Epidemiologic features of recovery from SARS-CoV-2 infection in a US population-based cohort: The Collaborative Cohort of Cohorts for COVID-19 Research (C4R)

Authors
Elizabeth C Oelsner
Yifei Sun
Pallavi P Balte
Akshaya Krishnaswamy
Norrina B Allen
Howard Andrews
R Graham Barr
April Carson
Shelley A Cole
Josef Coresh
David Couper
Mary Cushman
Martha Daviglus
Ryan T Demmer
Mitchell S V Elkind
Linda C Gallo
Jose Gutierrez
Virginia J Howard
Carmen R Isasi
Suzanne Judd
Alka M Kanaya
Namratha R Kandula
Robert Kaplan
Greg Kinney
Anna Kucharska Newton
Daniel Lackland
Joyce S Lee
Barry Make
Yuan-I Min
Joanne M Murabito
Arnita Norwood
Victor Ortega
Kelley Pettee Gabriel
Bruce M Psaty
Elizabeth Regan
Rafail Rustamov
Daniela Sotrez-Alvarez
David Schwartz
Neil Schneiderman
James M Shikany
Bharat Thyagarajan
Russell P Tracy
Jason G Umans
Ramachandran S Vasan
Sally Wenzel

Prescott Woodruff
Vanessa Xanthakis
Ying Zhang
Wendy S Post

Corresponding Author: Elizabeth C Oelsner, MD MPH

Target journal: JAMA

Word count, abstract: 329

Word count, manuscript: 3964

Key points

Question

How long does it take to recover from SARS-CoV-2 infection, and what are the major correlates of self-reported recovery by 90 days after infection?

Findings

Among 4,708 participants in 14 long-term US cohort studies who reported a history of SARS-CoV-2 infection, the median time to self-reported recovery was 20 days and an estimated 23% did not recover by 90 days post-infection. Sub-optimal pre-pandemic health was associated with longer time-to-recovery. In multivariable-adjusted models, recovery by 90 days was less likely among women and obese participants and more likely among participants with vaccination at the time of infection and infection during the Omicron variant wave. Associations were mediated in part by severity of acute infection.

Meaning

One in 5 adults infected with SARS-CoV-2 did not fully recover by 90 days after infection in a highly characterized US meta-cohort, and recovery was associated with socio-demographic and clinical factors, particularly female sex and obesity. Vaccination and infection during the Omicron variant wave were favorably associated with recovery, due in part to their association with lower risks of severe infection. Improved understanding of mechanisms underlying delayed recovery after SARS-CoV-2 infection is critical to guide prevention and treatment.

Abstract

Importance Persistent symptoms and disability following SARS-CoV-2 infection – often called "long COVID," are frequently reported and pose significant personal and societal burden.

Objective To determine time to recovery following SARS-CoV-2 infection and identify correlates of recovery by 90 days.

Design, Setting C4R is a prospective observational meta-cohort study performing standardized ascertainment of SARS-CoV-2 infection across 14 extant NIH-funded cohorts.

Participants Adults with self-reported SARS-CoV-2 infection.

Exposures Pre-infection health conditions and lifestyle factors assessed before, and during, the pandemic, via pre-pandemic examinations and pandemic-era questionnaires.

Main Outcome Self-reported time to recovery, assessed by questionnaire.

Results Of 4,708 participants with self-reported SARS-CoV-2 infection (mean (SD) age 61.2 (13.8) years; 62% women; 44% Hispanic/Latino; 34% non-Hispanic White, 13% non-Hispanic Black; 8% American Indian; 1% Asian), an estimated 22.5% (95%CI:21.2,23.7) did not recover by 90 days after onset of illness. The median time to recovery was 20 days (IQR:8,75). Based on Kaplan Meier curves censored at 90 days post-infection, there were significant differences in restricted mean recovery time according to socio-demographic, clinical, and lifestyle characteristics, and particularly by acute infection severity (outpatient: 32.9 days; non-critical hospitalization: 50.4 days; critical hospitalization: 57.6 days; P<.001). In multivariable-adjusted Cox proportional hazards models, recovery by 90 days post-infection was associated with vaccination prior to infection (HR 1.23; 95%CI:1.06-1.42) and infection during the sixth (Omicron) variant wave (versus first [Wild Type] wave: HR 1.24; 95%CI:1.05-1.46); these associations were 41.2% and 19.7% mediated, respectively, by reduced severity of acute infection. Recovery was adversely associated with female sex (HR 0.87; 95%CI: 0.82-0.94) and pre-pandemic obesity (HR 0.91; 95%CI: 0.82, 1.00). No significant multivariable-adjusted associations were observed for age, educational attainment, smoking history, diabetes, chronic kidney disease, clinical cardiovascular disease, asthma, COPD, or elevated depressive symptoms.

