Classifying COVID-19 Related Hospitalizations and Deaths in the Collaborative Cohort of Cohorts for COVID-19 Research (C4R) Study

Authors: Akshaya Krishnaswamy, Rafail Rustamov, Tauqeer Ali, Norrina B. Allen, Alexander Arynchyn, Pallavi Balte, Lori A. Bateman, Jianwen Cai, Harry Chang, Lucas Chen, Mitchell S.V. Elkind, James S. Floyd, Kelley P. Gabriel, Sina A. Gharib, Karen Hinckley Stukovsky, Carmen R. Isasi, Lauren Jager, Ling Jin, Suzanne E. Judd, Alka Kanaya, Maureen R. Kelly, Sadiya S. Khan, Anna Kucharska-Newton, Joyce S. Lee, Emily B. Levitan, Cora E. Lewis, Barry J. Make, Dave Mauger, Charles G. Murphy, Victor E. Ortega, Ashmi A. Patel, Amber Pirzada, Elizabeth A. Regan, Kimberly B. Ring, Wayne Rosamond, David Schwartz, Daniela Sotres-Alvarez, Cheryl Tarlton, Janis Tse, Elman Urbina, Maya Vankineni, Sally E. Wenzel, Prescott Woodruff, Jihyun Yang, Neil A. Zakai, Ying Zhang, Wendy S. Post, Elizabeth C. Oelsner

Affiliations:

Neil A. Zakai*[To be added in]* Cora E. Lewis, MD, **MSc:** MSPH, Department of **Medicine, Department of Pathology and Laboratory Medicine, Larner College of Medicine at the University of Vermont, Burlington, VT, USA.** *[To be added in]*Epidemiology, School of Public Health, University of Alabama at Birmingham, Birmingham, Alabama, USA

Lori A. Bateman, MS, Collaborative Studies Coordinating Center, Department of Biostatistics, Gillings School of Global Public Health, University of North Carolina at Chapel Hill, Chapel Hill, NC

Keywords: COVID-19, coronavirus disease 2019, epidemiology, cohort study

Abstract word count: 200/200

Word count: 3617/4000

Abstract

To disentangle risk factors versus sequelae of COVID-19, studies combining pre-pandemic and COVID-19 measures are needed. We evaluated a protocol to ascertain and adjudicate COVID-19 hospitalizations, deaths, and cardiopulmonary complications in the US general population-based Collaborative Cohort of Cohorts for COVID-19 Research (C4R) Study. This report describes findings from the first three years of the study. Among 49,790 participants, 1,974 potential COVID-19 hospitalizations and deaths were ascertained, 1,523 (77%) medical records were obtained, and 1,237 (63%) were adjudicated. Of 1,030 (88%) records with confirmed COVID-19 that were available for analysis, 77% were diagnosed as severe and 31% as critical COVID-19; COVID-19 was not the cause of hospitalization for 7%. During COVID-19 hospitalization, pneumonia was the most frequent complication (80%) compared to myocardial infarction (5%), pulmonary embolus (5%), deep venous thrombosis (3%), and stroke (2%); acute kidney injury occurred in 34%. Interrater reliability was high (kappa 0.85-1.00) except for myocardial infarction (0.60). Compared to adjudication, sensitivity of discharge diagnosis codes for pneumonia was higher (97%) than other complications (48-67%). Hence, burden of COVID-19 hospitalization and death was high in C4R, and protocolized review was effective to confirm and characterize outcomes, yet innovative approaches to expedite review of medical records are needed.

Introduction

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infections (herein referred to as coronavirus disease 19 [COVID-19] infections) have been a leading cause of hospitalizations and deaths since 2020.[1,2] Although rates of hospitalized and fatal COVID-19 have declined over the course of the pandemic, understanding risk factors for hospitalized COVID-19 illness remains of major public health importance. Furthermore, more severe acute illness has been associated with increased risks of post-acute sequelae of SARS-CoV-2 infection (PASC), which has been estimated to impact up to 57% of hospitalized COVID-19 patients.[3] Large population-based studies with accurate characterization of hospitalized COVID-19 and its acute complications – as well as clinical, biomarker, and lifestyle data from before and after infection – are urgently needed to understand the mechanisms of acute disease and post-acute sequelae of COVID-19.

Epidemiologic cohort studies provide important opportunities to untangle risk factors from sequelae of COVID-19. Many extant US cohort studies have collected extensive data on clinical and subclinical diseases and their risk factors, including behavior, cognition, biomarkers, and social determinants of health. Since enrollment for these studies was completed prior to the COVID-19 pandemic, their data support the analysis of incident COVID-19 while minimizing the referral, survival, and recall biases that are common to COVID-19 case series and disease-based studies. However, in the absence of a national healthcare system, most longstanding epidemiology studies have not established comprehensive linkages to electronic health records (EHRs) and have depended instead on complex systems of medical records ascertainment and adjudication.

To support robust epidemiology research on acute COVID-19 and PASC in NIH-funded cohort studies, we developed a protocol to ascertain and adjudicate hospitalized and fatal COVID-19 cases as part of the Collaborative Cohort of Cohorts for COVID-19 Research (C4R). This report describes the development of the protocol, which leveraged experience in the classification of cardiovascular events, and preliminary experience applying the protocol from February 2021 to June 2023, including comparison of ICD-based versus protocol-based classification for COVID-related fatal and non-fatal hospitalizations.

Methods

Study Design

C4Ris a national prospective meta-cohort study of adult participants from 14 long-standing cardiovascular, neurological, and respiratory cohort studies[4]: Atherosclerosis Risk in Communities (ARIC)[5], Coronary Artery Risk Development in Young Adults (CARDIA)[6], Genetic Epidemiology of COPD (COPDGene)[7], Framingham Heart Study (FHS)[8], Hispanic Community Health Study/Study of Latinos (HCHS/SOL)[9-11], Jackson Heart Study (JHS)[12-14], Mediators of Atherosclerosis in South Asians Living in America (MASALA)[15,16], Multi-Ethnic Study of Atherosclerosis (MESA)[17], Northern Manhattan Study (NOMAS)[18], Prevent Pulmonary Fibrosis (PrePF)[19], Reasons for Geographic and Racial Differences in Stroke (REGARDS)[20], Severe Asthma Research Program (SARP)[21], Subpopulations and Intermediate Outcome Measures in COPD Study (SPIROMICS)[22], and the Strong Heart Study (SHS)[23,24]. Details on each of the component cohorts are provided in **Supplementary Table 1**.

C4R received funding in 2020 to perform standardized prospective data collection on COVID-19 and to harmonize pre-pandemic deep phenotyping available in the cohorts. Data collection on COVID-19 was accomplished by two waves of questionnaires, a SARS-CoV-2 serosurvey, and ascertainment and adjudication of COVID-19 hospitalizations and deaths, hereafter referred to as "COVID-19 events," which is the subject of this report.

The study was approved by the Institutional Review Boards for all participating cohorts and all participants provided written informed consent for participation in each cohort.

Protocol Development and Coordination

The C4R Events Subcommittee was comprised of investigators and other cohort personnel with experience in the ascertainment of clinical events in at least one of the C4R cohorts. From October to December 2020, the Subcommittee met regularly to develop a protocol modeled on longstanding cohort protocols for the ascertainment and classification of cardiovascular, respiratory, and thromboembolic events[5,6,8,10,11,15,17,18,24-27]. The final protocol (available at www.c4r-nih.org) was approved by the C4R Cohort Coordinating Committee (CCC) in February 2021. Relevant administrative and review forms were coded into the C4R Events REDCap[28,29] toolkit, which was made available to the cohorts via a central (Columbia) instance or for local adaptation by cohort data coordinating centers (DCCs). For the central instance of the C4R Events REDCap, cohort-specific Data Access Groups (DAGs) ensured that each cohort was only able to enter or access its own participant data. Cohorts provided bimonthly tracking reports regarding all cohort-specific elements of protocol completion to the Data Coordination and Harmonization Center (DCHC), which integrated these reports and generated regular Events Protocol status reports to the Events Subcommittee, the CCC, and funding agencies. De-identified COVID-19 events classification data were made available for analysis on the C4R Analysis Commons as well as being transferred back to cohort-specific DCCs for cohort use and transfer to other data sharing and analysis platforms (e.g., NHLBI BioData Catalyst, Trans-Omics for Precision Medicine Project Exchange Area).

COVID-19 events Ascertainment

Potential COVID-19 hospitalizations and deaths were ascertained by cohorts via several mechanisms. C4R questionnaires, which were administered in two waves across all 14 cohorts, included questions regarding hospitalization for COVID-19 (**Supplementary Table 2**). If necessary, as in cases of participant dementia or death, these questionnaires were administered to a participant's proxy. Additional COVID-19-related hospitalizations and deaths were identified by regular non-C4R follow-up calls conducted in 8 cohorts (ARIC, CARDIA, FHS, HCHS/SOL, MESA, NOMAS, REGARDS, SHS) to collect information on all-cause hospitalization and vital status. These data were supplemented with information collected at in-person exams, which were conducted in all cohorts during the pandemic period, with the exception of NOMAS. Various non-questionnaire methods of ascertainment, such as active surveillance of local EHR systems and other sources (e.g., obituaries) were performed by ARIC, FHS, REGARDS, SARP, SHS, as well as selected clinical sites in CARDIA, COPDGene, and MESA. Most cohorts supplemented vital status data with National Death Index (NDI) searches, although these are subject to considerable reporting lags compared to other ascertainment methods and were not available for the pandemic period at the time of this report.

After confirming participant or next-of-kin consent, cohort staff requested copies of medical records for all potential COVID-19-related hospitalizations and deaths, including physician notes (admission, consultation, discharge), radiology and laboratory reports, electrocardiogram reports, and discharge diagnoses (including discharge diagnosis ICD codes). Where applicable, death certificates were obtained. Medical records submitted for central review at the DCHC were de-identified prior to secure file transfer.

Protocolized medical record review and adjudication

Hospitalizations (both fatal and non-fatal) and out-of-hospital deaths were eligible for protocolized medical record review if they were assigned, in any position, any of the following COVID-19-related discharge diagnoses based on the International Classification of Diseases, Tenth Revision (ICD-10), Clinical Modification codes[30]: confirmed COVID-19 (U07.1), post-infectious state after COVID-19 (U09.9), Multisystem inflammatory syndrome associated with COVID-19 (M35.81), Personal History of COVID-19 (Z86.16), Pneumonia due to coronavirus disease 2019 (J12.82), Other viral pneumonia (J12.89), or, for events occurring prior to May 1 2020, other coronavirus (B97.29). In the absence of these discharge diagnoses, records were also considered eligible for review if there was evidence of a positive COVID-19 test, physician suspicion of COVID-19, or next-of-kin interview indicating suspected or known COVID-19 infection. Episodes of treatment in the Emergency Department (ED) for over 24 hours were classified as hospitalizations since many hospitals used the ED for inpatient care during the heights of COVID-19 surges due to lack of sufficient inpatient capacity.

