

Supplementary Materials for
**Privacy-Protecting, Reliable Response Data Discovery Using COVID-19
Patient Observations**

Jihoon Kim[†], Larissa Neumann[†], Paulina Paul, Michael Aratow, Douglas S. Bell, Jason N. Doctor, Ludwig C. Hinske, Xiaoqian Jiang, Katherine K. Kim, Michael E. Matheny, Daniella Meeker, Mark J. Pletcher, Lisa M. Schilling, Spencer SooHoo, Hua Xu, Kai Zheng, Lucila Ohno-Machado* for the R2D2 Consortium

[†]These authors contributed equally to this work.

*Correspondence to: lohnomachado@health.ucsd.edu

This PDF file includes:

Materials and Methods
Figs. S1 to S5
Tables S1 to S5
R2D2 Consortium Members

Materials and Methods

Our approach is practical and generalizable: The network can be repurposed to any other disease of interest, as it is not based exclusively on data elements deemed relevant for COVID-19. Because privacy protection is at the core of our network, a wide range of institutions can participate. We provide a rapidly deployable and reproducible alternative or complement to centralized registries of EHR data that allows healthcare institutions to stay in control of their data. The development of our Q&A system involved the inclusion of new concept codes in local repositories, agreement on concept definitions (e.g., what constitutes a *COVID-19 hospitalization*, what codes should be included in the definition of *History of Coronary Heart Disease*, and how to map laboratory test records into LOINC, for which we developed a mapping tool) (21). Instead of centralizing data at the Consortium Hub, we focus on interpreting and clarifying the research questions in order to determine the data elements required. Our teams analyze these data elements to generate aggregate statistics at the various institutions, documenting the specific version of structured query language (SQL) code executed at a specific time point to generate their answers. In addition to basic counts and proportions, to adjust for confounders, we use distributed multivariate analyses to estimate risk-adjusted odds ratios. This is done in a synchronized fashion for iterative federated algorithms, such as one previously reported for building a logistic regression model (11). We have shown previously that a model obtained this way is identical to one built using data that are centralized in a single location.

Workflow

Fig. S1 shows our general workflow, including human interpretation and clarification of questions, human quality control of answers, and using graphs and related visualizations as much as possible. The responsibility of the Lead Site--to create a template query for all responding sites to use for rapid response--rotates among institutions (i.e., Health Systems). A more detailed workflow is illustrated in Fig. S2 using a swim-lane format with an emphasis on roles. The Q&A process starts when a user creates a request through the public web-site, <https://covid19questions.org>. Next, the data scientist at the Consortium Hub checks if this question had been answered before and pass it on to the clinician at the Consortium Hub to assess the feasibility (i.e., if the received question is answerable from the local data mart) and assigns it to one of the 12 institutions as a Lead Site. Throughout the whole process, the tracking system is used to report an issue to assignees, respond to the issue, update the code and results, and prompt to re-run the updated SQL. Next, another clinician at the Lead Site will work with database analyst to review and develop concept set. This is an iterative process within the Lead Site, to develop concept set, create SQL, generate results, and evaluate the results against the EHR records including chart reviews. The outputs of Lead Site level process are a template query (.sql format) and a template output (.csv format), which are uploaded to the shared code repository. Once 11 Responding Sites get notification emails about the template query and format for the results, their database analysts will run the template SQL to get preliminary results, and review these against their EHR data with clinicians.

This part of the process is where the Responding Site most frequently runs into errors, challenges, and requires troubleshooting. For example, when missing concepts like D-Dimer or blood type, illustrated in Fig. S3, are discovered, the database analyst at the Responding Site creates an issue in the tracking system and resolves this with the database analyst and the clinician at the Leading Site. Since there are 11 sites, this means the Lead Site coordinates the

concept set and SQL development through one-on-one sessions between the Lead Site and Responding Site. Through this iterative process among 12 sites, the concept set and SQL keep being updated and keep improving their sensitivity and specificity to identify the right patients and hospitalization encounter records. This involves rewriting and updating existing Extract-Transform-Load (ETL) scripts to map source EHR data to target the common data model (CDM), which in our case is the Observational Medical Outcomes Partnership -- OMOP). The institutions with the same EHR system or database management system share common experience and knowledge to help each other develop ETL scripts together and evaluate the OMOP query results against EHRs.

Once all Responding Sites have uploaded their site level results, the Data Scientist at Consortium Hub merges these results into a single file. A generic and extensible format for site-level summary result is used to answer general epidemiology and clinical research questions (Fig. S4). Then a data quality check is conducted. While use of a CDM in a large clinical data research network is a widely used approach to enable interoperable query development, a query formulated in one institution may not return accurate results in another due to variations in data integration and data quality differences. Several rounds of confirmations and checks with data analysts and clinical informaticians at each institution are often necessary to answer questions with confidence. There are many potential sources of errors and Table S1 displays selected examples of data quality checks with real issues. The check types are based on PEDSnet framework (22) and revised to fit our project's specific needs. The Data Scientist resolves issues together with the Lead Site and the Responding Site. When the aggregate results pass the quality control (QC) test, the Consortium Hub Clinician conducts the final review to ensure its clinical relevancy. Finally, if the clinician approves the release of the result, the Data Scientist uploads the answer to the public web-site (<https://covid19questions.org>) and notifies the requestor via an email.

Cohort and Concept Set

As questions frequently refer to the same subsets of patients, we developed electronic cohort definitions that facilitate our answers. We followed the United States Centers for Disease Control and Prevention (CDC) guideline (23) and the National COVID-19 Cohort Collaborative (9) and Observational Health Data Sciences and Informatics (OHDSI) approaches (24) to develop a cohort of hospitalization encounters for COVID-19 as a base for all inpatient questions. Fig. S3A displays the electronic phenotyping of adults hospitalized with COVID-19 derived by canonical SQL scripts. Hospitalization encounters were identified by using the following concepts stored in the OMOP <VISIT_OCCURRENCE> table: Emergency Room and Inpatient Visit (Concept Id 262), Inpatient Visit (Concept Id 92021) or Intensive Care (Concept Id 32037). To enter the COVID-19 hospitalization cohort, all four inclusion criteria needed to be met:

- 1) a minimum age of 18 years at the date of hospitalization,
- 2-3) a hospitalization without a length of stay requirement on or after January 1, 2020 and
- 4) at least one occurrence of
 - a) a positive viral test for SARS-CoV-2, or
 - b) a COVID-19 related diagnosis between the interval of 21 days prior to hospitalization and hospital encounter discharge.