Conclusions In a large, population-based, multi-ethnic US meta-cohort, one in five adults infected with SARS-CoV-2 did not fully recover by 90 days post-infection, particularly women and those with pre-pandemic obesity. Recovery time shortened following vaccination and the emergence of the Omicron variant, which may be related to reduced risks of severe acute infection.

Pandemic waves of SARS-CoV-2 infection have raised major concerns regarding lingering symptoms and disability, sometimes termed "long COVID" or post-acute sequelae of SARS-CoV-2 Infection (PASC).[1,2] Myriad symptoms, including fatigue, dyspnea, exercise intolerance, brain-fog, chest pain, and loss of taste and/or smell, are attributed to PASC.[3-5] There are many unanswered questions regarding the heterogeneity of PASC, which could constitute several discrete syndromes; variability in symptoms may emerge from different mechanisms of disease that could include persistent viral infection, sequalae of inflammation and immune activation including autoimmune conditions and reactivation of latent infections.[6-10] The potential public health impact of PASC has been staggering given the high incidence of SARS-CoV-2 infection across the world, despite high community rates of immunity from prior infection and effective vaccines against SARS-CoV-2.

Reports on the epidemiology of recovery from SARS-CoV-2 have varied considerably depending on the definition of PASC used, the sampling strategy employed, and the phase of the pandemic studied. For example, estimates for the proportion of participants with post-acute symptoms lasting 28 days or more have ranged from 17% to 81%.[11-14] Studies leveraging electronic health records (EHR) applied definitions for post-acute symptoms and diagnoses in large patient populations,[14-16] yet this approach is biased towards patients with more severe acute infection, particularly as self-testing has become a mainstream and recommended practice.[17] Also, EHRs often lack reliable measures of pre-infection health and lifestyle factors. Meanwhile, the estimated impact of vaccination on the risk of PASC has varied, with positive[18,19] and negative[20] associations in different reports, and there remains limited information on the relative risk of PASC after exposure to the Omicron variants, which rose to dominance in late 2021.

This report describes the epidemiologic features of recovery from SARS-CoV-2 over the US pandemic period in 14 longstanding cohort studies (meta-cohort) that included diverse participants from across the US who had robust pre-pandemic characterization of health and lifestyle factors.[21] Standardized questionnaires were used to identify SARS-CoV-2 infections of all severity levels and to define time to self-reported recovery after infection. Using these unique cohort data, we investigated temporal trends in, and pre-pandemic and pandemic era correlates of, time to recovery from SARS-CoV-2 infection.

Methods:

Study Design

The Collaborative Cohort of Cohorts for COVID-19 Research (C4R) is a nationwide prospective meta-cohort of adults comprising 14 established US prospective cohort studies funded by the National Institutes of Health (NIH; **eMethods, eTable 1**).[22] Starting as early as 1971, cohort investigators collected data on clinical and subclinical diseases and their risk factors, including lifestyle, cognition, laboratory biomarkers, and social determinants of health. C4R includes participants from across the United States, and across the adult life course, with diverse racial, ethnic, and socioeconomic backgrounds. For this report, we included participants with self-reported non-fatal SARS-CoV-2 infection and information on recovery status ascertained by questionnaires administered from April 1, 2020 to February 28, 2023 (**eFigure 1**).

Infection

C4R ascertained SARS-CoV-2 infection using two waves of standardized questionnaires administered across the entire C4R population via telephone interview, mailed pamphlet, electronic survey, and/or in-person examinations (**eTables 2-3**).