Eligible COVID-19-related hospitalizations (fatal and non-fatal) and out-of-hospital deaths were subjected to structured review by physicians with experience in treating hospitalized COVID-19. Reviewers were trained via webinar and individualized instruction. Three cohorts (FHS, REGARDS, SARP) elected to perform review at the cohort level by local adjudicators, while records from the remaining eleven cohorts (ARIC, CARDIA, COPDGene, HCHS/SOL, JHS, MASALA, MESA, NOMAS, PrePF, SPIROMICS, SHS) were reviewed centrally at the C4R DCHC.

The structured review included data abstraction, including information on oxygenation levels and medication administration, followed by assignment of COVID-19 diagnoses. Criteria for defining probable or definite diagnoses are shown in **Table 1**. Definitions of severe and critical disease were based on NIH COVID-19 treatment guidelines.[31] The certainty of diagnosis is partly determined by reviewer judgement, yet there are explicit instructions on types of evidence that should be used to support probable versus definite diagnoses. Treating physician notes may be sufficient for a probable diagnosis, but definite diagnosis generally requires additional objective evidence. For example, definite diagnoses for COVID-19 pneumonia, pulmonary embolus, deep vein thrombosis, or stroke required radiologic evidence in the medical record. Definite diagnosis of acute kidney injury was based on reported creatinine or new initiation of renal replacement therapy. Definite COVID-19 MI was defined to identify cases of MI caused by acute atherothrombotic coronary artery disease, or "Type 1" MI, in the context of COVID-19 infection.[32] Probable COVID-19 MI was defined more broadly and likely captures Type 1 and Type 2 MI. For comparison, myocardial injury was defined as a maximum recorded troponin level that was greater than two times the upper limit of normal (ULN).

Following completion of the review, reviewers were offered the option of requesting a second independent review for challenging cases. A random 10% subset of records was also submitted for a second independent review.

Analysis

Characteristics of C4R participants with and without an ascertained or adjudicated COVID-19 event were tabulated. The incidence of events ascertained and adjudicated per month (based on date of event) was plotted. Since, upon medical records review, it was found that out-of-hospital deaths did not contain sufficient data for adjudication, further analyses were limited to fatal and non-fatal COVID-19 hospitalizations. Hospitalizations among participants who did not consent to data sharing on the C4R Analysis Commons were also excluded from analysis. Among fatal and non-fatal hospitalizations that were successfully adjudicated, the incidence of COVID-19-related diagnoses was tabulated, by level of certainty. Interrater agreement was assessed via positive agreement, negative agreement, and the Cohen's κ-statistic[33]. The performance of discharge diagnosis ICD code-based classification was compared against definite C4R diagnoses, which were treated as the reference standards; probable diagnoses were excluded from these comparisons because they included discharge diagnosis ICD codes in the diagnostic criteria. Sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV) were calculated. Also, myocardial injury was compared to adjudicated diagnoses of definite COVID-19 MI. Analyses were performed in SAS Studio software (SAS Institute, Cary, NC) on the C4R Analysis Commons. Adjudicated events data are available on the C4R Analysis Commons for investigators with the necessary C4R and cohort approvals.

Results

COVID-19 events ascertainment

Among 49,790 C4R participants, 1,974 potential COVID-19-related hospitalization and/or deaths among 1,772 (3.6%) participants were ascertained between January 2020 and June 2023 (**Figure 1**), for an estimated incidence density rate of 11.9 per 1,000 person-years of follow-up. Overall, the cohorts reported that half (53%) of these events were identified by C4R questionnaire. As of June 2023, 1,768 (90%) medical records were requested, of which 1,523 (86%) were obtained. The most common reasons for failure to obtain medical records were inability to obtain participant consent and non-response from the hospital.

As of the writing of this report (October 1st, 2023), 1,237 of 1,974 (63%) events were adjudicated over 28 months. After exclusion of 57 out-of-hospital deaths due to incomplete information and 13 non-fatal hospitalizations with consent restrictions, there were 1,167 (59% of the 1,974) events available for analysis on the C4R Analysis Commons, of which 1,030 had evidence of COVID-19 infection (52% of potential COVID-19 hospitalizations and/or deaths) (**Figure 2**).

Table 2 compares socio-demographic and clinical characteristics of 1,772 participants among whom a potential COVID-19 hospitalization was ascertained, including those for whom events were (N=1098) and were not yet (N=674) adjudicated as of October 2023, versus those with no COVID-19 hospitalization or death ascertained through June 2023 (N=48,018). Compared to participants without a potential COVID-19 hospitalization or death, those with an ascertained event were more likely to be male, to report American Indian ancestry, and to have a history of pre-pandemic smoking, diabetes, or hypertension; they were less likely to have education beyond college and less likely to be vaccinated compared to those without ascertained potential COVID-19 events. Notably, among participants with

ascertained potential COVID-19 events, characteristics of those with adjudicated records were generally similar to those pending adjudication. More than one COVID-19 hospitalization was ascertained in 50 (3%) participants.

Adjudication of medical records for potential COVID-19 hospitalizations

There were a total of 7 reviewers at the C4R DCHC, 7 at FHS, 3 at REGARDS, and 2 at SARP. The median length of medical records was 59 pages (IQR: 37, 105). The average time to review was approximately 30 minutes per record.

SARS-CoV-2 infection was confirmed as probable in 1% and definite in 87% of adjudicated events. Among 1,030 hospitalized events in which SARS-CoV-2 infection was confirmed, COVID-19 illness was diagnosed as the cause of 952 (93%) of these events. **Table 3** describes the adjudicated probable and definite diagnoses of SARS-CoV-2 infection, acute COVID-19 illness severity, and acute COVID-19 complications following the criteria described in **Table 1**. 77% were diagnosed with severe COVID-19, 31% were diagnosed with critical COVID-19, 80% with COVID-19-associated pneumonia, and 34% with COVID-19-associated renal failure. Fatal hospitalization due to COVID-19 and other cardiopulmonary complications such as COVID-19-related pulmonary embolism (PE), deep venous thrombosis (DVT), stroke, and myocardial infarction (MI) were less common.

Of 139/1237 (11%) fatal and non-fatal hospitalizations that underwent a second review, there was almost perfect agreement for classification of COVID-19 infection, critical illness, stroke, pulmonary embolism, deep venous thrombosis, and fatal hospitalizations, as shown in **Table 4**. There was also strong agreement for hospitalization, severe illness, pneumonia, and renal failure, and moderate agreement for MI.

Compared to adjudicated diagnoses with a definite certainty level, discharge ICD code-based classification was highly sensitive but relatively non-specific for pneumonia (sensitivity=98%, specificity=72%), and insensitive for cardiovascular and renal complications (sensitivity 57-81%) (**Table 5**). The PPV of ICD-based classification for COVID-19-related pneumonia (ICD code J12.82, pneumonia due to coronavirus disease 2019 or J12.89, other viral pneumonia), acute kidney injury (ICD code N17, Acute kidney injury), pulmonary embolism (ICD Code I26, pulmonary embolism), DVT (ICD Code I82, deep venous thrombosis) and stroke (ICD code I63, Cerebral infarction) was excellent, ranging from 81-93%; however, the PPV for COVID-19-related myocardial infarction (ICD code I21, Acute myocardial infarction) was 33%.

A substantial number of adjudicated COVID-19-related hospitalizations and deaths were found to have myocardial injury, but many of them were not confirmed as COVID-19-related MI by adjudication. Among the 1,030 hospitalized events in which SARS-CoV-2 infection was confirmed, 170 (25%) demonstrated myocardial injury. Of these 170 events with myocardial injury, 25 (15%) were assigned an ICD code for MI, 5 (3%) were adjudicated as definite COVID-19-related MI, and 38 (22%) were adjudicated as probable COVID-19-related MI; the majority (75%) were not confirmed as definite or probable COVID-19-related MI.

Discussion

COVID-19-related hospitalizations and deaths were ascertained and adjudicated in a large US meta-cohort of adults using a new classification protocol. Despite challenges in obtaining medical records, implementation of the adjudication protocol was feasible, interrater agreement was excellent, and comparison of adjudicated diagnoses with questionnaire-based ascertainment and discharge diagnosis codes suggested advantages of adjudication with respect to false-positive and false-negative misclassification. Hence, protocolized review may be an effective and robust approach to defining COVID-19-related hospitalizations in epidemiologic studies, yet innovative approaches to expedite records ascertainment are needed.

Over three years, C4R ascertained at least one potential COVID-19-related hospitalization or out-of-hospital death in 3.6% of participants. This is likely to be an underestimate of the actual cumulative incidence of severe COVID-19 outcomes in the sample, since we experienced limitations with respect to obtaining information on out-of-hospital deaths; we are endeavoring to overcome this issue by leveraging NDI data, yet much of this data was not yet available at the time of this report. Nonetheless, this highlights the major impact of COVID-19 on cohort participants. For comparison, the incidence per 1,000 person-years of follow-up of severe COVID-19 in C4R was estimated at 11.9, which is similar to an incidence of 9.7[34] for atherosclerotic cardiovascular disease events in a recent report using 8 of the C4R cohorts. The adjudication of severe COVID-19 events in C4R will support important research on risk and resilience factors for SARS-CoV-2 infection, and potentially other viral respiratory pathogens. It will also allow for robust investigation of potential long-term sequelae of severe COVID-19 in cohort participants who continue to undergo longitudinal follow-up via in-person exams and cardiopulmonary events surveillance, with additional information available regarding pre-pandemic conditions ascertained prospectively.

Our adjudication results highlight certain limitations of standard cohort events surveillance for ascertainment of COVID-19-related hospitalizations. SARS-CoV-2 infection could not be confirmed in one in eight events that were identified as potentially COVID-19-related. These cases could have been due to incorrect self-reporting of a COVID-19 hospitalization on a C4R or cohort questionnaire, or active surveillance methods that were more sensitive than specific. Of note, some cases where infection could not be confirmed may have been true infections, but the medical records lacked sufficient detail for adjudication. Hence, rather than censoring all non-confirmed events, C4R is assigning a certainty level for all its COVID-19 outcomes, so that investigators can select the outcome most suitable for their specific research needs.

Our findings also illustrate the limitations of discharge diagnosis codes to define hospitalization "for" versus "with" SARS-CoV-2 infection. Of hospitalizations with confirmed SARS-CoV-2 infection, symptoms of COVID-19 illness did not contribute to the hospitalization in 7% of cases. These findings are similar to prior reports examining the role of SARS-CoV-2 in hospitalizations using EHR systems. Adjudicated C4R endpoints will allow investigators to exclude these cases of hospitalization with incidental COVID-19 from studies designed to examine risks and sequelae of severe COVID-19 illness.