The following concepts of the OMOP <MEASUREMENT> table for the definition of a positive viral test for SARS-COV-2 were used:

- 1) the occurrence of the pre-coordinated measurement Concept (Concept Name: 2019 novel coronavirus detected, Concept Id: 37310282), or
- 2) the occurrence of at least one concept for a SARS-CoV-2 viral test (e.g. Concept Name: SARS-CoV-2 (COVID19) RNA [Presence] in Respiratory specimen by NAA with probe detection, Concept Id: 706163) and at least one *value_as_concept_id* for a positive result (e.g. Concept Name: Positive, Concept Id: 45884084).

For identification of COVID-19 related diagnoses, we included the following ICD-10-CM Codes: Other coronavirus as the cause of diseases classified elsewhere (B97.29), COVID-19, virus identified (U07.1), Pneumonia (J12.89), Acute Bronchitis (J20.8), Lower Respiratory Infection (J22, J98.8) and Acute Respiratory Distress Syndrome (J80). Following two ICD-10-CM Official Coding and Reporting Guidelines released by CDC before and at/after April 1, 2020, we used diagnosis code aggregations to define a COVID-19 related diagnosis. An illness due to COVID-19 was specified if one of the ICD-10-CM codes (J12.89, J20.8, J22, J98.8, J80) was recorded in combination with either B97.29 (before April 1, 2020), or in combination with U07.1 (on/after April 1, 2020). These joint diagnosis codes needed to occur during the same hospitalization encounter, with a look-back period of 21 days prior to hospitalization. We applied the same logic for mapped SNOMED Concepts (261326, 260139, 4307774, 256451, 4195694, 320136, 4100065, 37311061). More ICD codes are detailed in Table S2. Pre-coordinated diagnoses codes (SNOMED, OMOP Extension) are shown in Table S3, followed by excluded codes in Tables S4 and S5. Refinement of phenotypes was guided by chart review.

Use cases of concept set are shown in Fig. S3B. As the Responding Sites' OMOP databases are not accessible to the Lead Site, a query developed at the Lead Site might miss a concept used in other sites. In such case, the database analyst at the Responding Site notifies the Lead Site by creating a GitHub issue, with zero or unexpectedly low count or proportion in the results generated by the initial template query authored by the Lead Site. For example, in Fig. S1B, during the Concept Set development for the quantitative laboratory measurement D-dimer, the responding site notified the Lead site about using another concept for D-dimer (Concept ID: 3048540 instead of Concept ID: 3051714), returning values with a different measurement unit than the ones of the Lead site (n/L instead of mg/L). Therefore, the Lead site had to add the missing concept to Concept Set and implemented logic to cover a measurement unit transformation. In the case of the Concept Set development for blood type, a responding site was missing concepts for blood type in its local OMOP CDM database. An ETL script was implemented to map EHR data to OMOP CDM. Sources of discrepancy were diverse; examples included unit differences in measurement values, differently mapped concepts, and non-compliance to the coding guideline.

GLORE

In addition to count queries, we also applied Grid Binary LOGistic REgression (GLORE) (10) to compute the effect of the exposure variable on the outcome, adjusted for confounders, without sharing patient-level data, as this would increase the risk for a privacy breach. We rewrote the

Newton-Raphson (NR) method to find the maximizer of the likelihood function of the parameters in logistic regression. Since the first and the second derivatives of the log likelihood functions are separable (i.e., they can be partially calculated at each site), in each NR iteration, each client institute calculated a $(p+1)$ dimensional vector of parameters, where p is the number of features in the model such as *age*, *gender*, and *race* and a $(p+1)$ by $(p+1)$ variance-covariance matrix, and sent JSON files containing these two objects to the Consortium Hub. Fig. S5 shows the screen shot of the real example JSON file used the GLORE run to answer the in-hospital mortality question. No patient level information was shared or transferred between institutions. At each iteration, the Consortium Hub automatically updated the global coefficient vector and the variance-covariance matrix and sent them back to clients. All clients repeatedly sent the updated JSON file to the Consortium Hub, until the estimates stabilized or reached a pre-defined number of iterations. To enhance the security, the Consortium Hub server allowed (i.e., “whitelisted”) only the pre-registered IP addresses of client machines and opened the port only during the scheduled time window.

