) and August 20, 2020. (study exit date). This includes patients who were discharged within this timeframe, as well as patients whose hospitalizations may have extended past August 20, 2020.

Step 4. Extract data for and calculate phase 1 variables

Table 1 identifies the variables needed for Phase 1 data extraction, including matching, outcome and other variables. Detailed definitions for each variable and its supporting data are provided in Appendix A – Phase 1 Data Definitions, Terminology and File Structure.

Table 1. List of Phase 1 Variables

Data Element Category	Data Element
Matching	Hospital name
	Age
	Administrative gender
	Severity of respiratory illness 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)
	Admission epoch
Outcome/time variable	Admission date
	Date of convalescent plasma administration
	Mechanical ventilation start date
	Mechanical ventilation end date
	Discharge date
	Date of death
Other variables	COVID-19 positive laboratory test
	Patient study ID

Severity of illness refers to the severity of the patient's respiratory symptoms. The goal is to measure worst severity on the day of treatment with CP infusion. The logic for calculating worst severity for any given day is provided in Appendix A – Phase 1 Data Definitions.

Severity of respiratory illness is needed for the first 10 days of the hospital stay. Table 2 shows the specific definitions of each of 4 severity levels used in the study. Reporting hospitals should calculate and report just one score for each day. If a patient qualifies for more than one severity level, choose the most severe level experienced during a given calendar day. Daily severity is needed on all patients (those with and without CP infusion) since the intent is to match controls by first day of treatment in the treated population, and that matching cannot be known ahead of time for non-CP patients.

Table 2. Severity of Illness Categories

Severity of Illness Allowable Value	Description	Definition
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 >= 30L/min OR 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, FiO2 flowsheet documentation) or ECMO

Step 5. Validate data extraction output

We recommend conducting validation of the extracted data using a small random sample of 5% of each cohort (control pool and CP-infused patients) and validate the automated data extraction output against visual inspection of the patient's record. This validation step will provide a quality control mechanism, ensuring the completeness and accuracy of the transmitted data.

If there is non-concordance in any data element, you should identify the root cause for the discrepancy and work with your IT team and, if appropriate, your EHR technical support to make adjustments to the data extraction code, if necessary.

Step 6. De-identify data including disguising dates

Review data to insure there are no patient identifiers included in the data set. There should be no patient names, DOB, MRNs, insurance numbers, phone numbers, home zip codes, home addresses, email addresses, SSN's or other personal identifiers in the data set.

Additionally, we ask hospitals to report disguised dates using an integer. The patient's admission date should be assigned the number zero (0). All subsequent dates should be computed as days from admission date. For example, if the patient was admitted on March 25th, was intubated on March 29th, received CP on April 2nd and was extubated on April 5th, and discharged on April 10th, those dates would be recorded as 4 (0+4), 8, 11, and 16.

Note: if your organization has already implemented the date disguising method described in the study protocol and previous versions of this guide, you don't need to implement this new method. However, you will need to report the random number used to disguise the admission date so that the researchers are able to calculate the time intervals (measured in days) needed in the study.

Step 7. Assign a study ID to each patient

To facilitate future data updates, including corrections and the transmission of phase 2 variables, hospitals should assign each patient should be assign a randomly generated unique identifier (cannot contain any personally identifiable information). Hospitals should keep a record of the crosswalk between the patient's medical record number and their study identifiers for future reference.

Step 8. Create Phase 1 CSV/TSV data set

Convert your data file to a CSV or TSV format ready for transmission. See Appendix A – Phase 1 Data Definitions, Terminology and File Structure.

Step 9. Transmit Phase 1 data to Mayo Clinic

Mayo will be the data aggregation center for the study. Participating hospitals can upload their data using secure SharePoint. Mayo will share a link to the appropriate SharePoint website.

When submitting Phase 2 data, please follow the file naming conventions described in FAQ 1.13 What naming conventions should I follow

2.2 Phase 2

This section outlines the steps to extract data for Phase 2 data submission. For complete details on Phase 1 variable definitions, please consult Appendix B - Phase 2 Data Definitions and File Structure.

Step 1. Identify patient population for Phase 2 data extraction

Phase 2 data will only be required for a subset of the population your health system submits to the research team. Phase 1 patient data will be used to create a matched cohort from the pool of potential control patients, with a goal to have at least one matched patient – but more, if possible – per CP-treated patient. This will result in Phase 2 data being needed only for CP patients for which it was possible to identify a matched control and control patients who were matched to a CP patient.

The research team will provide you with a list of patient study IDs for whom you should extract and submit Phase 2 data.

Step 2. Extract data for and calculate Phase 2 variables

Table 3 identifies the variables for Phase 2 data extraction, which consist mostly of variables that can be used to conduct sub-analyses and possibly adjust the results of Phase 1 data analysis. Detailed definitions for each variable and its supporting data are provided in Appendix B - Phase 2 Data Definitions and File Structure.