We also found that discharge diagnosis codes were insensitive for cardiopulmonary and renal complications. We confirmed a substantial number of cases of COVID-19 pneumonia, acute kidney injury, and MI among records without a corresponding discharge diagnosis code. This may be related, in part, to the major challenges of charting during COVID-19 surges, when documentation standards were modified to prioritize direct patient care activities. Nonetheless, before the pandemic period, previous

investigations on the use of administrative data to identify cases of pneumonia have found that ICD codes are imprecise and can result in a substantial number of pneumonia cases going undetected[35,36]. These results are supported by earlier findings that ICD-10 code N17 often misses acute kidney injury during hospitalizations for kidney transplant patients[37]. Potential underestimation of the incidence of MI using claims or administrative data has also been well-described.[38-41] This has been one justification for longstanding—albeit, labor- and time-intensive—ASCVD events adjudication programs in many of the C4R cohorts.

As expected, we found a substantial proportion of hospitalized events included evidence of myocardial injury, defined by troponin values two times greater than the upper limit of normal; however, only a small subset of these cases were confirmed as COVID-19 MI upon physician review using our C4R protocol definitions. We elected to restrict definite MI to confirmed cases of type 1 (ST-elevation MI) or type 3 MI (MI resulting in death with pathological evidence of MI). It may be difficult to differentiate type 2 MI/supply-demand mismatch, which requires documentation of a rise and fall in troponin levels and symptoms of ischemia, from myocardial injury (at least one elevated troponin level), based solely on review of obtained medical records. Prior studies demonstrated that elevated troponin values are common in patients hospitalized with COVID-19,[42-44] and may occur with conditions other than myocardial infarction, such as cardiomyopathy, acute cor pulmonale, arrhythmias, or cardiogenic shock.[45] Our results suggest that the true incidence of type 1 MI, due to plaque disruption and thrombosis leading to coronary occlusion, is rare, supporting the utility of physician review in validating these endpoints for COVID-19 cardiovascular research.

Limitations

Strengths of study include prospective ascertainment of potential COVID-19-related hospitalizations and deaths in a well-characterized, multi-ethnic, US community-based sample of adults that is relatively free of referral, survival, and recall biases. The adjudication protocol, which was modeled on gold-standard cardiopulmonary events adjudication in many of the C4R cohorts, was fully standardized across the cohorts and implemented at scale to generate robust events data for analysis within three years of program initiation. Nonetheless, certain limitations must be considered. Medical records have not yet been obtained for 37% of cases due to lack of participant consent, non-response from hospital systems, or other operational delays that are common to cohort events ascertainment operations. This highlights the need for novel approaches for cohorts to access medical records, such as the consenting of participants to share their own electronic medical records—now available to patients via the 21st Century Cures Act[46]—with cohorts.[46] The characteristics of participants with missing versus available medical records in C4R were generally comparable, so there was no clear evidence of selection bias. Some medical records had incomplete data available for adjudication, which could be related to relaxation of documentation requirements during the pandemic. Few out-of-hospital deaths were ascertained, and medical records for out-of-hospital deaths were often incomplete, due to various reasons, including lack of court documents from the family of the deceased to obtain records, or hospital decision to not grant access to the decedent's medical records for research purposes. To address this limitation, additional information on out-of-hospital deaths associated with COVID-19 will be obtained from the NDI, which provides complete ascertainment of deaths in the US, including ICD-codes. As of October 2023, 37% of records had not yet been adjudicated, reflecting the time commitment needed for this activity; however, adjudication activities are continuing.

Conclusions

Ascertainment and adjudication of COVID-19 hospitalizations and deaths in longstanding NIH-funded cohort studies was feasible, although there were significant obstacles to obtaining medical records from the pandemic period. Adjudication of potential COVID-19 hospitalizations and deaths confirmed SARS-CoV-2 infection in 88% of events and found that infection may have been incidental to the hospitalization in 7%. Compared to adjudication, discharge diagnosis codes were insensitive for acute cardiovascular and renal complications of COVID-19. Hence, protocolized review may be useful to identify and validate COVID-19-related hospitalizations and inpatient complications, but novel approaches to medical records access and linkage would augment opportunities for COVID-19 and other health outcomes research in NIH-funded epidemiologic cohorts.

Table 1. Criteria for classification of COVID-19 diagnoses as definite or probable based on medical record review.

Classification	Probable	Definite
SARS-CoV-2 Infection	Any of the following: • Physician suspicion • Informant suspicion • Acute COVID-19-like illness (fever, cough, and/or shortness of breath without other identifiable cause)	Any of the following • ICD code for SARS-CoV-2 infection or related illness (U07.1, U09.9, M35.81, J12.82) • Positive PCR or antigen test in medical record
COVID-19 illness severity		
COVID-19 Hospitalization	Probable infection AND any of the following: • Admitted to the hospital for COVID-19-related signs or symptoms; or, developed COVID-19-related signs or symptoms during hospitalization • Deceased in Emergency Department	Definite infection AND any of the following: • Admitted to the hospital for COVID-19-related signs or symptoms; or, developed COVID-19-related signs or symptoms during hospitalization • Deceased in Emergency Department
COVID-19 Severe Illness	Probable infection AND any of the following: • Oxygen saturation < 94% (on room air or on O2 supplementation) at any time, including pre-admission • Respiratory rate > 30 breaths/minute • Required BiPAP, HFNC, MV, and/or ECMO	Definite infection AND any of the following: • Oxygen saturation < 94% (on room air or on O2 supplementation) at any time, including pre-admission • Respiratory rate > 30 breaths/minute • Required BiPAP, HFNC, MV, and/or ECMO
COVID-19 Critical Illness	Probable infection AND any of the following: • High-flow oxygen • Mechanical ventilation or ECMO • Inotropes or pressors • ICU or step down unit admission	Definite infection AND any of the following: • High-flow oxygen • Mechanical ventilation or ECMO • Inotropes or pressors
Fatal COVID-19	Probable infection AND: • Probable COVID-19 hospitalization ending in death, *OR* • Physician or other informant suspicion of COVID-19 contributing to death, *OR* • Acute COVID-19-like illness (fever, cough, and/or shortness of breath without other identifiable cause) without positive COVID-19 test within 28 days of death, *OR* • Judged by reviewer that COVID-19 is probable cause of death	Definite infection AND: • Death certificate with ICD: o U07.1, Confirmed COVID-19 o U09.9, Post-infectious state after COVID-19 o M35.81, Multisystem inflammatory syndrome associated with COVID-19 o J12.82, Pneumonia due to coronavirus disease 2019, *OR* • Definite COVID-19 hospitalization ending in death, *OR* • Judged by reviewer that COVID-19 is definite cause of death
COVID-19 complications		
COVID-19 Pneumonia	Probable infection AND any of the following: • ICD code for COVID-19 pneumonia (J12.82, J12.89)	Definite infection AND any of the following: • Chest CT with pneumonia • Chest X ray with pneumonia

	• Physician documentation of pneumonia	
COVID-19 MI	Probable infection AND: • ICD-10 I21, *OR* • Physician documentation of NSTEMI, *OR* • Physician documentation of STEMI	Definite infection AND: • ST Elevation on electrocardiogram consistent with STEMI and troponin greater than the upper limit of normal (if available), *or* • Pathologic evidence of acute myocardial infarction
COVID-19 Stroke	Probable infection AND: • ICD-10 I63, *OR* • Physician documentation of stroke	Definite infection AND: • CT head consistent with stroke, *OR* • MRI brain consistent with stroke
COVID-19 PE	Probable infection AND: • ICD-10 I26, *OR* • Physician note documenting presumptive PE with initiation of anticoagulation	Definite infection AND: • CTA indicative of PE, *OR* • V/Q indicative of PE
COVID-19 DVT	Probable infection AND: • ICD-10 I82, *OR* • Physician note documenting new venous thrombosis	Definite infection AND: • Lower extremity Doppler showing DVT
COVID-19 Renal Failure	Probable infection AND: • ICD-10 N17, *OR* • Physician documentation of acute kidney injury (AKI)	Definite infection AND: • Max/min or Max/baseline creatinine > 1.5, *OR* • New initiation of renal replacement therapy

Table 2. Baseline characteristics of C4R participants according to ascertainment and adjudication of COVID-19 hospitalization or death, February 2020 – June 2023.

Characteristic	No event	Ascertained event, pending adjudication	Adjudicated event
Participants, n (%)	48018	674	1098
COVID-19 hospitalizations and deaths, n		737	1237
Sex			
Male	19932 (42)	304 (48)	456 (46)
Female	28019 (58)	326 (52)	543 (54)
Race/Ethnicity			
Asian	1136 (2)	9 (1)	8 (1)
American Indian and Alaskan Native	1456 (3)	38 (6)	248 (23)
Black	10733 (22)	210 (33)	209 (19)
White	22347 (47)	174 (27)	457 (43)
Hispanic	12192 (25)	207 (32)	248 (23)
Other	76 (0)	1 (0)	0 (0)
Age			
18-30 years	565 (1)	5 (1)	3 (0)
31-64 years	16673 (35)	167 (28)	182 (21)
65+ years	29888 (63)	417 (71)	695 (79)
Cohort			
ARIC	5560 (12)	53 (8)	311 (28)
CARDIA	2798 (6)	29 (4)	27 (2)
COPDGene	4109 (9)	33 (5)	49 (4)
FHS	3300 (7)	15 (2)	52 (5)
HCHS/SOL	11021 (23)	132 (20)	191 (17)
JHS	2279 (5)	100 (15)	0 (0)
MASALA	572 (1)	1 (0)	0 (0)
MESA	3268 (7)	189 (28)	47 (4)
NOMAS	909 (2)	23 (3)	67 (6)
PrePF	622 (1)	3 (0)	3 (0)
REGARDS	10126 (21)	20 (3)	179 (16)
SARP	508 (1)	2 (0)	4 (0)
SPIROMICS	1553 (3)	23 (3)	9 (1)
SHS	1393 (3)	51 (8)	159 (14)
Geographic Region			
Middle Atlantic	7804 (16)	145 (22)	274 (25)
Midwest	9605 (20)	140 (21)	154 (14)
New England	3162 (7)	15 (2)	56 (5)
South	16155 (34)	210 (32)	297 (28)
Southwest	857 (2)	5 (1)	3 (0)
West	3500 (7)	63 (10)	242 (22)
Smoking Status			

Current	6610 (14)	97 (16)	128 (13)
Former	17258 (37)	247 (40)	413 (43)
Never	23339 (49)	270 (44)	413 (43)
Body mass index			
<25 kg/m^2	11115 (24)	90 (15)	169 (18)
25-29.9 kg/m^2	17031 (37)	210 (35)	309 (33)
30-34.9 kg/m^2	10727 (23)	165 (27)	262 (28)
>35 kg/m^2	7493 (16)	139 (23)	206 (22)
Education			
Less than high school	5608 (12)	109 (18)	178 (19)
High school	10820 (23)	151 (25)	273 (29)
College	9081 (20)	108 (18)	180 (19)
Beyond college	20969 (45)	239 (39)	316 (33)
Medical History			
Diabetes	9099 (19)	173 (28)	301 (31)
Hypertension	23580 (50)	380 (62)	605 (63)
Cardiovascular disease	4712 (11)	80 (14)	165 (19)
Chronic obstructive pulmonary disease	3361 (9)	76 (13)	82 (12)
Asthma	4856 (14)	73 (13)	119 (17)
Received COVID-19 vaccine	28837 (91)	265 (87)	455 (82)
Infection wave during which event occurred			
First wave (WT, NE)	1133 (2)	155 (23)	173 (16)
Second wave (WT, rest of US)	573 (1)	100 (15)	136 (12)
Third wave (Alpha, winter)	1487 (3)	159 (24)	338 (31)
Fourth wave (MI, spring)	385 (1)	42 (6)	45 (4)
Fifth wave (Delta)	809 (2)	64 (10)	131 (12)
Sixth wave (Omicron)	904 (2)	17 (2)	45 (4)
Unknown wave	1093 (2)	87 (13)	191 (17)
No infection	41634 (87)	50 (7)	39 (4)

There were missing data for some pre-pandemic characteristics, in which case percentages are reported based on the number of non-missing observations.