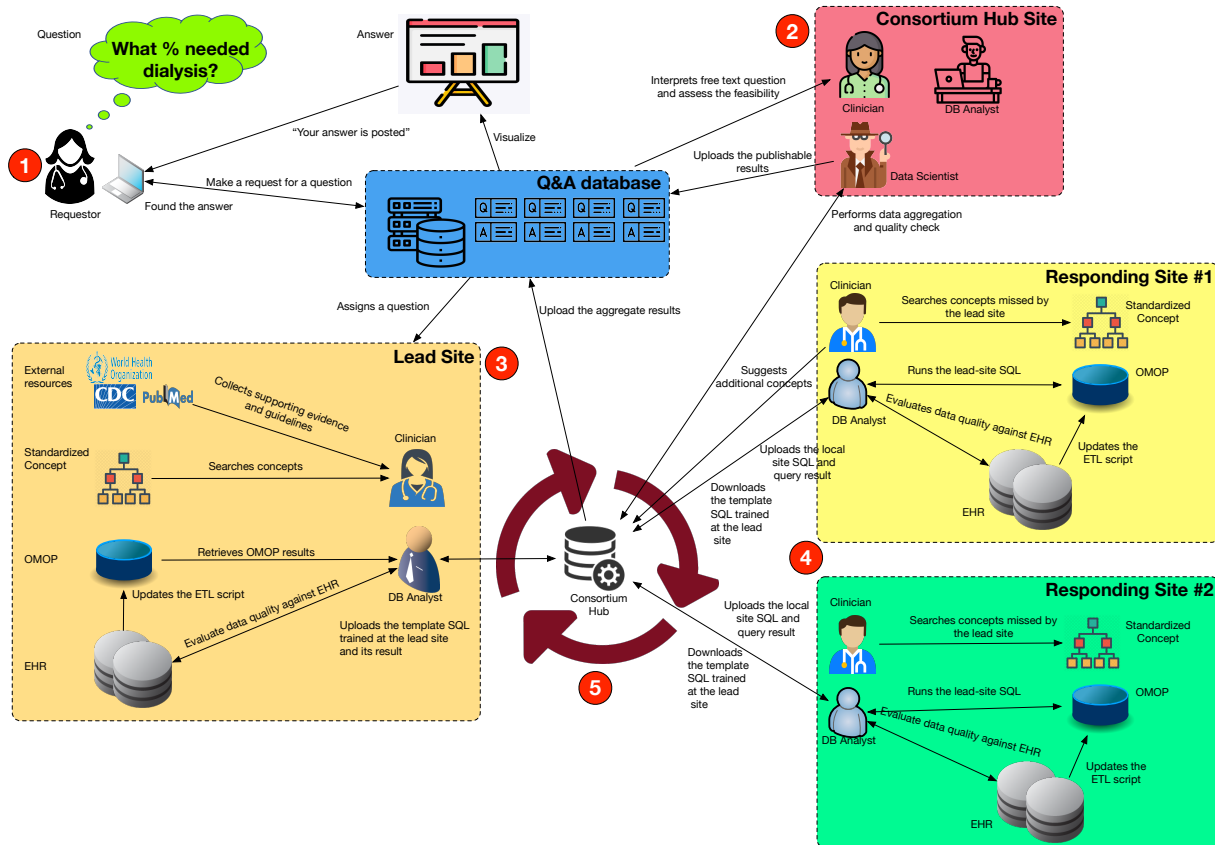


Fig S1. What happens behind the scenes: From questions to answers. Our Consortium involves humans in the clinical interpretation of questions and data quality control. The responsibility for formulating the code for a question is delegated to a different Lead Site for each question. Step 1. Users access a public web portal and search/browse if there is an answer posted or a pending response to their question. If not, they propose a question. Step 2. The questions get triaged to a Consortium Hub Clinical Informatician who determines whether (a) the answer will be of general interest and (b) how the question needs to be revised so that it can be answered by most sites. If these criteria are met, the Consortium Hub Clinical Informatician assigns the question to a Clinical Informatician at the site that will serve as the Lead Site, delegating the question and end-to-end process of query and response development. Step 3. The Lead site Clinical Informatician works together with the Lead Site Data Analyst to create concept sets, design a query, and check local results. Some iterations may happen before the Lead Site Clinical Informatician approves the query and submits to the Consortium Hub, where it is distributed to the Responding Sites. Step 4. The Responding Site Clinical Informatician and the Responding Site Data Analyst adjust the concept set, inclusion logic, and database query code in structured query language (SQL) for local implementation, obtain and quality controls the site-level results, and submit results to the Consortium Hub. For distributed multivariate analytics, a table is kept locally to be processed by the client software that transfers the aggregate results to the server software running at the Hub. Iterative algorithms are run in agreed-upon synchronous sessions using a server that is operational only during the analytics run and a firewall that allows only white-listed internet protocol (IP) addresses. Step 5. Site-level results are aggregated at the Hub and submitted to the Analytics Lead, who checks for

discrepancies across sites. If an error is detected, the local site is informed to fix it. Once there are no discrepancies, the Lead Site Analyst generates the visualizations and submits to the Consortium Hub Clinical Informatician for approval and posting on the portal.