Table 3. List of Phase 2 Variables

Data Element Category	Data Element
Patient characteristics	Race
	Ethnicity
	Body Mass Index

Data Element Category	Data Element
	Weight
	Height
	Smoking status
Convalescent Plasma Administration	Total volume of convalescent plasma transfused
	Total number of units of convalescent plasma transfused
	Date of plasma unit administration (for units 1-5)
	Volume of plasma unit (for units 1-5)
Pre-hospital comorbidities	Asthma
	COPD
	Hypertension
	Heart failure
	Cancer
	Cardiac arrythmias
	Cerebrovascular disease
	Coronary artery disease
	Diabetes
	Obesity
	Chronic kidney disease (CKD) Stage 1 through 4
	Chronic kidney disease (CKD) Stage 5 and ESRD
	Chronic liver disease
Hospitalization	Discharge to home
Laboratory test results on day of admission or immediately preceding ED visit (result, units and reference range)	AST result value
	AST result units
	AST result reference range
	Cardiac troponin result value
	Cardiac troponin result units
	Cardiac troponin result reference range
	Creatinine result value
	Creatinine result units
	Creatinine result reference range
Date of first administration for other therapies administered during the hospitalization)	Angiotensin-converting enzyme (ACE) inhibitors
	Angiotensin II receptor blockers (ARBs)
	Azithromycin

Data Element Category	Data Element
	Dexamethasone
	Famotidine
	Systemic Corticosteroids other than Dexamethasone (Hydrocortisone, Methylprednisolone, Prednisone)
	Hydroxychloroquine
	Interleukin-6 inhibitors (tocilizumab, sarilumab)
	Remdesivir

Step 3. Validate data extraction output and deidentify dates

Similarly to Phase 1 data, follow the steps described in section 2.1 Phase 1 to validate the extraction output, disguise dates, as described in Step 5 and Step 6, respectively.

Step 4. Assign Study ID

The study ID transmitted along with Phase 2 data should match the ID sent along with Phase 1 data for the same patient.

Step 5. Submit data to Mayo Clinic through secure SharePoint site

Phase 2 data should be submitted to the same SharePoint site used to submit Phase 1 data. When submitting Phase 2 data, please follow the file naming conventions described in FAQ 1.13. What naming conventions should I follow when submitting data?

Appendix A – Phase 1 Data Definitions, Terminology and File Structure

Phase 1 Data Definitions

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. Vendor-specific implementations of these definitions may be available from your EHR vendor. If you have any questions or feedback on these definitions, please reach out to convplasma@mitre.org.

Data Element	Description	Logic
Administrative gender	The gender of the patient used for administrative purposes, typically the patient's sex as specified in their legal documents.	n/a
Admission date	The date the patient was admitted for inpatient care.	n/a
Admission epoch	The month and year the patient was admitted for inpatient care for the COVID-19 related hospitalization. Admission epoch is intended to provide an indication of when the patient was hospitalized over the course of the pandemic.	n/a

Data Element	Description	Logic
Age at COVID-19 positive date	The age (in years) of the patient on the date they were diagnosed with COVID-19.	IF COVID-19 positive date - date of birth> 90, ">90" ELSE COVID-19 positive date – date of birth
Convalescent plasma administration	Convalescent plasma administration regardless of the number of units or antibody titer.	Blood product order LIKE %COVID-19% OR Blood product order administration product code in ISBT 128 E codes for COVID-19 convalescent plasma)
Convalescent plasma administration date	The first date when the patient was administered convalescent plasma during the hospitalization.	Earliest Convalescent plasma administration
COVID-19 confirmed diagnosis	A clinical diagnosis (any encounter diagnosis, billing diagnosis or problem list entry) of Confirmed COVID-19 infection. See Terminology section below for inclusion and exclusion criteria for COVID-19 confirmed infection below.	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	A patient who has been clinically diagnosed with COVID-19 or who tested positive for COVID-19.	COVID-19 confirmed diagnosis **OR** COVID-19 confirmatory laboratory test

Data Element	Description	Logic
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[1], first COVID-19 confirmed diagnosis)
COVID-19-positive laboratory test	A laboratory test indicating that the patient has a COVID-19 infection. See Terminology section below for inclusion and exclusion criteria for COVID-19 laboratory tests.	Laboratory test in COVID-19 Qualitative Laboratory Test value set AND (laboratory test result ~ detected **OR** ~positive)
COVID-19-related hospitalization	An encounter for inpatient care that is associated with COVID-19. If a patient is diagnosed with COVID-19 prior to the hospitalization, a respiratory diagnosis is used as a proxy for symptomatic COVID-19 infection when patient is admitted for reasons unrelated to COVID-19. It is recognized that this definition may include COVID-related hospitalizations but also hospitalizations of COVID-19 patients who may be asymptomatic, and nosocomial COVID-19 infections. Admissions exclusively for hospice are or patients who were transferred in from another hospital (external to the health system) should be excluded,	Encounter class ~ INP **OR** ~ inpatient **OR** ~acute inpatient AND (COVID-19-positive date <= 7 days after hospital admission OR COVID-19-positive date <= 14 days prior to hospitalization **AND** respiratory diagnosis associated with hospitalization) AND NOT Encounter type ~ hospice AND NOT Encounter origin ~ acute care hospital

[1] The result date can be used when specimen collection date is not available

Data Element	Description	Logic
Date of death	The date the patient expired. It is only expected that deaths during hospitalization are recorded.	IF discharge status LIKE %deceased% OR %expired% (or similar) then discharge date OR IF vital status LIKE %deceased% or %expired% (or similar) then date of death
Discharge date	The date the patient was discharged from inpatient care. The discharge date is expected to be the same as the death date for deceased patients.	n/a
Mechanical ventilation start date	The first date the patient was mechanically ventilated. See definition of “On invasive mechanical ventilation” for more details.	Earliest (vent mode OR vent activity) indicative of invasive mechanical ventilation.
Mechanical ventilation end date	The last date the patient was mechanically ventilated following the first intubation. See definition of “On invasive mechanical ventilation” for more details.	Most recent (vent more OR vent activity) NOT LIKE ~ CPAP% OR ~ %BiPAP% OR ~ %High Flow Nasal Canula% for 1st intubation
Respiratory diagnosis associated with hospitalization	Any diagnosis for a respiratory condition associated with a hospitalization	Condition code ICD-10-CM J00-J99 AND Condition.type ~ (encounter diagnosis, discharge diagnosis, final diagnosis, chief complaint, primary diagnosis, billing diagnosis, problem list entry)