Abbreviations: ARIC=Atherosclerosis Risk in Communities Study; CARDIA=Coronary Artery Risk Development in Young Adults, COPDGene=Genetic Epidemiology of Chronic Obstructive Pulmonary Disease; FHS=Framingham Heart Study; HCHS/SOL=Hispanic Community Health Study/Study of Latinos; JHS=Jackson Heart Study; MASALA=Mediators of Atherosclerosis in South Asians Living in America; MESA=Multi-Ethnic Study of Atherosclerosis; NOMAS=Northern Manhattan Study; PrePF=Preclinical Pulmonary Fibrosis; REGARDS=Reasons for Geographic and Racial Differences in Stroke; SARP=Severe Asthma Research Program; SPIROMICS=Subpopulations and Intermediate Outcome Measures in COPD Study; SHS=Strong Heart Study.

Table 3. Distribution of COVID-19-related diagnoses for adjudicated non-fatal and fatal hospitalizations.

Adjudicated COVID-19 Related Diagnoses	Eligible for Adjudication of Diagnosis, N	Certainty Level of Adjudicated Diagnoses[a]	
		Definite, N (%)	Probable, N (%)
Event with SARS-CoV-2 infection present	1167	1013 (87)	17 (1)
Event caused by COVID-19 illness	1030	913 (89)	39 (4)
Event caused by severe COVID-19 illness	1030	726 (70)	65 (6)
Event caused by critical COVID-19 illness	1030	275 (27)	44 (4)
Fatal COVID-19 hospitalization	1030	194 (19)	16 (2)
Pneumonia caused by COVID-19	1030	711 (69)	111 (11)
Renal failure caused by COVID-19	1030	240 (23)	110 (11)
PE caused by COVID-19	1030	32 (3)	16 (2)
DVT caused by COVID-19	1030	31 (3)	2 (0)
Stroke caused by COVID-19	1030	13 (1)	12 (1)
Myocardial infarction caused by COVID-19	1030	7 (1)	43 (4)

[a]Percentages displayed are of the total number of events eligible for adjudication of the COVID-19-related diagnosis.

Table 4. Interrater agreement for determination of COVID-19-related hospitalizations, deaths, and cardiopulmonary complications.

COVID-19-Related Event	Agreement*		Disagreement†	K-statistic	95% CI
	Positive	Negative			
Infection	129	9	1	0.94	0.83 – 1.00
Hospitalization	120	15	3	0.90	0.78 – 1.00
Severe illness	89	40	9	0.85	0.76 – 0.94
Critical illness	40	94	3	0.95	0.89 – 1.00
Fatal hospitalization	25	110	2	0.95	0.89 – 1.00
Pneumonia	102	29	7	0.86	0.76 – 0.96
Renal failure	45	88	6	0.90	0.83 – 0.98
Pulmonary embolism	8	130	0	1.00	1.00 – 1.00
Deep venous thrombosis	4	134	0	1.00	1.00 – 1.00
Stroke	4	134	0	1.00	1.00 – 1.00
Myocardial Infarction	4	128	5	0.60	0.28 – 0.92

Definition of abbreviations: CI = confidence interval.

Percentages are calculated out of the total number of events that underwent two reviews.

*Positive agreement is defined as both reviewers confirming the event as at least a probable COVID-19 event of interest; negative agreement is defined as both reviewers determining that the event was not a COVID-19-related event of interest.

†Disagreement is defined as one reviewer classifying the event as at least a probable COVID-19 event of interest and the other determining that it was not consistent with a COVID-19 event of interest.

Table 5. Sensitivity, specific, and predictive values for administrative codes with respect to definite diagnoses of COVID-19 and relevant cardiopulmonary complications.[a]

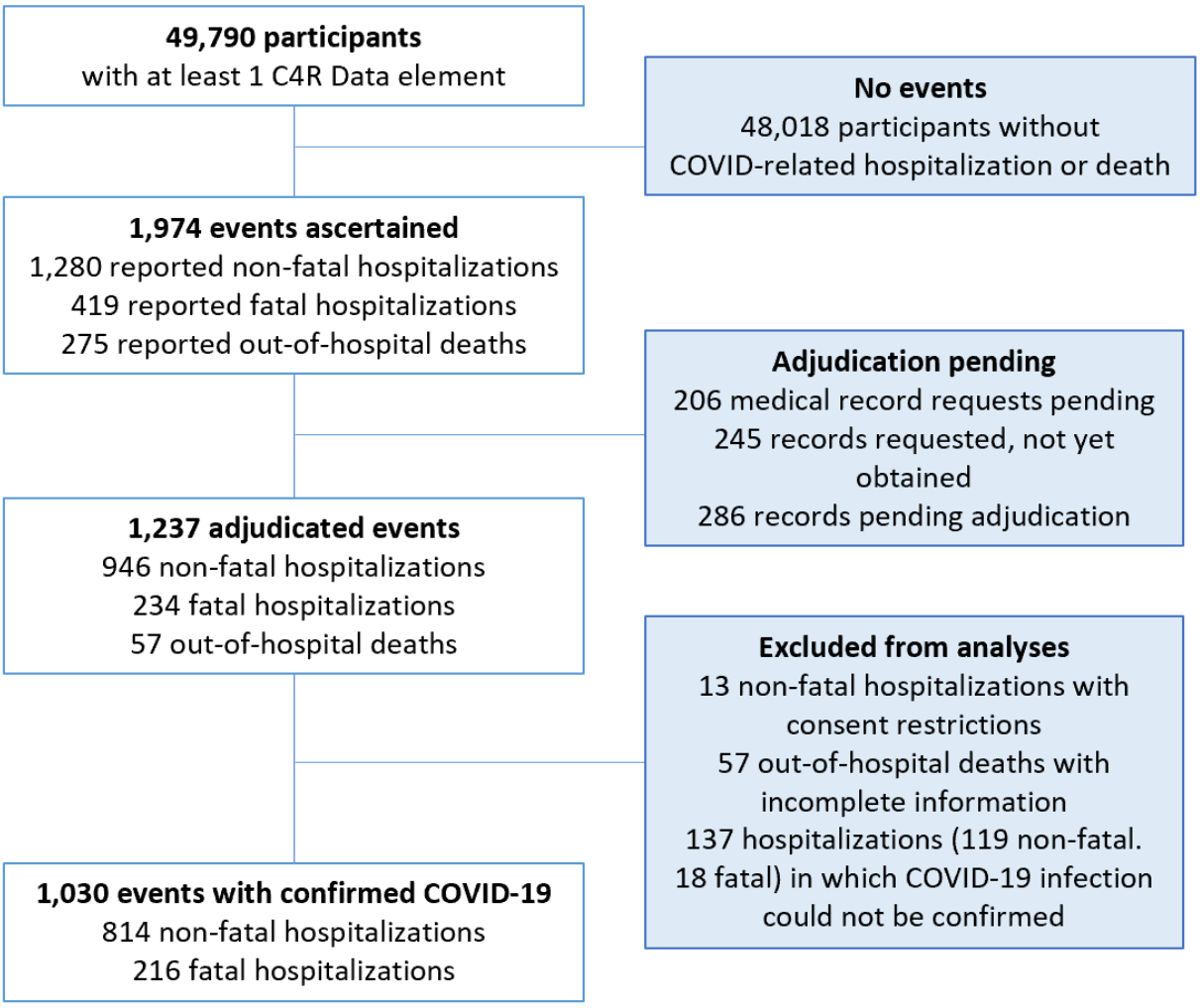
		Agreement and disagreement of ICD diagnosis with adjudicated definite diagnosis, N				Performance of ICD versus adjudicated definite diagnosis, %			
ICD code *Label*	Reference standard *C4R diagnosis*	ICD diagnosis confirmed by review	False positive by ICD diagnosis vs review	False negative by ICD diagnosis vs review	Diagnosis not confirmed by ICD or review	Sensitivity	Specificity	PPV	NPV
J12.82, *Pneumonia due to coronavirus disease 2019* J12.89*, Other viral pneumonia*	COVID-19-related Pneumonia	578	19	110	170	98%	72%	90%	95%
N17, *Acute kidney injury*	COVID-19-related Renal failure	146	23	88	617	62%	96%	86%	88%
I21, *Acute myocardial infarction*	COVID-19-related MI	4	8	3	920	57%	99%	33%	100%
I26, *Pulmonary embolism*	COVID-19-related PE	26	3	6	926	81%	100%	90%	99%
I82, *Deep venous thrombosis*	COVID-19-related DVT	21	5	9	937	70%	99%	81%	99%
I63, *Cerebral infarction*	COVID-19-related Stroke	8	0	5	952	62%	96%	86%	88%

[a]Table does not include comparisons between ICD-based classification and adjudicated diagnoses with a certainty level of 'probable'.

Abbreviations: PPV, positive predictive value; NPV, negative predictive value; MI=myocardial infarction; PE=pulmonary embolism; DVT=deep venous thrombosis.

Figure 1. Incidence of COVID-19-related hospitalization and death over C4R follow-up, January 2020 – June 2023. Incidence is calculated per month and based on date of event.

Figure 2. Consort diagram of participants with COVID-19-related hospitalizations and deaths in C4R, 2020-2023.