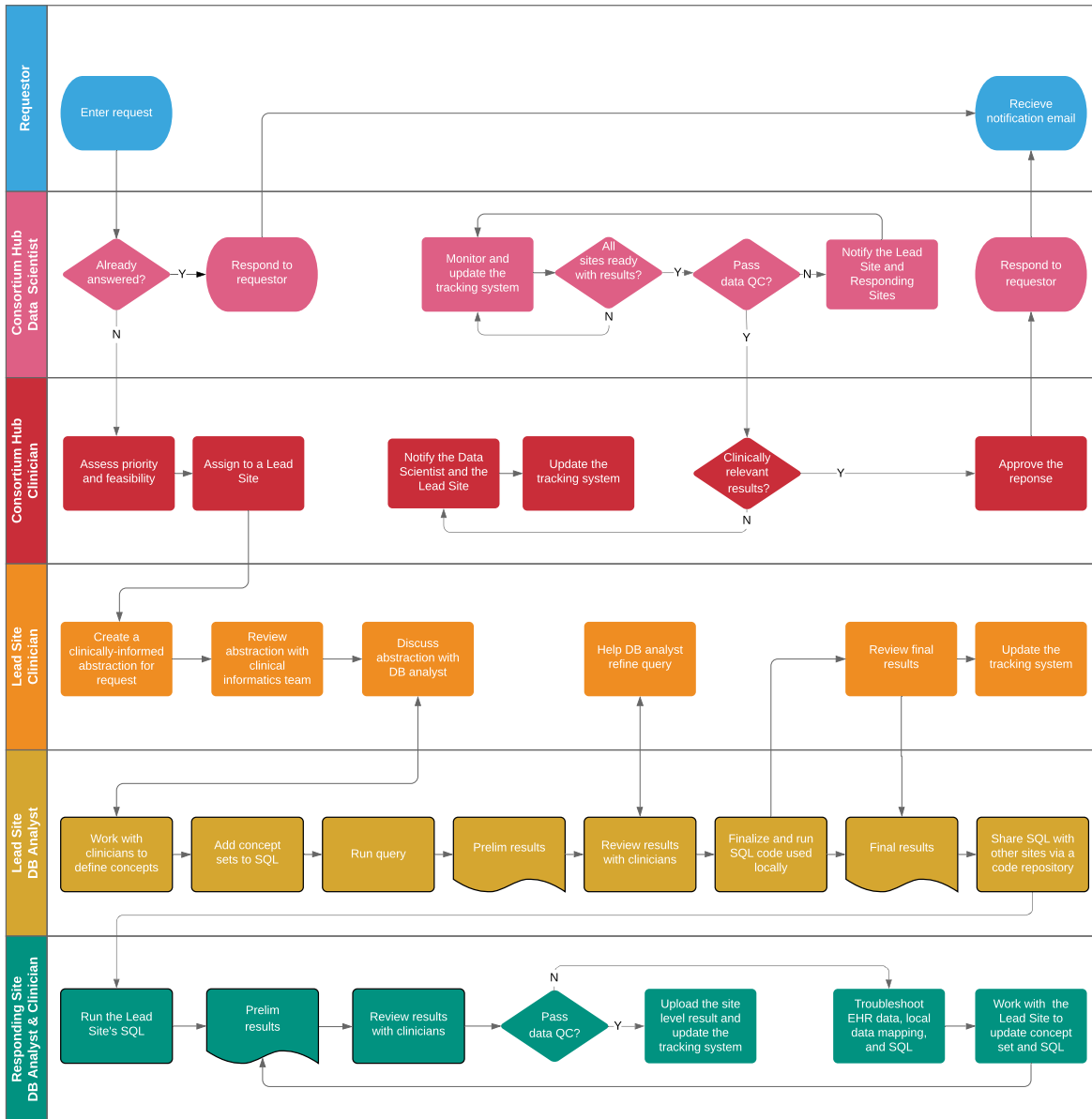


Fig. S2. Swimlane diagram. A Q&A process flow starts from a user entering a request and ends with the user receiving email notification about a response. Both clinicians and database analysts work together to answer the user’s question.

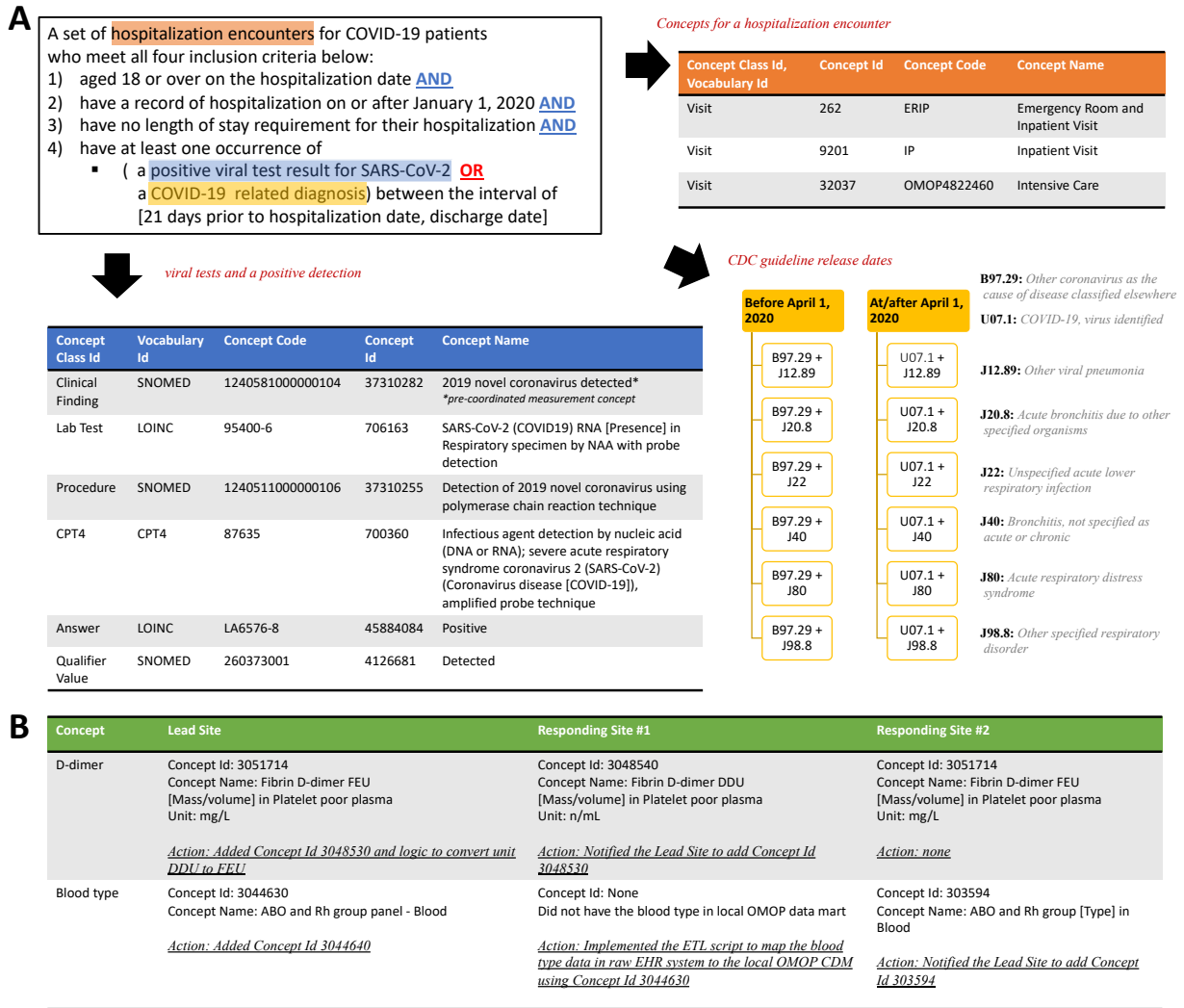


Fig. S3. Cohort definition and Concept set development. Defining a cohort of patients that is frequently used to answer questions helps us reuse code. In this example, defining the cohort of patients hospitalized with COVID-19 involves use of SARS-CoV-2 test results or diagnosis codes (A). In (B), we illustrate how a laboratory test is defined differently at two sites, and how blood type had yet to be harmonized into OMOP at one site.