Data Element	Description	Logic
Not on supplemental oxygen	Indication that the patient did not receive any form of supplemental oxygen on a given hospital day (midnight to midnight).	**NOT** on invasive mechanical ventilation **AND NOT** on high flow invasive mechanical ventilation **AND NOT** on invasive mechanical ventilation
On conventional supplemental oxygen therapy	Evidence that the patient received conventional oxygen therapy on a given hospital day (midnight to midnight)	On nasal cannula or oxygen facial mask < 30L/min
On high-flow supplemental oxygen	Evidence that the patient received high-flow supplemental oxygen on a given hospital day (midnight to midnight). CPAP use between 8am and 9pm is excluded to rule out regular home CPAP use.	(On high-flow nasal cannula (HFNC) or oxygen facial mask >= 30L/min **OR** Non-invasive positive pressure ventilation (NIPPV), including BiPAP **OR** CPAP (between 8am and 9pm) **AND**
On invasive mechanical ventilation	Evidence that the patient was mechanically ventilated on a given hospital day (midnight to midnight). See Terminology section below for inclusion and exclusion criteria for invasive mechanical ventilation.	n/a

Data Element	Description	Logic
Severity of respiratory illness	The worst severity of the patient's respiratory symptoms on a given hospitalization day (with the day of admission as hospital day 0), as evidenced by the patient's oxygenation requirements. A day is considered an individual date, i.e. midnight to midnight.	IF On invasive mechanical ventilation THEN 2 ELSE IF On high-flow supplemental oxygen THEN ELSE IF On conventional supplemental oxygen therapy THEN 4 ELSE IF Not on conventional supplemental oxygen THEN 5

Terminology

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

SARS-CoV-2 Laboratory Tests	
Includes	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

Excludes	Human coronavirus tests (non-SARS/SARS-like tests) and MERS tests Antibody and antigen tests Quantitative results (e.g. cycle threshold #)

Invasive Mechanical Ventilation

Includes	• ICU flowsheet documentation of vent mode or vent activity, evidence of positive end-expiratory pressure (PEEP) documentation. • Mechanical ventilation-associated procedure performed (CPT 94002 or 94003) – if flowsheet documentation isn't reliable • Extracorporeal membrane oxygenation (ECMO)
Excludes	• Intubation procedure • High-flow oxygen delivered through nasal cannula, BiPAP, or CPAP • Sedation meds without evidence of ventilator management • Conditions indicative of respiratory failure without evidence of ventilator management

Phase 1 Reporting File Structure

Hospital Level Data

This file should include information about the individual facilities where convalescent plasma has been administered. Data should not be aggregated at a health system level, because matching will be done at the facility level.

VARIABLE CODE	VARIABLE NAME	DATA TYPE	FORMAT/ALLOWABLE VALUES	OCCURS	NOTES FOR EXTRACTION
hospital	Hospital name	string	Any valid string	1-20	Each individual facility should be listed, as matching will occur at the facility level.
address	Hospital address	String	Any valid string	1-20	The address for each facility.
date_first_cp	First plasma patient	Date	YYYY-MM	1-20	The month associated the first CP patient treated at an individual facility was admitted.
date_extraction	Data extraction date	Date	YYYY-MM-DD	1	The date the data was extracted from the EHR.

Phase 1 Patient-level Data

This file should include variables for Phase 1, for all patients identified in Phase 1 Identify all COVID-19 hospitalized patients. This includes both patients who were treated with convalescent plasma as well as those who were not. Please consult the data definitions in Appendix A – Phase 1 Data Definitions for detailed definitions and inputs for each reported variable.

Variable Code	Variable Name	Data Type	Allowable Values	Occurs	Notes for Extraction
hospital	Hospital name	string	Any valid string	1	The name of the hospital where the COVID-19 hospitalization occurred. This should match the hospital name included in the hospital level data file.
id	Patient study ID	Integer	Any valid integer	1	Instructions on how to generate a patient study ID are available in section 2.1 Phase 1, Step 7: Assign a study ID to each patient.
age	Age	Integer OR string	18-89, “>90”	1	For purposes of deidentification, if patient age is >90, populate ">90" instead of the numeric age
admin_gender	Administrative gender	string	M = male F = female U = unknown/other	1	
severity_day[x]	Severity of respiratory illness	Integer	5 4 3 2	10	Consult Appendix A – Phase 1 Data Definitions for definitions of each allowable value.
admission_epoch	Admission epoch	Date	YYYY-MM	1	

Variable Code	Variable Name	Data Type	Allowable Values	Occurs	Notes for Extraction
date_admission	Admission date	Integer	any valid integer	1	Only needs to be reported if using the date disguising method described in the study protocol or earlier versions of this guide.
covid_positive_test	COVID-19-positive laboratory test	Boolean	TRUE = positive test result FALSE = negative test result NULL = patient was not tested, result not available (e.g. pending) or unknown (e.g. inconclusive, invalid, equivocal, indeterminate)	0 -1	This variable is not mandatory.
date_cp_admin	Date of convalescent plasma administration	boolean or integer	NULL = not administered convalescent plasma any valid integer = disguised date	1	Follow the instructions on section 2.1 Phase 1, Step 6 De-identify data including disguising dates.
start_date_vent	Mechanical ventilation start date	boolean or integer	NULL = not mechanically ventilated at any point from admission until discharge or data extraction date (if still hospitalized) any valid integer = disguised date	1	Follow the instructions on section 2.1 Phase 1, Step 6 De-identify data including disguising dates.