Funding: The Collaborative Cohort of Cohorts for COVID-19 Research (C4R) Study is supported by National Heart, Lung, and Blood Institute (NHLBI)–Collaborating Network of Networks for Evaluating COVID-19 and Therapeutic Strategies (CONNECTS) grant OT2HL156812, with cofunding from the National Institute of Neurological Disorders and Stroke (NINDS) and the National Institute on Aging (NIA). The Atherosclerosis Risk in Communities Study has been funded in whole or in part by the NHLBI, National Institutes of Health (NIH), US Department of Health and Human Services, under contracts 75N92022D00001, 75N92022D00002, 75N92022D00003, 75N92022D00004, and 75N92022D00005. Neurocognitive data are collected under grants U01 2U01HL096812, 2U01HL096814, 2U01HL096899, 2U01HL096902, and 2U01HL096917 from the NHLBI, the NINDS, the NIA, and the National Institute on Deafness and Other Communication Disorders. Ancillary studies funded additional data elements. The Blood Pressure and Cognition Study is supported by the NINDS (grant R01 NS102715). The Coronary Artery Risk Development in Young Adults Study (CARDIA) is conducted and supported by the National Heart, Lung, and Blood Institute (NHLBI) in collaboration with the University of Alabama at Birmingham (75N92023D00002 & 75N92023D00005), Northwestern University (75N92023D00004), University of Minnesota (75N92023D00006), and Kaiser Foundation Research Institute (75N92023D00003). The Genetic Epidemiology of COPD (COPDGene) Study was supported by awards U01 HL089897 and U01 HL089856 from the NHLBI. COPDGene is also supported by the COPD Foundation through contributions made to an industry advisory board comprised of AstraZeneca AB (Cambridge, United Kingdom), Boehringer-Ingelheim (Ingelheim am Rhein, Germany), Genentech, Inc. (South San Francisco, California), GlaxoSmithKline plc (London, United Kingdom), Novartis International AG (Basel, Switzerland), Pfizer, Inc. (New York, New York), Siemens AG (Berlin, Germany), and Sunovion Pharmaceuticals Inc. (Marlborough, Massachusetts). The Framingham Heart Study has received support from the NHLBI (grant N01-HC-25195, contract HHSN268201500001I, and grant 75N92019D00031). The Hispanic Community Health Study/Study of Latinos (HCHS/SOL) is a collaborative study supported by contracts between the NHLBI and the University of North Carolina (contract HHSN268201300001I/N01-HC-65233), the University of Miami (contract HHSN268201300004I/N01-HC-65234), Albert Einstein College of Medicine (contract HHSN268201300002I/N01-HC-65235), the University of Illinois at Chicago (contract HHSN268201300003I/N01-HC-65236 (Northwestern University)), and San Diego State University (contract HHSN268201300005I/N01-HC-65237). The following institutes/centers/offices have contributed to the HCHS/SOL through a transfer of funds to the NHLBI: the National Institute on Minority Health and Health Disparities, the National Institute on Deafness and Other Communication Disorders, the National Institute of Dental and Craniofacial Research, the National Institute of Diabetes and Digestive and Kidney Diseases, the NINDS, and the NIH Office of Dietary Supplements. The Jackson Heart Study is supported by and conducted in collaboration with Jackson State University (contract HHSN268201800013I), Tougaloo College (contract HHSN268201800014I), the Mississippi State Department of Health (contract HHSN268201800015I), the University of Mississippi Medical Center (contracts HHSN268201800010I, HHSN268201800011I, and HHSN268201800012I), the NHLBI, and the National Institute on Minority Health and Health Disparities. The Mediators of Atherosclerosis in South Asians Living in America (MASALA) Study was supported by grant R01HL093009 from the NHLBI, the National Center for Research Resources, and the National Center for Advancing Translational Sciences, NIH, through University of California, San Francisco–Clinical and Translational Science Institute grant UL1RR024131. The Multi-Ethnic Study of Atherosclerosis (MESA) and the MESA SNP Health Association Resource (SHARe) are conducted and supported by the NHLBI in collaboration with the MESA investigators. Support for MESA is provided by grants and contracts 75N92020D00001,

HHSN268201500003I, N01-HC-95159, 75N92020D00005, N01-HC-95160, 75N92020D00002, N01-HC-95161, 75N92020D00003, N01-HC-95162, 75N92020D00006, N01-HC-95163, 75N92020D00004, N01-HC-95164, 75N92020D00007, N01-HC-95165, N01-HC-95166, N01-HC-95167, N01-HC-95168, N01-HC-95169, R01-HL077612, R01-HL093081, R01-HL130506, R01-HL127028, R01-HL127659, R01-HL098433, R01-HL101250, and R01-HL135009 from the NHLBI; grant R01-AG058969 from the NIA; and grants UL1-TR-000040, UL1-TR-001079, and UL1-TR-001420 from the National Center for Advancing Translational Sciences. Funding for SHARe genotyping was provided by NHLBI contract N02-HL-64278. This publication was developed under Science to Achieve Results (STAR) research assistance agreements RD831697 (MESA Air) and RD-83830001 (MESA Air Next Stage), awarded by the Environmental Protection Agency. Whole genome sequencing for the Trans-Omics in Precision Medicine (TOPMed) Program was supported by the NHLBI. Whole genome sequencing for the MESA component of the TOPMed Study (Database of Genotypes and Phenotypes accession no. phs001416.v1.p1) was performed at the Broad Institute of MIT and Harvard (grant 3U54HG003067-13S1). Centralized read mapping and genotype calling, along with variant quality metrics and filtering, were provided by the TOPMed Informatics Research Center (grant 3R01HL-117626-02S1 and contract HHSN268201800002I) (Broad RNA Seq, Proteomics HHSN268201600034I, UW RNA Seq HHSN268201600032I, USC DNA Methylation HHSN268201600034I, Broad Metabolomics HHSN268201600038I). Phenotype harmonization, data management, sample-identity quality control, and general study coordination were provided by the TOPMed Data Coordinating Center (grants 3R01HL-120393 and U01HL-120393 and contract HHSN268180001I). The provision of genotyping data was supported in part by the National Center for Advancing Translational Sciences, Clinical and Translational Science Institute grant UL1TR001881, and National Institute of Diabetes and Digestive and Kidney Diseases Diabetes Research Center grant DK063491 to the Southern California Diabetes Endocrinology Research Center. The NHLBI Pooled Cohorts Study was supported by grants R21HL153700, K23HL130627, R21HL129924, and R21HL121457 from the NIH/NHLBI. The Northern Manhattan Study was supported by grants R01 NS29993 and R01 NS48134 from the NINDS and grant R01 AG066162 from the NIA. The Prevent Pulmonary Fibrosis cohort study was established in 2000 and has been supported by NIH awards Z01-ES101947, R01-HL095393, RC2-HL1011715, R21/33-HL120770, R01-HL097163, Z01-HL134585, UH2/3-HL123442, P01-HL092870, UG3/UH3-HL151865, and DoD W81XWH-17-1-0597. The Reasons for Geographic and Racial Differences in Stroke (REGARDS) Study is supported by cooperative agreement U01 NS041588, cofunded by the NINDS and the NIA. Research by the principal and co–principal investigators of the Severe Asthma Research Program was funded by the NIH/NHLBI (grants U10 HL109164, U10 HL109257, U10 HL109146, U10 HL109172, U10 HL109250, U10 HL109168, U10 HL109152, and U10 HL109086). Additional support was provided through industry partnerships with the following companies: AstraZeneca, Boehringer-Ingelheim, Genentech, GlaxoSmithKline, MedImmune, Inc. (Gaithersburg, Maryland), Novartis, Regeneron Pharmaceuticals, Inc. (Tarrytown, New York), Sanofi S.A. (Paris, France), and Teva Pharmaceuticals USA (North Wales, Pennsylvania). Spirometers used in Severe Asthma Research Program III were provided by nSpire Health, Inc. (Longmont, Colorado). The Subpopulations and Intermediate Outcome Measures in COPD Study (SPIROMICS) has been funded by contracts with the NIH/NHLBI (contracts HHSN268200900013C, HHSN268200900014C, HHSN268200900015C, HHSN268200900016C, HHSN268200900017C, HHSN268200900018C, HHSN268200900019C, and HHSN268200900020C) and grants from the NIH/NHLBI (grants U01 HL137880 and U24 HL141762) and supplemented through contributions made to the Foundation for the NIH and the COPD Foundation by AstraZeneca, MedImmune, Bayer Corporation (Whippany, New Jersey), Bellerophon Therapeutics (Warren, New Jersey), Boehringer-Ingelheim, Chiesi

Farmaceutici S.p.A. (Parma, Italia), the Forest Research Institute, Inc. (Jersey City, New Jersey), GlaxoSmithKline, Grifols Therapeutics, Inc. (Research Triangle Park, North Carolina), Ikaria, Inc. (Hampton, New Jersey), Novartis, Nycomed Pharma GmbH (Zurich, Switzerland), ProterixBio, Inc. (Billerica, Massachusetts), Regeneron, Sanofi, Sunovion, Takeda Pharmaceutical Company (Tokyo, Japan), Theravance Biopharma, Inc. (South San Francisco, California), and Mylan N.V. (White Sulphur Springs, West Virginia). The Strong Heart Study has been funded in whole or in part by the NHLBI (contracts 75N92019D00027, 75N92019D00028, 75N92019D00029, and 75N92019D00030). The Strong Heart Study was previously supported by research grants R01HL109315, R01HL109301, R01HL109284, R01HL109282, and R01HL109319 and cooperative agreements U01HL41642, U01HL41652, U01HL41654, U01HL65520, and U01HL65521. NJM was supported by U01CA260508 from the National Cancer Institute.

The views expressed in this manuscript are those of the authors and do not necessarily represent the views of the National Heart, Lung, and Blood Institute; the National Institutes of Health; or the U.S. Department of Health and Human Services.

Supplementary Methods: Cohort Descriptions

Atherosclerosis Risk in Communities (ARIC): The ARIC study began in the mid 1980s with initial aims for its cohort component being to describe the presence of subclinical atherosclerosis (mainly via carotid ultrasound), the progression of atherosclerosis to clinical cardiovascular disease (CVD), and the association of novel risk factors with CVD. ARIC recruited its cohort of 15,792 men and women aged 45-64 in 1987-89 from four communities: Forsyth County, NC; Jackson, MS; suburban Minneapolis, MN; and Washington County, MD. The investigators used probability sampling to obtain a community wide sample, exclusively sampling African Americans in Jackson and oversampling African Americans in Forsyth County. ARIC conducted a baseline examination of cohort participants and up to seven subsequent examinations; performed annual or semi-annual telephone follow-up interviews; and throughout has identified and validated incident CVD and other outcomes, particularly cognitive decline in recent years.

Coronary Artery Risk Development in Young Adults (CARDIA): CARDIA is a study examining the development and determinants of clinical and subclinical CVD and their risk factors. It began in 1985-1986 with a cohort of 5115 Black and White men and women aged 18-30 years. The participants were selected so that there would be approximately the same number of people in subgroups of race (Black and White), gender (female and male), education (high school or less and more than high school) and age (18-24 and 25-30 years) in each of 4 field centers: Birmingham, AL; Chicago, IL; Minneapolis, MN; and Oakland, CA. These same participants were asked to participate in follow-up examinations during 1987-1988 (Year 2), 1990-1991 (Year 5), 1992-1993 (Year 7), 1995-1996 (Year 10), 2000-2001 (Year 15), 2005-2006 (Year 20), 2010-2011 (Year 25), 2015-2016 (Year 30), and 2020-2022 (Year 35). A majority of the group has been examined at each of the follow-up examinations (91%, 86%, 81%, 79%, 74%, 72%, 72%, 71%, and 67% [despite the impact of the COVID-19 pandemic on Year 35], respectively). While the specific aims of each examination have varied, data have been collected on a variety of factors believed to be related to heart disease. These include conditions with clear links to heart disease such as blood pressure, cholesterol and other lipids, and glucose. Data have also been collected on physical measurements such as weight and body composition as well as lifestyle factors such as dietary and activity patterns (self-report and objective in some), substance use (including tobacco and alcohol), behavioral and psychological variables, medical and family history, and other chemistries (e.g., insulin). In addition, subclinical atherosclerosis has been measured via echocardiography during Years 5, 10, 25, and 30, a chest CT scan during Years 15, 20, 25, and 35, an abdominal CT scan during Years 25 and 35, and carotid ultrasound during Year 20. A brain MRI was performed on a subset of participants at Years 25, 30, and 35. The CARDIA cohort, born between 1955 and 1968, has been influenced substantially by the obesity epidemic at ages younger than participants in other established NHLBI cohorts. Further investigation of the mechanisms linking obesity to derangements in cardiovascular structure and function and the etiology of clinical events promises to generate important new knowledge to inform health promotion and disease prevention efforts. CARDIA is in the process of obtaining approval to conduct National Death Index (NDI) searches, and currently conducts searches every five years after cohort-specific exams.