Fig. S4. Extensible output format of site level results. A 2-by-2 table is converted to SQL-friendly 4-row format to store the binary exposure and the binary outcome question. Then it expands to add a covariate, second outcome, or another exposure. A name-value format was adopted for a clarity and efficiency for data quality check and aggregation.

```

{
  "SD_matrix": "[0.5077726510299974, 0.6324728142599232, 0.18220658774632076, 0.37292021986559626, 0.43743776771472637, 0.1960131777242006, 0.19834144545189025]",
  "Beta 1": "[-6.571584585029573, 4.898163226230124, 0.28996329717199837, 0.11797777998361085, -0.7386399038721124, 0.6327913437876899, 0.12963251924656277]",
  "Covariance matrix": [
    "[0.2578330651340315, -0.29910112576665693, -0.030721970027503372, -0.02629726036935407, -0.027684288243071647, -0.02612168473239688, -0.03248897336604138]",
    "[-0.29910112576665693, 0.4000218607778673, 0.011436929894340265, 0.006584197762780007, 0.006928118984794617, 0.01878901528913741, 0.026788335914024304]",
    "[-0.030721970027503383, 0.011436929894340272, 0.03319924061815768, 0.00343311678189749, 0.005303086831563236, 6.284006568122062E-4, 0.002145829717076881]",
    "[-0.026297260369354033, 0.006584197762779952, 0.0034331167818974825, 0.13906949038460464, 0.01962845021821445, 0.012384909942492117, 0.011932499137359035]",
    "[-0.027684288243071616, 0.006928118984794577, 0.0053030868315632335, 0.019628450218214452, 0.19135180062324292, 0.012535724405576027, 0.01162818126021859]",
    "[-0.026121684732396865, 0.01878901528913739, 6.284006568122079E-4, 0.012384909942492119, 0.01253572440557602, 0.03842116584153905, -0.010711730049463304]",
    "[-0.03248897336604139, 0.026788335914024314, 0.0021458297170768715, 0.011932499137359038, 0.011628181260218596, -0.010711730049463295, 0.03933932898394516]"
  ]
}

```

Fig. S5. Transferred file in the federated logistic regression. A real example of a transferred file in JSON format among the Consortium Hub and participating sites is shown to illustrate how patient level data are protected. The first row is a vector of standard deviations of the coefficients of the global logistic regression. The second row is a vector of coefficients. The third row is a covariance matrix between features. At each iteration of the federated logistic regression, this JSON file is being transferred among sites until convergence or the process reaches the predetermined number of iterations. The values in JSON file above are used in the forest plot in Fig. 2.

Table S1. Data quality checks and issues. Different data quality check types are enumerated together with real issues identified with this COVID-19 project.

Check Type	Example of data quality issue
Datetime reversal	A condition/observation was recorded after discharge date
Extreme outlier	The hospital length of stay was greater than 80 days. The median length of stay ranged between 11 and 15 days in China and US studies
Gaps in data transformation	Discharge disposition and ICU departments were not transformed to OMOP
Loss of granularity during mapping	Invasive and non-invasive mechanical ventilation mapped to the same concept
Impossible events	Multiple death events occurred in different time points from multiple hospital encounters
Non-compliance to the output format	Header was missing in the predefined output .csv format, missing columns, shifted columns, and duplicate rows
Unexpected proportion	The percentage of current smokers was 65% at a certain site. The national percentage of smoking was 15.6% among male adults in 2018 US CDC data
Unexpected zero count	The number of patients who were taking any anti-hypertensives was zero
Unmatched group sum	The total sums of patient count in age groups and race groups were different even when all cell counts were greater than 10
Version mismatch	The version of the template query was revised after the query result was uploaded

Table S2. Concept Relationships between ICD10CM and SNOMED Concepts. ICD10CM Concepts and their mapped SNOMED Concepts from the CONDITION_OCCURRENCE table. In OMOP CDM, ICD10CM Concepts are non-Standard Concepts. Therefore, ICD10CM Concepts are mapped to SNOMED-based Standard Concepts (9). These relationships are stored in the OMOP CDM CONCEPT_RELATIONSHIP table. In this case each ICD10CM Concept got the relationship_id = 'Maps to', which directs to one SNOMED Concept.

Concept Code 1 (ICD10CM)	Concept Name 1	Concept Id 1	Relationship Id	Concept Id 2	Concept Name 2	Concept Code 2 (SNOMED)
J12.89	Other viral pneumonia	45572161	'Maps to'	261326	Viral pneumonia	75570004
J20.8	Acute bronchitis due to other specified organisms	35207965	'Maps to'	260139	Acute bronchitis	10509002
J22	Unspecified acute lower respiratory infection	35207970	'Maps to'	4307774	Acute lower respiratory tract infection	195742007
J40	Bronchitis, not specified as acute or chronic	35208013	'Maps to'	256451	Bronchitis	32398004
J80	Acute respiratory distress syndrome	35208069	'Maps to'	4195694	Acute respiratory distress syndrome	67782005
J98.8	Other specified respiratory disorders	35208108	'Maps to'	320136	Disorder of respiratory system	50043002
B97.29	Other coronavirus as the cause of diseases classified elsewhere	45600471	'Maps to'	4100065	Disease due to Coronaviridae	27619001
U07.1	Emergency use of U07.1 Disease caused by severe acute respiratory syndrome coronavirus 2	702953	'Maps to'	37311061	Disease caused by 2019-nCoV	840539006

Table S3. Pre-coordinated diagnosis codes. Diagnosis codes (OMOP Extension, SNOMED) used to identify patients with COVID-19. At least one occurrence of the diagnosis codes during a hospital encounter with a look back period of 21 days prior to hospitalization captured the patient having a COVID-19 related diagnosis. In contrast to the joint diagnosis codes (ICD-10-CM, SNOMED) mentioned in Fig 3A and Table S3, there was no further applied logic.

Concept Class Id	Vocabulary Id	Concept Code	Concept Id	Concept Name
Clinical Finding	OMOP Extension	OMOP4873906	756023	Acute bronchitis due to COVID-19
Clinical Finding	OMOP Extension	OMOP4873911	756044	Acute respiratory distress syndrome (ARDS) due to COVID-19
Clinical Finding	OMOP Extension	OMOP4873910	756061	Asymptomatic COVID-19
Clinical Finding	OMOP Extension	OMOP4873909	756031	Bronchitis due to COVID-19
Clinical Finding	SNOMED	1240561000000108	37310284	Encephalopathy caused by 2019 novel coronavirus
Clinical Finding	SNOMED	1240571000000101	37310283	Gastroenteritis caused by 2019 novel coronavirus
Clinical Finding	OMOP Extension	OMOP4873908	756081	Infection of lower respiratory tract due to COVID-19
Clinical Finding	SNOMED	1240541000000107	37310286	Infection of upper respiratory tract caused by 2019 novel coronavirus
Clinical Finding	SNOMED	1240531000000103	37310287	Myocarditis caused by 2019 novel coronavirus
Clinical Finding	SNOMED	1240521000000100	37310254	Otitis media caused by 2019 novel coronavirus
Clinical Finding	SNOMED	1240551000000105	37310285	Pneumonia caused by 2019 novel coronavirus
Clinical Finding	OMOP Extension	OMOP4873907	756039	Respiratory infection due to COVID-19

Table S4. Excluded diagnosis codes. These ICD-10-CM Codes, mentioned in CDC guideline (29), were excluded given the high count of false positive COVID-19 patients.