Self-reported infections were confirmed, when possible, via adjudication of medical records for hospitalizations and deaths linked to SARS-CoV-2, and a SARS-CoV-2 serosurvey conducted via dried blood spots (**eTable 4**). The present analysis includes all participants with a history of SARS-CoV-2 infection as defined by self-report in response to a questionnaire, with sensitivity analyses limited to cases that were confirmed either by self-report of a positive test, medical record review, or the C4R serosurvey. Re-infections with SARS-CoV-2 were excluded from all analyses.

With respect to defining acute infection severity, participants with self-reported infections were asked if they had any symptoms at the time of SARS-CoV-2 infection. Need for hospitalization and critical care were ascertained by questionnaires and confirmed by adjudication of medical records. Fatal infections were ascertained by proxy questionnaire or active surveillance and relevant medical records or death certificates were adjudicated. Altogether, these data were used to define the severity of acute infection as non-hospitalized (sub-categorized as asymptomatic, symptomatic, or symptomatic status unknown), hospitalized (sub-classified as non-critical or critical illness[23]), or fatal. Asymptomatic and fatal cases were excluded from the primary analyses.

Date of self-reported infection was ascertained by questionnaire and confirmed by medical records, where possible. The calendar date of infection was used to infer the relevant pandemic wave, which was associated with specific viral variants.[24]

Questionnaires also assessed the time of vaccination and, together with information on infection date, these data were used to define vaccination status at the time of the acute infection.

Recovery

Participants reporting infection with SARS-CoV-2 were asked, “Following your COVID-19 infection, would you say you are completely recovered from COVID-19 now?” Participants responding “yes” were asked, “How long did it take for you to recover?” The response was used to define time to recovery in days. Among participants who had not completely recovered at the time of inquiry, the time-to-non-recovery was defined as the number of days from the onset of the infection to the date of questionnaire completion, the timing of which was independent of infection history. Non-recovery at 90 days was classified if time to recovery or time to non-recovery was 90 days or greater. Of note, 3 cohorts (COPDGene, REGARDS, SHS) only included questions on time to recovery in the second wave of questionnaire administration (**eTable 2**).

Correlates studied in relation to recovery from SARS-CoV-2 infection

Potential correlates of time to recovery were selected *a priori* from data available in the majority of cohorts and included socio-demographic factors and pre-pandemic behavioral, physiologic, and clinical characteristics.[22]

Age, sex, and educational attainment were self-reported at enrollment into each cohort study. Age at enrollment into C4R in March 2020 was calculated from age at cohort enrollment. Race and ethnicity were self-reported according to fixed categories that differed by cohort, hence they were harmonized

into a single classification of race and ethnicity (American Indian, Asian, Non-Hispanic Black, Hispanic, Non-Hispanic White). Race and ethnicity were examined as potential correlates since this study aimed to evaluate factors associated with recovery in a multiethnic US population-based setting, where race and ethnicity have been associated with major disparities in COVID-19 outcomes.[25]

Pre-pandemic health measures were obtained for each individual cohort from the examination closest to the time of C4R enrollment, with a median interval of 5 years (IQR: 4, 12).[22] Height, weight, and systolic and diastolic blood pressure were measured using standardized methods. Blood glucose and creatinine were measured in fasting blood samples. BMI was categorized by current CDC standards as underweight or normal weight, overweight, or obese. Current smoking status was assessed at the most recent cohort examination by self-report, with biochemical verification in a subset. Medication use was self-reported or assessed via medication inventories. Diabetes was self-reported or defined by fasting blood glucose (≥126 mg/dL) or the use of hypoglycemic medications. Hypertension was defined by blood pressure (≥140/90 mmHg) or the use of anti-hypertensive medications. The estimated glomerular filtration rate (eGFR) was calculated by the Chronic Kidney Disease Epidemiology Collaboration equation. Chronic Kidney Disease (CKD) was defined by an eGFR < 45 ml/min/1.73m^2.[26] A history of clinical cardiovascular disease (CVD) was defined as a prior self-report of myocardial infarction, coronary heart disease, or heart failure, or the occurrence of relevant, adjudicated health events over cohort follow-up prior to the pandemic. A history of asthma, COPD, emphysema, and chronic bronchitis was self-reported. A history of elevated depressive symptoms was defined by a score of 10 or greater on the 10-item Centers for Epidemiologic Depression (CES-D) scale on the most recent pre-pandemic assessment.