Variable Code	Variable Name	Data Type	Allowable Values	Occurs	Notes for Extraction
end_date_vent	Mechanical ventilation end date	boolean or integer	NULL = never mechanically ventilated or still mechanically ventilated any valid integer = disguised date	1	If start_date_vent = integer AND end_date_vent = FALSE, then the patient will be considered to be mechanically ventilated at the time of data extraction. If start_date_vent = FALSE and end_date_vent = FALSE, then the patient will be considered to never have been mechanically ventilated from admission to the date of data extraction.
date_death	Date of death	boolean or integer	NULL = not deceased as of date of data extraction any valid integer = disguised date	1	Follow the instructions on section 2.1 Phase 1, Step 6 De-identify data including disguising dates.
date_discharge	Date of discharge	boolean or integer	NULL = patient is currently hospitalized any valid integer = disguised date	1	Note this variable subsumes the "hospitalization at the time of data extraction" variable as allowable value FALSE.

Appendix B - Phase 2 Data Definitions and File Structure

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. Vendor-specific implementations of these definitions may be available from your EHR vendor. If you have any questions or feedback on these definitions, please reach out to convplasma@mitre.org.

These definitions should be used to produce Phase 2 variables. Please note the variable definitions are grouped by category for convenience only. All Phase 2 variables should be submitted in a single .csv file.

Terminology used in Phase 2 Data Definitions

Terminology (i.e., code lists or value sets) for comorbidities (ICD-10-CM and SNOMED CT), laboratory tests (LOINC) and other therapies (RxNorm) will be provided in a companion .csv file. Please contact kornatow@msu.edu or convplasma@mitre.org to obtain this companion file as it is critical to the consistent extraction of data across organizations.

Please note that each variable may reference more that one code list or value set in its calculation logic. Also note that there is specific guidance in the Comorbidities section, Laboratory Tests section, and Other Therapies section on how the code lists should be used, including how to use the code lists in the logic for variable calculation. Code lists are referenced in "logic and notes for extraction" as ***bold italicized text***).

Patient Demographics

VARIABLE NAME (variable code)	DEFINITION	DATA TYPE	FORMAT/ALLOWABLE VALUES	OCCURS	LOGIC AND NOTES FOR EXTRACTION
Study ID (id)	A unique identifier for the patient.	Integer	Any valid integer	1	The study ID submitted with Phase 2 data should be the same as the study ID submitted with Phase 1 data for the same patient. See section 2.2 Phase 2 Step 4 Assign Study ID for more details. If submitting Phase 1 and Phase 2 data in the same file, you only need to include the study ID once.

VARIABLE NAME (variable code)	DEFINITION	DATA TYPE	FORMAT/ALLOWABLE VALUES	OCCURS	LOGIC AND NOTES FOR EXTRACTION
Race (race)	The race of the patient.	string	1002-5 American Indian or Alaska Native 2028-9 Asian 2054-5 Black or African American 2076-8 Native Hawaiian or Other Pacific Islander 2131-1 Other Race 2106-3 White NULL = unknown or missing	1	IF multiracial THEN "2131-1"
Ethnicity (ethnicity)	The ethnicity of the patient.	string	2135-2 Hispanic or Latino 2186-5 Not Hispanic or Latino NULL = unknown or missing	1	
Body Mass Index (bmi)	The most recently documented body mass index within the two years preceding and up to the COVID-19 related hospitalization start date.	decimal	0.0 to 100.0 NULL = unknown or missing in the timeframe of interest	1	Most recent BMI documented <= 2 years before or on hospitalization start date.
Weight (weight)	The most recently documented patient's weight (in kg), within the two years preceding and up to the COVID-19 related hospitalization start date.	decimal	40.0 to 300.0 NULL = unknown or missing in the timeframe of interest	1	Most recent weight documented <= 2 years before or on hospitalization start date.
Height (height)	The most recently documented patient height (in cm).	integer	100 to 250 NULL = unknown or missing	1	Most recent height available.
Smoking status (smoking_status)	The most recently documented patient status as it related to the use of smoking tobacco.	string	current smoker former smoker never smoker NULL = unknown or missing	1	Most recent smoking status documented on or before discharge date. If you are using the SNOMED CT smoking status value set outlined in the Common Clinical Data Set, please see Table 4. Mapping

VARIABLE NAME (variable code)	DEFINITION	DATA TYPE	FORMAT/ALLOWABLE VALUES	OCCURS	LOGIC AND NOTES FOR EXTRACTION
					for smoking status for mapping to this variable's allowable values.

Table 4. Mapping for smoking status

Common Clinical Data Set (CCDS) Smoking Status Allowable Values (SNOMED CT codes)	CP EHR Study Allowable Value for Smoking Status Variable
449868002 | Current every day smoker	Current smoker
428041000124106 | Current some day smoker	Current smoker
8517006 | Former smoker	Former smoker
266919005 | Never smoker	Never smoker
266919005 | Smoker, current status unknown	Current smoker
266927001 | Unknown if ever smoked	NULL
428071000124103 | Heavy tobacco smoker	Current smoker
428061000124105 | Light tobacco smoker	Current smoker

Convalescent Plasma Administration

Data elements related to plasma volume are optional and the corresponding columns can be omitted in the reporting file.