ICD-10-CM Code	Description
R05	Cough
R06.02	Shortness of breath
R50.9	Fever, unspecified
Z20.828	Contact with and (suspected) exposure to other viral communicable diseases

Table S5. Other excluded diagnosis codes. These ICD10CM and SNOMED Concepts were mentioned in N3C - COVID-19 Phenotype Documentation, Version 1.6 (Last updated 6/5/2020) but excluded from our study, as the count of false positives COVID-19 patients was too high.

Concept Class Id	Vocabulary Id	Concept Code	Concept Id	Concept Name
Condition	ICD10CM	J96.2	35208101	Acute and chronic respiratory failure
Condition	ICD10CM	J96.22	45581868	Acute and chronic respiratory failure with hypercapnia
Condition	ICD10CM	J96.21	45543283	Acute and chronic respiratory failure with hypoxia
Condition	ICD10CM	J96.20	45596290	Acute and chronic respiratory failure, unspecified whether with hypoxia or hypercapnia
Condition	ICD10CM	J21	1569471	Acute bronchiolitis
Condition	ICD10CM	J21.1	920135	Acute bronchiolitis due to human metapneumovirus
Condition	ICD10CM	J21.8	35207968	Acute bronchiolitis due to other specified organisms
Condition	ICD10CM	J21.0	35207967	Acute bronchiolitis due to respiratory syncytial virus
Condition	ICD10CM	J21.9	35207969	Acute bronchiolitis, unspecified
Condition	ICD10CM	J20	1569470	Acute bronchitis
Condition	ICD10CM	J20.3	35207960	Acute bronchitis due to coxsackievirus
Condition	ICD10CM	J20.7	35207964	Acute bronchitis due to echovirus
Condition	ICD10CM	J20.1	35207958	Acute bronchitis due to Hemophilus influenzae
Condition	ICD10CM	J20.0	35207957	Acute bronchitis due to Mycoplasma pneumoniae
Condition	ICD10CM	J20.8	35207965	Acute bronchitis due to other specified organisms
Condition	ICD10CM	J20.4	35207961	Acute bronchitis due to parainfluenza virus
Condition	ICD10CM	J20.5	35207962	Acute bronchitis due to respiratory syncytial virus
Condition	ICD10CM	J20.6	35207963	Acute bronchitis due to rhinovirus
Condition	ICD10CM	J20.2	35207959	Acute bronchitis due to streptococcus
Condition	ICD10CM	J20.9	35207966	Acute bronchitis, unspecified
Condition	ICD10CM	R06.03	1326788	Acute respiratory distress
Condition	ICD10CM	J80	35208069	Acute respiratory distress syndrome
Condition	ICD10CM	J96.0	35208099	Acute respiratory failure
Condition	ICD10CM	J96.02	45596289	Acute respiratory failure with hypercapnia
Condition	ICD10CM	J96.01	45567283	Acute respiratory failure with hypoxia
Condition	ICD10CM	J96.00	45605906	Acute respiratory failure, unspecified whether with hypoxia or hypercapnia
Condition	ICD10CM	J06.9	35207929	Acute upper respiratory infection, unspecified
Condition	ICD10CM	J12.0	35207932	Adenoviral pneumonia
Condition	ICD10CM	R43.0	35211351	Anosmia
Condition	ICD10CM	J40	35208013	Bronchitis, not specified as acute or chronic

Condition	ICD10CM	J18.0	35207952	Bronchopneumonia, unspecified organism
Condition	ICD10CM	R07.1	35211284	Chest pain on breathing
Condition	ICD10CM	R68.83	45577807	Chills (without fever)
Condition	ICD10CM	J96.1	35208100	Chronic respiratory failure
Condition	ICD10CM	J96.12	45572177	Chronic respiratory failure with hypercapnia
Condition	ICD10CM	J96.11	45538489	Chronic respiratory failure with hypoxia
Condition	ICD10CM	J96.10	45548131	Chronic respiratory failure, unspecified whether with hypoxia or hypercapnia
Observation	ICD10CM	Z20.828	45542411	Contact with and (suspected) exposure to other viral communicable diseases
Condition	ICD10CM	B34.2	35205800	Coronavirus infection, unspecified
Condition	ICD10CM	R05	35211275	Cough
Condition	ICD10CM	R50.2	35211385	Drug induced fever
Condition	ICD10CM	R06.0	1572191	Dyspnea
Condition	ICD10CM	R06.00	45587496	Dyspnea, unspecified
Condition	ICD10CM	R50.84	45597190	Febrile nonhemolytic transfusion reaction
Condition	ICD10CM	R50	1572254	Fever of other and unknown origin
Condition	ICD10CM	R50.81	45606818	Fever presenting with conditions classified elsewhere
Condition	ICD10CM	R50.9	35211387	Fever, unspecified
Condition	ICD10CM	J12.3	35207935	Human metapneumovirus pneumonia
Condition	ICD10CM	J18.2	35207954	Hypostatic pneumonia, unspecified organism
Condition	ICD10CM	J18.1	35207953	Lobar pneumonia, unspecified organism
Condition	ICD10CM	R06.01	45597165	Orthopnea
Condition	ICD10CM	R06.09	45548944	Other forms of dyspnea
Condition	ICD10CM	J18.8	35207955	Other pneumonia, unspecified organism
Condition	ICD10CM	R50.8	35211386	Other specified fever
Condition	ICD10CM	J98.8	35208108	Other specified respiratory disorders
Condition	ICD10CM	J12.8	35207936	Other viral pneumonia
Condition	ICD10CM	R43.2	35211353	Parageusia
Condition	ICD10CM	J12.2	35207934	Parainfluenza virus pneumonia
Condition	ICD10CM	J12.81	45567260	Pneumonia due to SARS-associated coronavirus
Condition	ICD10CM	J18.9	35207956	Pneumonia, unspecified organism
Condition	ICD10CM	J18	1569469	Pneumonia, unspecified organism
Condition	ICD10CM	R50.82	45597189	Postprocedural fever
Condition	ICD10CM	R50.83	45592424	Postvaccination fever
Condition	ICD10CM	J96	1569515	Respiratory failure, not elsewhere classified
Condition	ICD10CM	J96.9	35208102	Respiratory failure, unspecified
Condition	ICD10CM	J96.92	45533563	Respiratory failure, unspecified with hypercapnia
Condition	ICD10CM	J96.91	45605907	Respiratory failure, unspecified with hypoxia
Condition	ICD10CM	J96.90	45567284	Respiratory failure, unspecified, unspecified whether with hypoxia or hypercapnia