Genetic Epidemiology of COPD (COPDGene): COPDGene is a non-interventional, multicenter, longitudinal, case-control study at 21 US sites of smokers with a ≥10 pack-year history with and without COPD and healthy never smokers. The goal was to characterize disease-related phenotypes and explore associations with susceptibility genes. COPDGene research participants were extensively phenotyped with the use of comprehensive symptom and comorbidity questionnaires, spirometry, chest CT scans, and genetic and biomarker profiling. The study enrolled

10,198 participants. COPDGene has had 3 exams that include spirometry, diffusing capacity, lung CT scans and other measures; its current exam is ongoing. COPDGene examines the influence of age, sex, and race on the natural history of COPD, and the impact of comorbid conditions, chronic bronchitis, exacerbations, and asthma/COPD overlap.

Framingham Heart Study (FHS): FHS was initiated in 1948. Researchers recruited 5,209 men and women between the ages of 30 and 62 from the town of Framingham, Massachusetts, and began the first round of extensive physical examinations and lifestyle interviews that they would later analyze for common patterns related to CVD development. Since 1948, the subjects have returned to the study every two years for an examination consisting of a detailed medical history, physical examination, and laboratory tests, and in 1971, the study enrolled a second-generation cohort – 5,124 of the original participants' adult children and their spouses – to participate in similar examinations. The second examination of the Offspring cohort occurred eight years after the first examination, and subsequent examinations have occurred approximately every four years thereafter. In April 2002 the Study entered a new phase: the enrollment of a third generation of participants, the grandchildren of the original cohort. The first examination of the Third Generation Study was completed in July 2005 and involved 4,095 participants. Thus, the FHS has evolved into a prospective, community-based, three generation family study. In addition to research studies focused on risk factors, subclinical CVD and clinically apparent CVD, Framingham investigators have also collaborated with leading researchers from around the country and throughout the world on projects involving some of the major chronic illnesses in men and women, including dementia, osteoporosis and arthritis, nutritional deficiencies, eye diseases, hearing disorders, and chronic obstructive lung disease.

Hispanic Community Health Study/Study of Latinos (HCHS/SOL): HCHS/SOL is an ongoing population based prospective cohort study of 16,415 community dwelling Hispanic/Latino adults aged 18-74 years at baseline, recruited from four urban field centers with large populations of Hispanics/ Latinos (Bronx, NY; Chicago, IL; Miami, FL; and San Diego, CA). A two-stage area probability sample of households was selected, with stratification and over-sampling at each stage to ensure a diverse and representative sample.[47] Participants self-identified as Hispanic/Latino and of Cuban, Dominican, Mexican, Puerto Rican, Central American, South American, or other/more than one heritage. Study participants underwent an extensive clinic exam and assessments to determine baseline risk factors (2008-2011),[48] and annual telephone follow-up interviews for ascertainment of cardiovascular and pulmonary events. A second clinic visit was conducted in 2014-2017, and a third clinic visit is now in process (2020-2022). The overall retention rate as of December 2019 was 81.9%. The primary goals of the HCHS/SOL are to describe: (1) the prevalence and incidence of cardiovascular, pulmonary, and other major chronic conditions (2) the risk and/or protective factors associated with these conditions; and (3) the relationships between the initial sociodemographic and health profiles and future health events in the target population. The study to date has revealed a high prevalence of cardiovascular risk factors, with significant variability by Hispanic/Latino heritage and sociodemographic factors such as income and time in the United States.[49]

Jackson Heart Study (JHS): The JHS is a community-based cohort study evaluating risk factors for cardiovascular and related diseases among adult African Americans residing in the three counties (Hinds, Madison, and Rankin) that make up the Jackson, Mississippi metropolitan area. Data and biologic materials have been collected from 5,306 participants, including a nested family cohort of 1,498 members of 264 families. The age at enrollment for the unrelated cohort was 35-84 years; the family cohort included related individuals >21 years old. Participants have provided extensive medical and psychosocial histories and had an array of physical and biochemical measurements and diagnostic

procedures during a baseline examination (2000-2004) and two follow-up examinations (2005-2008 and 2009-2012). Samples for genomic DNA were collected during the first two examinations. Annual follow-up interviews and cohort surveillance of cardiovascular events and mortality are continuing and a fourth examination is in progress.

Mediators of Atherosclerosis in South Asians Living in America (MASALA) study[15,16]: South Asians comprise almost one-quarter of the world's population and are the second fastest growing ethnic group in the US. The MASALA Study is a prospective cohort of South Asians called the MASALA study, which is closely tied to the Multi-Ethnic Study of Atherosclerosis (MESA), for valid cross-ethnic comparisons.[15] MASALA enrolled 906 South Asians in 2010-2013 and then added a new wave of 258 South Asian participants from 2017-2018, for a full cohort size of 1,164.[16] The original MASALA cohort has been followed for approximately 8.5 years, and completed a second clinical exam in early 2018. A third MASALA clinical exam is planned for 2022-2024. 75 papers have been published from MASALA to date, and the findings clearly show that the US South Asian population has a distinct phenotype compared to the other four race/ethnic groups studied in MESA. Major findings have included a higher prevalence of diabetes, ectopic adiposity and coronary artery calcium compared to MESA. The MASALA study findings have influenced guidelines for diabetes screening, lipid management, and raised awareness of South Asian CVD risk. MASALA is filling a large gap in scientific knowledge about CVD in a large, growing Asian American subgroup.

Multi-Ethnic Study of Atherosclerosis (MESA): MESA is a study of the characteristics of subclinical CVD (disease detected non-invasively before it has produced clinical signs and symptoms) and the risk factors that predict progression to clinically overt cardiovascular disease or progression of the subclinical disease. MESA researchers study a diverse, population-based sample of 6,814 men and women aged 45-84 without known clinical cardiovascular disease. Thirty-eight percent of the recruited participants are white, 28 percent African-American, 22 percent Hispanic, and 12 percent of Chinese descent. Participants were recruited from six field centers across the United States: Wake Forest University, Columbia University, Johns Hopkins University, University of Minnesota, Northwestern University and University of California – Los Angeles. At baseline, each participant received an extensive physical exam and determination of coronary artery calcification, ventricular mass and function, flow-mediated endothelial vasodilation, carotid intimal-medial wall thickness and presence of echogenic lucencies in the carotid artery, lower extremity vascular insufficiency, arterial wave forms, electrocardiographic (ECG) measures, standard coronary risk factors, sociodemographic factors, lifestyle factors, and psychosocial factors. Selected repetition of subclinical disease measures and risk factors at follow-up visits allows study of the progression of disease. Blood samples have been assayed for putative biochemical and genetic risk factors and stored for case-control studies. Participants are being followed for identification and characterization of cardiovascular disease events, including acute myocardial infarction and other forms of coronary heart disease (CHD), stroke, and heart failure; for CVD interventions; and for mortality. The first examination took place over two years, from July 2000 – July 2002. It has been followed by six examination periods that were 17-20 months in length. Participants have been contacted every 9 to 12 months throughout the study to assess clinical morbidity and mortality.

Northern Manhattan Study (NOMAS): NOMAS began in 1993 as a population-based incidence and case-control study. In 1998 (cycle 2) the study evolved into a prospective cohort study of 3,298 stroke-free, tri-ethnic, community subjects followed annually to detect stroke, MI, and death. Starting in 2003 (cycle 3), subclinical measures (brain MRI & carotid ultrasound) and the first complete neuropsychological (NP) battery were collected on 1290 members (MRI cohort). The project has remained productive through

subsequent cycles. As the cohort aged, the specific aims grew to include not only vascular determinants of stroke but also cognitive decline, mild cognitive impairment (MCI) and dementia. NOMAS participates in collaborative studies on genetics, stroke, MRI markers, Alzheimer Disease and neurodegenerative diseases. One of the major interests of the study has been the exploration of inflammatory and infectious contributors to stroke risk, subclinical atherosclerotic and cerebrovascular disease, and cognitive decline. The NOMAS community cohort of 3,298 subjects was assembled from a population-based, random sample based on the following criteria: (1) resident of at least 3 months of Northern Manhattan; (2) randomly derived from a household with a telephone; (3) age 40 or older at baseline (changed to age 55 in 1998); and (4) no history of stroke. The 1,290 subjects in the MRI cohort (median age 70 at MRI; 60% women, 15% non-Hispanic White, 17% non-Hispanic Black, 66% Hispanic, 2% Other) were evaluated with a standardized brain MRI and NP battery between 2003-08. The cohort has been prospectively followed with annual telephone contacts, including the Telephone Interview for Cognitive Status (TICS), and 3 in depth neuropsychological evaluations at 5 year intervals in the MRI cohort. The aging cohort is representative of an elderly, urban, diverse community at risk for cognitive decline. A wealth of data was collected during baseline enrollment and at time of MRI and 1st NP visit, including socio-demographics, psychosocial and socioeconomic status (education, occupational attainment, insurance status), medical history, medications, risk factors, family history and other health data, behavioral/environmental factors, subclinical vascular measures, serum biomarkers (infectious burden, neuroimmune markers using a novel multiplex assay, HOMA index for insulin resistance, adiponectin, CRP, homocysteine), carotid imaging, echocardiographic imaging (LV, LA size), ambulatory BP and cardiac rhythm monitoring, brain MRI biomarkers (regional brain volumes, regional white matter lesion burden, hippocampal volumes, cortical thickness, covert infarcts, cerebral microbleeds, perivascular spaces, brain arterial diameters), and genetic markers (GWAS, ApoE4). Fasting blood was collected and stored at baseline and at MRI. Subjects had complete blood count, chemistry profile, total protein, albumin, calcium, markers of mineral metabolism (fibroblast growth factor 23, parathyroid hormone, 1,25OH and 25OH vitamin D, and phosphate), CRP, TNF receptor levels, IL-6, and serologies against some viral and bacterial pathogens. Fasting plasma levels were assayed for total and HDL cholesterol, lipoprotein (a), HDL particle size, triglycerides, lipoprotein-associated phospholipase A2, homocysteine, serum insulin levels, and adiponectin. Buffy coats and DNA were stored on 2433 subjects and ApoE4 genotype is available on the MRI cohort. We continue to follow the cohort with annual telephone contacts and a 4th NP assessment to track cognitive trajectories and adjudicate MCI and dementia. Cognitive, functional, quality of life, and social situation questions are assessed annually. The National Death Index is consulted periodically for those with unknown vital status. An I surveillance system at CUIMC detects hospitalizations, ED visits, and clinical visits. Remarkably, only 3 (0.38%) subjects are lost, and 11 (1.4%) have withdrawn from active participation.