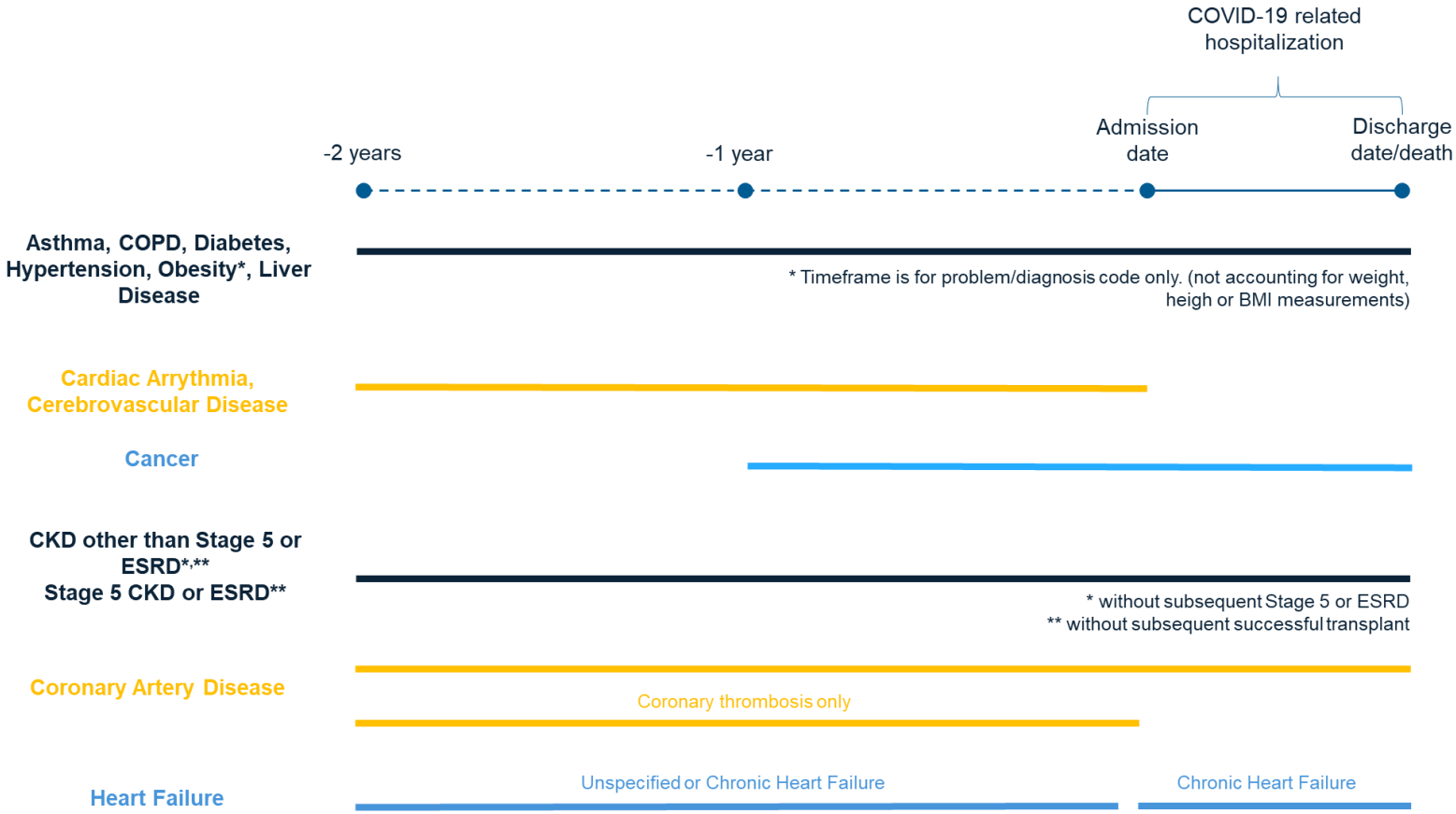
DATA ELEMENT	DEFINITION	DATA TYPE	FORMAT/ALLOWABLE VALUES	OCCURS	LOGIC AND NOTES FOR EXTRACTION
Total number of units of plasma administered (total_cp_units)	The number of units of COVID-19 convalescent plasma administered to the patient over the course of the COVID-19 related hospitalization.	Integer	0-20 NULL = unknown or missing	1	Include all units administered over the entire hospitalization, regardless of whether they were administered in a single or multiple dates or transfusion events. Data source should be blood product administration records as opposed to orders.
Total volume of plasma administered (total_cp_vol)	Total volume, in milliliters (mL), of COVID-19 convalescent plasma administered to the patient over the course of the entire COVID-19 related hospitalization.	Integer	0-6000 NULL = unknown or missing	0-1	Add the volume of all units administered to the patient, regardless of whether they were administered in a single or multiple dates or transfusion events. Data source should be blood product administration records as opposed to orders. ***This data element is optional and may be omitted from the reporting file.***
Date of plasma unit administration (date_cp_unit[x])	The date on which the patient was administered the n^{th} (1-5) individual unit of COVID-19 convalescent plasma during their COVID-19 related hospitalization.	Integer	any valid integer = disguised date NULL = not administered n^{th} unit of convalescent plasma, unknown or missing	5	Follow the instructions on section 2.1 Phase 1, Step 6 De-identify data including disguising dates. If total number of units/volume = 0, populate all as NULL.

DATA ELEMENT	DEFINITION	DATA TYPE	FORMAT/ALLOWABLE VALUES	OCCURS	LOGIC AND NOTES FOR EXTRACTION
Plasma unit volume (vol_cp_unit[x])	The volume, in milliliters (mL) of the n^{th} (1-5) individual unit of COVID-19 convalescent plasma the patient received during their COVID-19 related hospitalization.	Integer	0-1000 NULL = not administered nth unit of convalescent plasma, unknown or missing	0-5	This should include the volume of an individual plasma unit only. If total number of units/volume = 0, populate all as NULL. ***These data elements are optional and may be omitted from the reporting file.***

Comorbidities

Comorbidity data sources include the problem list as well as encounter and billing diagnoses. Please note that the timeframe applicable to each comorbidity may vary and may or may not include problems and/or diagnoses documented as part of the COVID-19 related hospitalization. In addition, when "hospitalization start date" and "discharge date" are mentioned in the logic, they are always in reference to the COVID-19 related hospitalization. A summary of the timeframe of interest for each comorbidity is provided in Figure 2.