Condition	ICD10CM	J12.1	35207933	Respiratory syncytial virus pneumonia
Condition	ICD10CM	R06.02	45534422	Shortness of breath
Condition	ICD10CM	J12	1569465	Viral pneumonia, not elsewhere classified
Condition	ICD10CM	J12.9	35207937	Viral pneumonia, unspecified
Condition	SNOMED	75483001	442555	Breathing painful
Condition	SNOMED	161940008	4059022	Breathless - mild exertion
Condition	SNOMED	161939006	4059021	Breathless - moderate exertion
Condition	SNOMED	161855003	4059003	C/O shivering
Condition	SNOMED	274664007	4168213	Chest pain on breathing
Condition	SNOMED	43724002	434490	Chill
Condition	SNOMED	49727002	254761	Cough
Condition	SNOMED	135883003	4048098	Cough with fever
Condition	SNOMED	11833005	4038519	Dry cough
Condition	SNOMED	267036007	312437	Dyspnea
Condition	SNOMED	161941007	4060052	Dyspnea at rest
Condition	SNOMED	60845006	4263848	Dyspnea on exertion
Observation	SNOMED	840546002	37311059	Exposure to 2019 novel coronavirus
Condition	SNOMED	103001002	4011766	Feeling feverish
Condition	SNOMED	386661006	437663	Fever
Measurement	SNOMED	426000000	4141062	Fever greater than 100.4 Fahrenheit
Condition	SNOMED	274640006	4164645	Fever with chills
Condition	SNOMED	23141003	4047610	Gaspings for breath
Condition	SNOMED	409702008	4260205	Hyperpyrexia
Condition	SNOMED	44169009	4185711	Loss of sense of smell
Condition	SNOMED	36955009	4289517	Loss of taste
Condition	SNOMED	426976009	4140453	Pain provoked by breathing
Condition	SNOMED	247410004	4090569	Painful cough
Condition	SNOMED	284523002	4109381	Persistent cough
Condition	SNOMED	2237002	4330445	Pleuritic pain
Condition	SNOMED	28743005	4102774	Productive cough

R2D2 Consortium Members			
Last Name	First Name	Middle Name	Affiliation(s)
Anderson	David	M	San Mateo County Health, Health Information Technology, San Mateo, CA, USA
Aratow	Michael		San Mateo Medical Center, San Mateo, CA, USA
Balacha	Chandrasekar		University of Southern California, Viterbi School of Engineering, University of Southern California, Los Angeles, CA, USA
Bath	Tyler		UC San Diego Health Department of Biomedical Informatics, University of California, San Diego, La Jolla, CA, USA
Baxter	Sally	L	UC San Diego Health Department of Biomedical Informatics, University of California, San Diego, La Jolla, CA, USA; UCSD Viterbi Family Department of Ophthalmology and Shiley Eye Institute, University of California, San Diego, La Jolla, CA, USA
Becker-Pennrich	Andrea		LMU Klinikum, Department of Anaesthesiology, Ludwig Maximilian University of Munich, Munich, Bavaria, Germany
Bell	Douglas	S	Biomedical Informatics Program, UCLA Clinical and Translational Science Institute (CTSI), Los Angeles, CA, USA
Bernstam	Elmer	V	School of Biomedical Informatics, The University of Texas Health Science Center at Houston, Houston, TX, USA
Carter	William	A	Data Science to Patient Value Program, University of Colorado Anschutz Medical Campus, Aurora, CO, USA
Ngan	Chau		Biomedical Informatics Program, UCLA Clinical and Translational Science Institute (CTSI), Los Angeles, CA, USA
Day	Michele	E	UC San Diego Health Department of Biomedical Informatics, University of California, San Diego, La Jolla, CA, USA
Doctor	Jason	N	USC Schaeffer Center for Health Policy and Economics, Price School of Policy, University of Southern California, Los Angeles, CA, USA
DuVall	Scott		Salt Lake City VA, Salt Lake City, UT, USA; Division of Epidemiology, University of Utah, Salt Lake City, UT, USA
El-Kareh	Robert		UC San Diego Health Department of Biomedical Informatics, University of California, San Diego, La Jolla, CA, USA
Florian	Renato		Departments of General Internal Medicine and Enterprise Information Services, Cedars Sinai Medical Center, Los Angeles, CA, USA
Follett	Robert	W	Biomedical Informatics Program, UCLA Clinical and Translational Science Institute (CTSI), Los Angeles, CA, USA
Geisler	Benjamin	P	Institute for Medical Information Processing, Biometry, and Epidemiology (IBE), Ludwig Maximilian University of Munich, Munich, Bavaria, Germany
Ghigi	Alessandro		Institute for Clinical and Translational Science, University of California, Irvine, Irvine, CA, USA
Gottlieb	Assaf		School of Biomedical Informatics, The University of Texas Health Science Center at Houston, Houston, TX, USA
Hinske	Ludwig	C	Institute for Medical Information Processing, Biometry, and Epidemiology (IBE), Ludwig Maximilian University of Munich, Munich, Bavaria, Germany; LMU Klinikum, Department of Anaesthesiology, Ludwig Maximilian University of Munich, Munich, Bavaria, Germany
Hu	Zhaoxian		Institute for Clinical and Translational Science, University of California, Irvine, Irvine, CA, USA
Ir	Diana		Data Science to Patient Value Program, University of Colorado Anschutz Medical Campus, Aurora, CO, USA
Jiang	Xiaoqian		School of Biomedical Informatics, The University of Texas Health Science Center at Houston, Houston, TX, USA
Kim	Katherine	K	Betty Irene Moore School of Nursing, University of California, Davis, Sacramento, CA, USA; School of Medicine, Department of Public Health Sciences, Health Informatics Division, University of California Davis, Sacramento, CA, USA
Kim	Jihoon		UC San Diego Health Department of Biomedical Informatics, University of California, San Diego, La Jolla, CA, USA