Prevent Pulmonary Fibrosis (PrePF): PrePF has been investigating the clinical, physiologic and genetic phenotypes of interstitial lung disease (ILD) by focusing on families with two or more cases of ILD and individuals with sporadic IPF. It has recruited over 1200 families with two or more cases of pulmonary fibrosis. These families with pulmonary fibrosis include 2837 individuals with probable or definite idiopathic interstitial pneumonia (IIP) and 2404 unaffected FDRs. In addition, PrePF recruited over 10,000 individuals with sporadic idiopathic pulmonary fibrosis (IPF).

REasons for Geographic and Racial Differences in Stroke (REGARDS): the REGARDS cohort is one of the nation's largest, most comprehensive population-based cohorts, its innovative home- and telephone-based data collection is nimble and cost-efficient. REGARDS centrally recruited and initially examined 30,239 non-Hispanic Black and White men and women aged ≥45 years in 2003-7 by telephone and in participant homes across the 48 contiguous US states (62% of US counties). Over 17 years,

REGARDS has collected follow-up data by computer-assisted telephone interviews (CATI), participant collaboration in at-home tasks (i.e., actigraphy), and a 2nd in-home visit. REGARDS oversampled Black individuals and residents of the southeastern United States known as the Stroke Belt and 17% reside in rural areas. REGARDS currently follows ~11,000 surviving participants. Comprehensive available data include adjudicated health events, social determinants of health (SDOH), cognition, biomarkers and genomics. Participants currently have mean age 76.9 (range 57-105), are 37% Black, have high cardiovascular risk, and 54% reside in the southeast — all factors associated with COVID-19 risk and adverse outcomes. Participants are geocoded, and linked to administrative data such as EPA and Medicare. Biorepositories were assembled in 2003-2007 and 2013-2016. In REGARDS, all deaths are reviewed and a cause of death adjudicated using the following information: death certificates, interviews with proxies or next of kin, and review of the last hospitalization.

Severe Asthma Research Program (SARP): SARP has been investigating the clinical, physiologic and molecular phenotypes of asthma since 2000. It is currently following ~400 deeply phenotyped asthma patients (60% severe), most with sputum samples, bronchoscopies, lung CTs, allergy status, spirometry and biobanking.

Subpopulations and Intermediate Outcome Measures in COPD Study (SPIROMICS): SPIROMICS is a multi-center, observational, longitudinal case-control study designed to guide future development of therapies for COPD by 1) providing robust criteria for sub-classifying COPD participants into groups most likely to benefit from a given therapy during a clinical trial, thereby improving the chances of successful outcome; and 2) identifying biomarkers and phenotypes that can be used as intermediate outcomes to reliably predict clinical benefit during therapeutic trials. The baseline exam included morphometric measures, spirometry, six-minute walk, an inspiratory and expiratory chest CT, and a set of standardized questionnaires. Biospecimens, including plasma, serum, DNA, urine and induced sputum, have been collected and stored. SPIROMICS has recruited 2,983 COPD cases and controls, 40-80 years old with 20+ pack-years of smoking at 12 US sites in 2010–2015. SPIROMICS has 5 follow-up exams, that include spirometry, lung CT scans, sputum induction and, in a subset, bronchoscopies; its current exam is ongoing.

Strong Heart Study (STRONG): STRONG was designed to respond to the recommendations from the Subcommittee on Cardiovascular and Cerebrovascular Disease of the Secretary of Health and Human Service's Task Force on Black and Minority Health that concluded that information on cardiovascular disease (CVD) in American Indians was inadequate. In its initial stages, the STRONG included three components. The first was a survey to determine cardiovascular disease mortality rates from 1984 to 1994 among tribal members aged 35-74 years of age residing in the 3 study areas (the community mortality study). The second was the clinical examination of 4,500 eligible tribal members. The third component is the morbidity and mortality (M&M) surveillance of these 4,500 participants. STRONG has completed three clinical examinations of the original Cohort in Phase I 1989-1991; Phase II: 1993-1995; 1998-1999, respectively. In Phases III-V, STRONG expanded to include genetic epidemiologic studies and family-based genetics studies due to the importance of genetics in the occurrence of CVD. Phase VI was a surveillance of the original STRONG cohort and of the STRONG family study participants to better understand CVD, cancer, liver disease, and inflammation in American Indians. Phase VII is currently underway with continued surveillance beginning February 2019 for a seven-year duration. The STRONG Phase VII exam serves as a platform for in-depth ancillary studies that are funded outside of the STRONG contracts. Additional sources utilized in morbidity and mortality surveillance processes for SHS include death certificates, the parent study examination questionnaire, the retrieval of medical records from healthcare facilities, annual follow-up calls conducted by the surveillance coordinator with participants,

and potential events identified by the study field staff during their interactions with participants during recruitment activities.

SUPPLEMENTARY TABLES

Supplementary Table 1. Characteristics of participants in C4R cohorts, United States, March 1, 2020.

Cohort	N	Original enrollment	Current age range, years	Sex, %	Race/ethnicity, %						Original research focus
				Female	NHW	B	H/L	As	Am Ind	Other	
ARIC	6,690	1987-89	75-97	63	77	23	0[a]	0	0	0	Cardiovascular
CARDIA	4,590	1985-86	53-66	56	50	50	0	0	0	0	Cardiovascular
COPDGene	7,731	2007-12	50-90	48	65	35	0	0	0	0	Pulmonary
FHS	7,339	1971-2005	26-108	56	86	3	4	0	0	7	Cardiovascular
HCHS/SOL	13,142	2008-11	30-87	60	0	0	100	0	0	0	Cardiovascular
JHS	2,444	2000-04	38-102	63	0	100	0	0	0	0	Cardiovascular
MASALA	1,132	2010-13	50-94	47	0	0	0	100	0	0	Cardiovascular
MESA	4,683	2000-02	65-103	56	38	27	24	12	0	0	Cardiovascular
NOMAS	1,256	1993-2003	62-106	65	12	14	72	0	1	0	Neurologic
PrePF	5,000	2000-13	40-80	55	92	3	3	0	0	0	Pulmonary
REGARDS	12,766	2003-07	57-105	58	62	38	0	0	0	0	Neurologic
SARP	397	2000-present	18-80	65	75	25	0	0	0	0	Pulmonary
SPIROMICS	2,273	2010-15	47-87	48	82	4	4	0	0	0	Pulmonary
SHS	2,915	1984-94	31-105	62	0	0	0	0	100	0	Cardiovascular

Am Ind = American Indian; As = Asian American; B = Black; H/L = Hispanic/Latinx; NHW = Non-Hispanic White.
ARIC = Atherosclerosis Risk in Communities Study; C4R = Collaborative Cohort of Cohorts for COVID-19 Research; CARDIA = Coronary Artery Risk Development in Young Adults; COPDGene= Genetic Epidemiology of COPD; FHS = Framingham Heart Study; HCHS/SOL = Hispanic Community Health Study/Study of Latinos; JHS = Jackson Heart Study; MASALA = Mediators of Atherosclerosis in South Asians Living in America; MESA = Multi-Ethnic Study of Atherosclerosis; NOMAS = Northern Manhattan Study; PrePF = Prevent Pulmonary Fibrosis; REGARDS = Reasons for Geographic and Racial Differences in Stroke; SARP = Severe Asthma Research Program; SPIROMICS = Subpopulations and Intermediate Outcome Measures in COPD Study; SHS = Strong Heart Study
[a] ARIC did not inquire regarding Hispanic/Latino ethnicity, hence White participants cannot be definitely defined as non-Hispanic.

Supplementary Table 2. Selected COVID-19 questionnaire elements and their inclusion by cohorts in wave 1 questionnaires for C4R, April 2020 – September 2021.

Question	N, Cohorts
Do you think that you have had COVID-19?	14
Did a healthcare provider ever tell you that you had COVID-19?	13 [a]
Have you ever had a test that showed you had COVID-19?	14
When you knew or thought that you had COVID-19 the first time, did you have any symptoms?	14
Since March 1, 2020, have you had an overnight stay in a hospital for any illness related to COVID-19?	14
While in the hospital, did you have any of the following treatments: A breathing tube or ventilator? ICU monitoring?	10 [b]

[a]Not included in Wave 1 Questionnaire in REGARDS.

[b]Not included in Wave 1 Questionnaire in ARIC, CARDIA, HCHS/SOL, or REGARDS.

Supplementary Table 3. Selected COVID-19 outcomes ascertained in the C4R wave 2 questionnaires.

Question	N, Cohorts
Since March 2020, have you had an overnight stay in a hospital for any illness related to COVID-19?	14
How many times have you been admitted to the hospital for COVID-19 or COVID-19 complications?	14
When was the first (second, third) time you were hospitalized for COVID-19 or complications thereof?	14
For the first (second, third) hospital admission, how many nights did you stay in the hospital?	14