Figure 2. Timeframes of interest for comorbidity variables. Determining whether a patient has a comorbidity based on active problem list entries is dependent only on the status of the entry at the end of the timeframe of interest, irrespective of the start/onset date of the problem. Please see logic and notes for extraction for individual comorbidities for more details.



DATA ELEMENT	DEFINITION	DATA TYPE	ALLOWABLE VALUES	OCCURS	LOGIC AND NOTES FOR EXTRACTION[2]
Asthma (asthma)	Indicator for whether the patient had pre-existing COPD at the time of the COVID-19 related hospitalization.	Boolean	TRUE FALSE	1	Active problem list entry of [***Asthma, COPD, Diabetes, Hypertension, Obesity, Chronic Liver Disease***] on discharge date[3] **OR** Billing or encounter diagnosis of [***Asthma, COPD, Diabetes, Hypertension, Obesity, Chronic Liver Disease***] <= 2 years prior to or on discharge date
COPD (copd)	Indicator for whether the patient had pre-existing COPD at the time of the COVID-19 related hospitalization.	Boolean	TRUE FALSE	1	
Hypertension (htn)	Indicator for whether the patient had pre-existing hypertension at the time of the COVID-19 related hospitalization.	Boolean	TRUE FALSE	1	
Diabetes (diabetes)	Indicator for whether the patient had pre-existing diabetes at the time of the COVID-19 related hospitalization.	Boolean	TRUE FALSE	1	
Obesity (obesity)	Indicator for whether the patient had an obesity condition or diagnosis at the time of the COVID-19 related hospitalization.	Boolean	TRUE FALSE	1	
Liver disease (liver_disease)	Indicator for whether the patient had pre-existing chronic liver disease at the time of the COVID-19 related hospitalization.	Boolean	TRUE FALSE	1	
Cancer (cancer)	Indicator for whether the patient had cancer at the time of the COVID-19 related hospitalization.	Boolean	TRUE FALSE	1	Active problem list entry of ***Cancer*** on discharge date[3] **OR** Billing diagnosis OR encounter diagnosis of ***Cancer*** <= 1 year prior to or on discharge date

[2] See introductory notes for extraction at the top of the Comorbidities section for guidance applicable to all comorbidities.

[3] Problem list entry must be active as of the hospitalization discharge or death date. The start/onset date of the problem could be anytime before or even on the discharge/death date and the problem must either 1) remain unresolved (i.e. resolved date = NULL) or have a resolved date > discharge date.

DATA ELEMENT	DEFINITION	DATA TYPE	ALLOWABLE VALUES	OCCURS	LOGIC AND NOTES FOR EXTRACTION[2]
Chronic kidney disease (CKD) Stage 1 through 4 (ckd_stg1_4)	Indicator for whether the patient had pre-existing chronic kidney disease stage 1-4 at the time of the COVID-19 related hospitalization.	Boolean	TRUE FALSE	1	Active problem list entry of ***CKD Stage 1-4*** on discharge date[3] OR Billing or encounter diagnosis of ***CKD Stage 1-4*** <= 2 years prior to or on discharge date AND NOT Active problem list entry of ***CKD Stage 5 and ESRD*** on discharge date[3] OR Billing or encounter diagnosis of ***CKD Stage 5 and ESRD*** date after billing or encounter diagnosis of ***CKD Stage 1-4*** date OR ***Kidney Transplant*** on or before discharge date AND NOT ***Kidney Transplant Failure*** on or before discharge date AND after ***Kidney Transplant***
Chronic kidney disease (CKD) Stage 5 and ESRD (ckd_stg5_esrd)	Indicator for whether the patient had pre-existing chronic kidney disease stage 5 or end-stage renal disease at the time of the COVID-19 related hospitalization.	Boolean	TRUE FALSE	1	Active problem list entry of ***CKD Stage 5 and ESRD*** on discharge date[3] OR Billing or encounter diagnosis of ***CKD Stage 5 and ESRD*** on or before discharge date AND NOT ***Kidney Transplant*** on or before discharge AND NOT ***Kidney Transplant Failure*** on or before discharge date AND after ***Kidney Transplant***

DATA ELEMENT	DEFINITION	DATA TYPE	ALLOWABLE VALUES	OCCURS	LOGIC AND NOTES FOR EXTRACTION[2]
Cardiac arrythmia (cardiac_arryth)	Indicator for whether the patient had pre-existing cardiac arrythmia at the time of the COVID-19 related hospitalization.	Boolean	TRUE FALSE	1	Active problem list entry of [***Cardiac Arrythmia, Cerebrovascular Disease***] before hospitalization start date[4] OR Billing or encounter diagnosis of [***Cardiac Arrythmia, Cerebrovascular Disease***] <= 2 years prior to hospitalization start date
Cerebrovascular Disease (cvd)	Indicator for whether the patient had pre-existing cerebrovascular disease at the time of the COVID-19 related hospitalization.	Boolean	TRUE FALSE	1	
Coronary Artery Disease (cad)	Indicator for whether the patient had pre-existing coronary artery disease at the time of the COVID-19 related hospitalization.	Boolean	TRUE FALSE	1	Active problem list entry of ***Stable Ischemic Heart Disease*** on discharge date[3] OR Active problem list entry of ***Coronary Thrombosis*** before hospitalization start date[4] OR Billing or encounter diagnosis of ***Stable Ischemic Heart Disease*** on or before discharge date OR ***Coronary Thrombosis*** with start date <= 2 years before hospitalization start date

[4] Problem list entry must be active *before* the related to the hospitalization start date. The start/onset date of the problem could be anytime before the hospitalization start date and the problem must either 1) remain unresolved (i.e. resolved date = NULL) or have a resolved date >= hospitalization start date.