Statistical Analyses

Analyses of time to recovery used the date of infection as the baseline and recovery as the outcome. Participants who reported non-recovery prior to 90 days were censored at the time of report. Participants who had not recovered by 90 days were censored at 90 days on the conservative assumption that proportionality may not hold among outliers with recovery times significantly greater than 90 days. Median time to recovery was plotted according to infection during specific SARS-CoV-2 variant waves, stratified by vaccination status at time of infection.

Kaplan Meier curves were used to estimate probability of non-recovery by 90 days and to estimate restricted mean recovery time according to strata of each potential correlate; differences were assessed by the log-rank test.

Cox proportional hazards regression models were performed to assess multivariable-adjusted associations with recovery by 90 days. The proportional hazards assumption was confirmed by residual plots. All correlates were included as potential confounders in fully adjusted models with two exceptions. Acute infection severity was hypothesized to operate as a mediator and therefore was adjusted for in sensitivity analyses only. Race and ethnicity were collinear with source cohort and were not adjusted for in the main model, which treated cohort as a stratum term; this approach allows each cohort to have its own baseline hazard function and thereby accommodates potential cohort heterogeneity. Linearity of associations was tested with generalized additive models. Effect modification by SARS-CoV-2 vaccination at time of infection was assessed by interaction terms and stratification by vaccination status.

Mediation of associations by acute infection severity was tested with respect to significant correlates identified in the multivariable-adjusted Cox proportional hazards model. We applied a parametric model-based mediation analysis to estimate the average mediation effects, the average direct effects, and the average percent mediation.[27,28] Time to recovery was modeled using a Weibull accelerated failure time model, and ordinal logistic regression was used to model infection severity.

Multiple imputations were used to account for missing data on time of infection, vaccination status, and correlates; no information on recovery status or time to recovery were imputed. A 2-tailed alpha of 0.05 was considered statistically significant for all analyses. Analyses were completed using R 4.0.0.[29]

Sensitivity Analyses

Cox proportional hazards models were repeated with the inclusion of participants who experienced fatal COVID-19, which was treated as non-recovery at 90 days, and asymptomatic participants. Analyses were repeated with restriction to participants with definite SARS-CoV-2 infection based on evidence of confirmatory SARS-CoV-2 testing (as defined in **eTable 4**). Models that were not stratified by cohort were performed to explore associations with race, ethnicity, and cohort.

Results:

Participants

Of the original target population for C4R (N=53,143), there was a response rate for at least one C4R questionnaire of 92.8% (**eFigure 1, eTable 5**). There were 6,980 participants who reported a history of first non-fatal SARS-CoV-2 infection on a C4R questionnaire that was completed, on average, on August 16, 2021 (IQR: March 21, 2021; January 31, 2022). Of these participants, 4,974 (71.3%) provided information on time to recovery. After further excluding asymptomatic and fatal cases, there were 4,708 participants included in the main analysis (**eFigure 1**).

The socio-demographic and pre-pandemic health and lifestyle characteristics of the 4,708 participants are shown in **Table 1**. Included participants were younger and had fewer comorbidities compared to those infected participants who were excluded due to missing data on recovery (**eTable 6**); nonetheless, the distribution of characteristics was generally similar. The mean age (SD) of included participants was 61.2 (13.8) years, 61.9% were women, and the racial and ethnic composition was 44.4% Hispanic/Latino, 33.5% non-Hispanic White, 13.2% non-Hispanic Black, 7.9% American Indian, and 1.1% Asian.

Infections occurred across six distinct waves of the pandemic (**Figure 1**). Hospitalization occurred in 23.7% of infections, and 6.8% required critical care. There were 3825 (81.2%) infections confirmed as definite by self-reported positive test results, C4R serology, or medical records adjudication (**eTable 4**). 19.2% of participants reported SARS-CoV-2 vaccination prior to infection.