References

1. Leading Causes of Death. *National Center for Health Statistics, Centers for Disease Control and Prevention*
2. Shiels MS, Haque AT, Berrington de González A, Freedman ND. Leading Causes of Death in the US During the COVID-19 Pandemic, March 2020 to October 2021. *JAMA Internal Medicine*. 2022;182(8):883-886. doi:10.1001/jamainternmed.2022.2476
3. Chen C, Haupert SR, Zimmermann L, Shi X, Fritsche LG, Mukherjee B. Global Prevalence of Post-Coronavirus Disease 2019 (COVID-19) Condition or Long COVID: A Meta-Analysis and Systematic Review *The Journal of Infectious Diseases*. 2022;226(9):1593-1607. doi:10.1093/infdis/jiac136
4. Oelsner EC, Krishnaswamy A, Balte PP, et al. Collaborative Cohort of Cohorts for COVID-19 Research (C4R) Study: Study Design. *American Journal of Epidemiology*. 2022;191(7):1153-1173. doi:10.1093/aje/kwac032
5. The Atherosclerosis Risk in Communities (ARIC) Study: design and objectives. The ARIC investigators. *Am J Epidemiol*. Apr 1989;129(4):687-702.
6. Friedman GD, Cutter GR, Donahue RP, et al. CARDIA: study design, recruitment, and some characteristics of the examined subjects. *J Clin Epidemiol*. 1988;41(11):1105-16. doi:10.1016/0895-4356(88)90080-7
7. Regan EA, Hokanson JE, Murphy JR, et al. Genetic epidemiology of COPD (COPDGene) study design. *Copd*. Feb 2010;7(1):32-43. doi:10.3109/15412550903499522
8. Tsao CW, Vasan RS. Cohort Profile: The Framingham Heart Study (FHS): overview of milestones in cardiovascular epidemiology. *Int J Epidemiol*. Dec 2015;44(6):1800-13. doi:10.1093/ije/dyv337
9. Daviglus ML, Talavera GA, Avilés-Santa ML, et al. Prevalence of major cardiovascular risk factors and cardiovascular diseases among Hispanic/Latino individuals of diverse backgrounds in the United States. *Jama*. Nov 7 2012;308(17):1775-84. doi:10.1001/jama.2012.14517
10. Lavange LM, Kalsbeek WD, Sorlie PD, et al. Sample design and cohort selection in the Hispanic Community Health Study/Study of Latinos. *Ann Epidemiol*. Aug 2010;20(8):642-9. doi:10.1016/j.annepidem.2010.05.006
11. Sorlie PD, Avilés-Santa LM, Wassertheil-Smoller S, et al. Design and implementation of the Hispanic Community Health Study/Study of Latinos. *Ann Epidemiol*. Aug 2010;20(8):629-41. doi:10.1016/j.annepidem.2010.03.015
12. Carpenter MA, Crow R, Steffes M, et al. Laboratory, reading center, and coordinating center data management methods in the Jackson Heart Study. *Am J Med Sci*. Sep 2004;328(3):131-44. doi:10.1097/00000441-200409000-00001
13. Keku E, Rosamond W, Taylor HA, Jr., et al. Cardiovascular disease event classification in the Jackson Heart Study: methods and procedures. *Ethn Dis*. Autumn 2005;15(4 Suppl 6):S6-62-70.
14. Taylor HA, Jr., Wilson JG, Jones DW, et al. Toward resolution of cardiovascular health disparities in African Americans: design and methods of the Jackson Heart Study. *Ethn Dis*. Autumn 2005;15(4 Suppl 6):S6-4-17.
15. Kanaya AM, Kandula N, Herrington D, et al. Mediators of Atherosclerosis in South Asians Living in America (MASALA) study: objectives, methods, and cohort description. *Clin Cardiol*. Dec 2013;36(12):713-720. doi:10.1002/clc.22219
16. Kanaya AM, Chang A, Schembri M, et al. Recruitment and retention of US South Asians for an epidemiologic cohort: Experience from the MASALA study. *J Clin Transl Sci*. Jun 2019;3(2-3):97-104. doi:10.1017/cts.2019.371
17. Bild DE, Bluemke DA, Burke GL, et al. Multi-Ethnic Study of Atherosclerosis: objectives and design. *Am J Epidemiol*. Nov 1 2002;156(9):871-81. doi:10.1093/aje/kwf113

18. Sacco RL, Boden-Albala B, Gan R, et al. Stroke incidence among white, black, and Hispanic residents of an urban community: the Northern Manhattan Stroke Study. *Am J Epidemiol*. Feb 1 1998;147(3):259-68. doi:10.1093/oxfordjournals.aje.a009445
19. Mathai SK, Humphries S, Kropski JA, et al. MUC5B variant is associated with visually and quantitatively detected preclinical pulmonary fibrosis. *Thorax*. Dec 2019;74(12):1131-1139. doi:10.1136/thoraxjnl-2018-212430
20. Howard VJ, Cushman M, Pulley L, et al. The reasons for geographic and racial differences in stroke study: objectives and design. *Neuroepidemiology*. 2005;25(3):135-43. doi:10.1159/000086678
21. Teague WG, Phillips BR, Fahy JV, et al. Baseline Features of the Severe Asthma Research Program (SARP III) Cohort: Differences with Age. *J Allergy Clin Immunol Pract*. Mar-Apr 2018;6(2):545-554.e4. doi:10.1016/j.jaip.2017.05.032
22. Couper D, LaVange LM, Han M, et al. Design of the Subpopulations and Intermediate Outcomes in COPD Study (SPIROMICS). *Thorax*. May 2014;69(5):491-4. doi:10.1136/thoraxjnl-2013-203897
23. Lee ET, Welty TK, Fabsitz R, et al. The Strong Heart Study. A study of cardiovascular disease in American Indians: design and methods. *Am J Epidemiol*. Dec 1990;132(6):1141-55. doi:10.1093/oxfordjournals.aje.a115757
24. North KE, Howard BV, Welty TK, et al. Genetic and environmental contributions to cardiovascular disease risk in American Indians: the strong heart family study. *Am J Epidemiol*. Feb 15 2003;157(4):303-14. doi:10.1093/aje/kwf208
25. Wyatt SB, Diekelmann N, Henderson F, et al. A community-driven model of research participation: the Jackson Heart Study Participant Recruitment and Retention Study. *Ethn Dis*. Fall 2003;13(4):438-55.
26. Oelsner EC, Loehr LR, Henderson AG, et al. Classifying Chronic Lower Respiratory Disease Events in Epidemiologic Cohort Studies. *Ann Am Thorac Soc*. Jul 2016;13(7):1057-66. doi:10.1513/AnnalsATS.201601-063OC
27. Zakai NA, McClure LA, Judd SE, et al. Racial and regional differences in venous thromboembolism in the United States in 3 cohorts. *Circulation*. Apr 8 2014;129(14):1502-9. doi:10.1161/circulationaha.113.006472
28. Harris PA, Taylor R, Minor BL, et al. The REDCap consortium: Building an international community of software platform partners. *J Biomed Inform*. Jul 2019;95:103208. doi:10.1016/j.jbi.2019.103208
29. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform*. Apr 2009;42(2):377-81. doi:10.1016/j.jbi.2008.08.010
30. Prevention CfDCa. ICD-10-CM Official Coding and Reporting Guidelines April 1, 2020 through September 30, 2020 2020 (Hyattsville, MD: National Center for Health Statistics)
31. COVID-19 Treatment Guidelines Panel. Coronavirus Disease 2019 (COVID-19) Treatment Guidelines. National Institutes of Health. Available at https://www.covid19treatmentguidelines.nih.gov/. Accessed October 27, 2023.
32. Thygesen K, Alpert JS, Jaffe AS, et al. Fourth Universal Definition of Myocardial Infarction (2018). *J Am Coll Cardiol*. Oct 30 2018;72(18):2231-2264. doi:10.1016/j.jacc.2018.08.1038
33. Chen G, Faris P, Hemmelgarn B, Walker RL, Quan H. Measuring agreement of administrative data with chart data using prevalence unadjusted and adjusted kappa. *BMC Med Res Methodol*. Jan 21 2009;9:5. doi:10.1186/1471-2288-9-5
34. Khera R, Pandey A, Ayers CR, et al. Performance of the Pooled Cohort Equations to Estimate Atherosclerotic Cardiovascular Disease Risk by Body Mass Index. *JAMA Network Open*. 2020;3(10):e2023242-e2023242. doi:10.1001/jamanetworkopen.2020.23242

35. Aronsky D, Haug PJ, Lagor C, Dean NC. Accuracy of administrative data for identifying patients with pneumonia. *Am J Med Qual*. Nov-Dec 2005;20(6):319-28. doi:10.1177/1062860605280358
36. van de Garde EM, Oosterheert JJ, Bonten M, Kaplan RC, Leufkens HG. International classification of diseases codes showed modest sensitivity for detecting community-acquired pneumonia. *J Clin Epidemiol*. Aug 2007;60(8):834-8. doi:10.1016/j.jclinepi.2006.10.018
37. Molnar AO, van Walraven C, McArthur E, Fergusson D, Garg AX, Knoll G. Validation of administrative database codes for acute kidney injury in kidney transplant recipients. *Can J Kidney Health Dis*. 2016;3:18. doi:10.1186/s40697-016-0108-7
38. Colantonio LD, Levitan EB, Yun H, et al. Use of Medicare Claims Data for the Identification of Myocardial Infarction: The Reasons for Geographic And Racial Differences in Stroke Study. *Med Care*. Dec 2018;56(12):1051-1059. doi:10.1097/mlr.0000000000001004
39. Faridi KF, Tamez H, Butala NM, et al. Comparability of Event Adjudication Versus Administrative Billing Claims for Outcome Ascertainment in the DAPT Study. *Circulation: Cardiovascular Quality and Outcomes*. 2021;14(1):e006589. doi:doi:10.1161/CIRCOUTCOMES.120.006589
40. Crane HM, Heckbert SR, Drozd DR, et al. Lessons Learned From the Design and Implementation of Myocardial Infarction Adjudication Tailored for HIV Clinical Cohorts. *American Journal of Epidemiology*. 2014;179(8):996-1005. doi:10.1093/aje/kwu010
41. Mahaffey KW, Harrington RA, Akkerhuis M, et al. Systematic adjudication of myocardial infarction end-points in an international clinical trial. *Curr Control Trials Cardiovasc Med*. Jul 17 2001;2(4):180-186. doi:10.1186/cvm-2-4-180
42. Efros O, Barda N, Meisel E, et al. Myocardial injury in hospitalized patients with COVID-19 infection-Risk factors and outcomes. *PLoS One*. 2021;16(2):e0247800. doi:10.1371/journal.pone.0247800
43. Del Prete A, Conway F, Della Rocca DG, et al. COVID-19, Acute Myocardial Injury, and Infarction. *Card Electrophysiol Clin*. Mar 2022;14(1):29-39. doi:10.1016/j.ccep.2021.10.004
44. Bavishi C, Bonow RO, Trivedi V, Abbott JD, Messerli FH, Bhatt DL. Special Article - Acute myocardial injury in patients hospitalized with COVID-19 infection: A review. *Progress in Cardiovascular Diseases*. 2020/09/01/ 2020;63(5):682-689. doi:https://doi.org/10.1016/j.pcad.2020.05.013
45. Gupta A, Madhavan MV, Sehgal K, et al. Extrapulmonary manifestations of COVID-19. *Nature Medicine*. 2020/07/01 2020;26(7):1017-1032. doi:10.1038/s41591-020-0968-3
46. 21st Century Cures Act, Pub. L. No. 114-255 (2016), https://www.govinfo.gov/app/details/PLAW-114publ255.
47. Lavange LM, Kalsbeek WD, Sorlie PD, et al. Sample design and cohort selection in the Hispanic Community Health Study/Study of Latinos. *Ann Epidemiol*. Aug 2010;20(8):642-9. doi:10.1016/j.annepidem.2010.05.006
48. Sorlie PD, Aviles-Santa LM, Wassertheil-Smoller S, et al. Design and implementation of the Hispanic Community Health Study/Study of Latinos. *Ann Epidemiol*. Aug 2010;20(8):629-41. doi:10.1016/j.annepidem.2010.03.015
49. Daviglus ML, Talavera GA, Aviles-Santa ML, et al. Prevalence of major cardiovascular risk factors and cardiovascular diseases among Hispanic/Latino individuals of diverse backgrounds in the United States. *JAMA*. Nov 7 2012;308(17):1775-84. doi:10.1001/jama.2012.14517