DATA ELEMENT	DEFINITION	DATA TYPE	ALLOWABLE VALUES	OCCURS	LOGIC AND NOTES FOR EXTRACTION[2]
Heart Failure (hf)	Indicator for whether the patient had pre-existing chronic heart failure at the time of the COVID-19 related hospitalization.	Boolean	TRUE FALSE	1	Active problem list entry of ***Chronic Heart Failure*** OR ***Exacerbation of Heart Failure*** on discharge date[3] OR Billing or encounter diagnosis of ***Chronic Heart Failure*** OR ***Exacerbation of Heart Failure*** with start date <= 2 years on or before discharge date OR Active problem list entry of ***Unspecified Heart Failure*** before hospitalization start date[4] OR Billing or encounter diagnosis of ***Unspecified Heart Failure*** on or before COVID-19 related hospitalization start date

Hospitalization

DATA ELEMENT	DEFINITION	DATA TYPE	ALLOWABLE VALUES	OCCURS	LOGIC AND NOTES FOR EXTRACTION
Discharge to Home (disch_home)	An indicator of whether the patient was discharge to their home.	Boolean	TRUE FALSE NULL = unknown or missing	1	Discharge disposition LIKE %home% AND NOT LIKE %hospice% Note: A patient's home may be, in certain cases, a healthcare facility (e.g. a patient living in a skilled nursing facility (SNF) prior to hospitalization was discharged back to the SNF).

Laboratory Tests

For each laboratory test, the result value, the result units and the reference range should be reported.

LOINC codes for each laboratory test are provided in a companion terminology file. While you may use the LOINC codes provided to facilitate the identification of the appropriate laboratory tests, it is important to note that their accuracy and completeness is dependent on your organization's available mappings to LOINC. We recommend you consider locally used terminologies and/or validation techniques to ensure the correct laboratory tests are extracted.

VARIABLE (VARIABLE CODE)	DEFINITION	DATA TYPE	ALLOWABLE VALUES	OCCURS	LOGIC AND NOTES FOR EXTRACTION[5]
Aspartate transaminase result value (ast_result)	The first aspartate transaminase laboratory test result available anytime between the COVID-19 related hospitalization start date and the admission date.	String	Any valid string NULL = no test done or no result available in the timeframe of interest	1	First RESULT value for ***Aspartate Transaminase*** with date >= hospitalization start date AND <= hospitalization admission date[6].
Aspartate transaminase result units (ast_units)	The units associated with the first result value reported for aspartate transaminase.	String	Any valid string NULL = no test done or no result available in the timeframe of interest	1	N/A
Aspartate transaminase result reference range (ast_ref_range)	The reference range associated with the first result value reported for aspartate transaminase.	String	Any valid string NULL = no test done or no result available in the timeframe of interest	1	N/A

[5] See introductory notes for extraction at the top of section Laboratory Tests for guidance applicable to all laboratory tests of interest.

[6] The hospitalization encounter start date may precede the admission date (e.g., when a patient comes in through the emergency department).

VARIABLE (VARIABLE CODE)	DEFINITION	DATA TYPE	ALLOWABLE VALUES	OCCURS	LOGIC AND NOTES FOR EXTRACTION[5]
Cardiac troponin result value (tropon_result)	The first cardiac troponin laboratory test result available anytime between the COVID-19 related hospitalization start date and the admission date.	String	Any valid string NULL = no test done or no result available in the timeframe of interest	1	First RESULT value for ***Cardiac Troponin*** with date >= hospitalization start date AND <= hospitalization admission date[6].
Cardiac troponin result units (tropon_units)	The units associated with the first result value reported for cardiac troponin.	String	Any valid string NULL = no test done or no result available in the timeframe of interest	1	N/A
Cardiac troponin result reference range (tropon_ref_range)	The reference range associated with the first result value reported for cardiac troponin.	String	Any valid string NULL = no test done or no result available in the timeframe of interest	1	N/A
Creatinine result value (creat_result)	The first cardiac troponin laboratory test result available anytime between the COVID-19 related hospitalization start date and the admission date.	String	Any valid string NULL = no test done or no result available in the timeframe of interest	1	First RESULT value for ***Creatinine*** with date >= hospitalization start date AND <= hospitalization admission date[6].
Creatinine result units (creat_units)	The units associated with the first result value reported for cardiac troponin.	String	Any valid string NULL = no test done or no result available in the timeframe of interest	1	N/A
Creatinine result reference range (creat_ref_range)	The reference range associated with the first result value reported for cardiac troponin.	String	Any valid string NULL = no test done or no result available in the timeframe of interest	1	N/A

Other Therapies

This section defines therapies other than convalescent plasma that may have been administered during the COVID-19 related hospitalization. Each therapy variable should be reported as the first (disguised) date there is evidence that the medication was administered to the patient over the course of the hospitalization.

Lists of RxNorm codes are provided in a companion document to help you identify specific medications. In certain cases, you may also use therapeutic classes available in your local drug database (e.g. Medispan or First DataBank). Additional constraints on administration route will also need to be applied for certain medications, as noted.