Last Name	First Name	Middle Name	Affiliation(s)
Knight	Tara	K	USC Schaeffer Center for Health Policy and Economics, Price School of Policy, University of Southern California, Los Angeles, CA, USA
Koola	Jejo	D	UC San Diego Health Department of Biomedical Informatics, University of California, San Diego, La Jolla, CA, USA
Kuo	Tsung-Ting		UC San Diego Health Department of Biomedical Informatics, University of California, San Diego, La Jolla, CA, USA
Lee	Nelson		Academic Research Systems, University of California, San Francisco, San Francisco, CA, USA
Mansmann	Ulrich		Institute for Medical Information Processing, Biometry, and Epidemiology (IBE), Ludwig Maximilian University of Munich, Munich, Bavaria, Germany
Matheny	Michael	E	GRECC, Tennessee Valley Healthcare System, Nashville, TN, USA; Department of Biomedical Informatics, Biostatistics, and Medicine, Vanderbilt University Medical Center, Nashville, TN, USA
Meeker	Daniella		Department of Preventive Medicine, Keck School of Medicine, Los Angeles, CA, USA; USC Schaeffer Center for Health Policy and Economics, Price School of Policy, University of Southern California, Los Angeles, CA, USA
Mou	Zongyang		UC San Diego Health Department of Biomedical Informatics, University of California, San Diego, La Jolla, CA, USA; Department of Surgery, UC San Diego School of Medicine, San Diego, CA, USA
Murphy	Robert	E	School of Biomedical Informatics, The University of Texas Health Science Center at Houston, Houston, TX, USA
Neumann	Larissa		UC San Diego Health Department of Biomedical Informatics, University of California, San Diego, La Jolla, CA, USA; Institute for Medical Information Processing, Biometry, and Epidemiology (IBE), Ludwig Maximilian University of Munich, Munich, Bavaria, Germany
Nguyen	Nghia	H	UC San Diego Health Department of Biomedical Informatics, University of California, San Diego, La Jolla, CA, USA; Division of Gastroenterology, UC San Diego Health, La Jolla, CA, USA
Nick	Anderson		School of Medicine, Department of Public Health Sciences, Health Informatics Division, University of California, Davis Health, Sacramento, CA, USA
Niedermayer	Sebastian		LMU Klinikum, Department of Anaesthesiology, Ludwig Maximilian University of Munich, Munich, Bavaria, Germany
Ohno-Machado	Lucila		UC San Diego Health Department of Biomedical Informatics, University of California, San Diego, La Jolla, CA, USA; Veteran Affairs San Diego Healthcare System, San Diego, CA, USA.
Park	Eunice		Enterprise Information Services, Cedars Sinai Medical Center, Los Angeles, CA, USA
Paul	Paulina		UC San Diego Health Department of Biomedical Informatics, University of California, San Diego, La Jolla, CA, USA
Perkins	Amy	M	Department of Biostatistics, Vanderbilt University Medical Center, Nashville, TN, USA; GRECC, Tennessee Valley Healthcare System, Nashville, TN, USA
Pletcher	Mark	J	Department of Epidemiology and Biostatistics, University of California, San Francisco, San Francisco, CA, USA
Post	Kai	W	UC San Diego Health Department of Biomedical Informatics, University of California, San Diego, La Jolla, CA, USA
Rieder	Clemens		LMU Klinikum, Department of Anaesthesiology, Ludwig Maximilian University of Munich, Munich, Bavaria, Germany
Scherer	Clemens		LMU Klinikum, Medizinische Klinik und Poliklinik I, Ludwig Maximilian University of Munich, Munich, Bavaria 81377, Germany
Schilling	Lisa	M	Data Science to Patient Value Program, University of Colorado Anschutz Medical Campus, Aurora, CO, USA
Soares	Andrey		Data Science to Patient Value Program, University of Colorado Anschutz Medical Campus, Aurora, CO, USA
SooHoo	Spencer		Division of Informatics, Department of Biomedical Sciences, Cedars Sinai Medical Center, Los Angeles, CA, USA

Last Name	First Name	Middle Name	Affiliation(s)
Soysal	Ekin		School of Biomedical Informatics, The University of Texas Health Science Center at Houston, Houston, TX, USA
Steven	Covington		IT Health Informatics, University of California Davis Health, Sacramento, CA, USA
Tep	Brian		Enterprise Information Services, Cedars Sinai Medical Center, Los Angeles, CA, USA
Toy	Brian		University of Southern California, Keck School of Medicine, Los Angeles, CA, USA
Wang	Baocheng		LMU Klinikum, Department of Anaesthesiology, Ludwig Maximilian University of Munich, Munich, Bavaria, Germany
Wu	Zhen	R	IT Health Informatics, University of California, Davis Health, Sacramento, CA, USA
Xu	Hua		School of Biomedical Informatics, The University of Texas Health Science Center at Houston, Houston, TX, USA
Yong	Choi		School of Medicine, Department of Public Health Sciences, Health Informatics Division, University of California, Davis, Sacramento, CA, USA
Zheng	Kai		Department of Informatics, Donald Bren School of Information and Computer Sciences, University of California, Irvine, Irvine, CA, USA
Zhou	Yujia		School of Biomedical Informatics, The University of Texas Health Science Center at Houston, Houston, TX, USA
Zucker	Rachel	A	Data Science to Patient Value Program, University of Colorado Anschutz Medical Campus, Aurora, CO, USA