DATA ELEMENT	DEFINITION	DATA TYPE	ALLOWABLE VALUES	OCCURS	LOGIC AND NOTES FOR EXTRACTION[7]
Angiotensin-converting enzyme (ACE) inhibitors (ace_inhibitors)	The first date during the COVID-19 related hospitalization when an ACE inhibitor was administered to the patient.	Integer	Any valid integer NULL = unknown, missing or medication not administered during inpatient stay	1	FIRST date >= admission date AND <= discharge date when Medication administered in ***ACE Inhibitors*** OR drug therapeutic class ~ ACE inhibitor OR simple generic name LIKE *pril AND NOT medication administered route ~ ophthalmic OR dermatological Follow the instructions on section 2.1 Phase 1, Step 6 De-identify data including disguising dates.
Angiotensin II receptor blockers (ARBs) (arbs)	The first date during the COVID-19 related hospitalization when an ARB was administered to the patient.	Integer	Any valid integer NULL = unknown, missing or medication not administered during inpatient stay	1	FIRST date >= admission date AND <= discharge date when Medication administered in ***ARBs*** OR drug therapeutic class ~ ARB OR OR simple generic name LIKE *sartan AND NOT medication administered route ~ ophthalmic OR dermatological Follow the instructions on section 2.1 Phase 1, Step 6 De-identify data including disguising dates.

[7] See introductory notes for extraction at the top of Laboratory Tests section for guidance applicable to all laboratory tests of interest.

DATA ELEMENT	DEFINITION	DATA TYPE	ALLOWABLE VALUES	OCCURS	LOGIC AND NOTES FOR EXTRACTION[7]
Azithromycin (azythromicin)	The first date during the COVID-19 related hospitalization when azythromicin was administered to the patient.	Integer	Any valid integer NULL = unknown, missing or medication not administered during inpatient stay	1	FIRST date >= admission date AND <= discharge date when Medication administered in ***Azythromicin*** OR Simple generic name LIKE azythromicin AND NOT medication administration route ~ ophthalmic OR dermatological Follow the instructions on section 2.1 Phase 1, Step 6 De-identify data including disguising dates.
Dexamethasone (dexamethasone)	The first date during the COVID-19 related hospitalization when dexamethasone was administered to the patient.	Integer	Any valid integer NULL = unknown, missing or medication not administered during inpatient stay	1	FIRST date >= admission date AND <= discharge date when Medication administered in ***Dexamethasone*** OR Simple generic name LIKE dexamet* AND medication administration route ~ oral OR ~ IV or ~ peg/g/j/og/other tube Follow the instructions on section 2.1 Phase 1, Step 6 De-identify data including disguising dates.
Famotidine (famotidine)	The first date during the COVID-19 related hospitalization when famotidine was administered to the patient.	Integer	Any valid integer NULL = unknown, missing or medication not administered during inpatient stay	1	FIRST date >= admission date AND <= discharge date when Medication administered in ***Famotidine*** OR Simple generic name LIKE famotidine AND NOT medication administration route ~ ophthalmic OR dermatological Follow the instructions on section 2.1 Phase 1, Step 6 De-identify data including disguising dates.

DATA ELEMENT	DEFINITION	DATA TYPE	ALLOWABLE VALUES	OCCURS	LOGIC AND NOTES FOR EXTRACTION[7]
Systemic Corticosteroids other than Dexamethasone (systemic_steroids)	The first date during the COVID-19 related hospitalization when systemic corticosteroids other than dexamethasone, including hydrocortisone, methylprednisone and prednisone, were administered to the patient.	Integer	Any valid integer NULL = unknown, missing or medication not administered during inpatient stay	1	FIRST date >= admission date AND <= discharge date when Medication administered in ***Systemic Corticosteroids other than Dexamethasone*** OR Simple generic name LIKE hydrocortisone OR methlypred* OR prednisone AND medication administration route ~ oral OR ~ IV or ~ peg/g/j/og/other tube Follow the instructions on section 2.1 Phase 1, Step 6 De-identify data including disguising dates.
Hydroxychloroquine (hydroxychloroquine)	The first date during the COVID-19 related hospitalization when hydroxychloroquine was administered to the patient.	Integer	Any valid integer NULL = unknown, missing or medication not administered during inpatient stay	1	FIRST date >= admission date AND <= discharge date when Medication administered in ***Hydroxychloroquine*** OR Simple generic name LIKE hydroxychloroquine AND NOT medication administration route ~ ophthalmic OR dermatological Follow the instructions on section 2.1 Phase 1, Step 6 De-identify data including disguising dates.
Interleukin-6 (IL-6) inhibitors (il6_inhibitors)	The first date during the COVID-19 related hospitalization when IL-6 inhibitors was administered to the patient.	Integer	Any valid integer NULL = unknown, missing or medication not administered during inpatient stay	1	FIRST date >= admission date AND <= discharge date when Medication administered in ***IL-6 inhibitors*** OR Simple generic name LIKE tocilizumab OR sarilumab AND NOT medication administration route ~ ophthalmic OR dermatological Follow the instructions on section 2.1 Phase 1, Step 6 De-identify data including disguising dates.

DATA ELEMENT	DEFINITION	DATA TYPE	ALLOWABLE VALUES	OCCURS	LOGIC AND NOTES FOR EXTRACTION[7]
Remdesivir (remdesivir)	The first date during the COVID-19 related hospitalization when remdesivir was administered to the patient.	Integer	Any valid integer NULL = unknown, missing or medication not administered during inpatient stay	1	FIRST date >= admission date AND <= discharge date when Medication administered in ***Remdesivir*** OR Simple generic name LIKE remdesivir AND NOT medication administration route ~ ophthalmic OR dermatological Follow the instructions on section 2.1 Phase 1, Step 6 De-identify data including disguising dates.