Correlates of time to recovery

At the time of questionnaire completion (mean [standard deviation] time-since-infection, 250 [201] days), 3,656 (77.6%) of participants reported that they had completely recovered from their infection. Of the 1,052 who had not completely recovered at the time of questionnaire, 334 completed questionnaires less than 90 days following infection and hence recovery status at 90 days could not be defined.

The median time to recovery was 20 days (IQR:8,75). There was a reduction in median time to recovery over time (**Figure 1**). Participants who were vaccinated against SARS-CoV-2 at the time of infection generally had shorter median time to recovery.

Using the Kaplan Meier framework, the probability of non-recovery by 90 days was 22.5% (95%CI: 21.2, 23.7) and differed for infections in pre-Omicron (23.3%; 95%CI: 22.0, 24.6) versus Omicron (16.8%; 95%CI: 13.3, 20.2) waves.

The overall restricted mean recovery time was 35.4 days (95%CI: 34.4, 36.4). Restricted mean recovery time was associated with socio-demographic factors (**Figure 2A**), pre-pandemic clinical and lifestyle characteristics (**Figure 2B**), and features of SARS-CoV-2 infection (**Figure 2C**). Pertinent negatives include a lack of significant differences by age group, educational attainment, or pre-pandemic CKD or asthma. Particularly large differences in restricted mean survival time were observed across the range of acute infection severity, with almost two-fold higher recovery times reported after critical hospitalization (57.6 days; 95%CI: 51.9, 63.3) versus after outpatient illness (32.9 days; 95%CI: 31.9, 33.9).

Multivariable-adjusted hazard ratios (HR) for recovery by 90 days are presented in **Table 2**. Improved (greater) hazard of recovery was associated with vaccination prior to the time of infection and infection during the Omicron variant wave. Lesser recovery was associated with female sex and pre-pandemic obesity. The generalized additive model suggested that hazard of recovery was inversely associated with greater BMI, plateauing after 30 kg/m^2, the threshold for categorically defined obesity (**eFigure 2**). There were no significant associations with age group, educational attainment, or pre-pandemic smoking, diabetes, hypertension, CVD, chronic kidney disease, asthma, COPD, or elevated depressive symptoms.

Models stratified by vaccination status at the time of infection generally yielded similar results (**eFigure 3**). The effect estimates for female sex were similar in vaccinated (N=906; HR: 0.86; 95%CI: 0.73, 1.02) versus unvaccinated (N=3802; HR: 0.88; 95%CI: 0.81, 0.95) cases, although the multiplicative interaction term was statistically significant (P-interaction < .001). Significant adverse associations with pre-pandemic obesity were observed in vaccinated (HR: 0.79; 95%CI: 0.62, 1.00) but not unvaccinated (HR: 0.94; 95%CI: 0.84, 1.05) cases (P-interaction = .048). Infection during the Omicron wave was favorably associated with recovery in unvaccinated (HR: 1.31; 95%CI: 1.03, 1.66) but not vaccinated (HR: 1.15; 95%CI: 0.65, 2.05) cases (P-interaction = .004).

Associations with severity of acute infection

When added to the multivariable Cox proportional hazards model reported in **Table 2**, severity of infection was strongly associated with recovery by 90 days (**eTable 7**). Compared to outpatient infection, hospitalized infections showed a HR of 0.58 (95%CI: 0.51, 0.66) and hospitalizations for critical illness showed a HR of 0.45 (95% CI: 0.36, 0.57). Other effect estimates were comparable to the main model.

Using a model-based mediation approach, there was evidence to suggest partial mediation of associations by acute infection severity (**Table 3**). Favorable associations of recovery with vaccination prior to infection and infection during the omicron wave were partially mediated by associations with reduced severity of acute infection. The adverse association of recovery with obesity was also partially mediated by its association with greater severity of acute infection, although confidence intervals were wide. There was also significant mediation of sex effects, but in the reverse direction (-26.9%): the

association of male sex with more severe disease offset the association of male sex with shorter time to recovery.

Sensitivity analyses

Results were similar with the inclusion of fatal COVID-19 cases, which were treated as cases of non-recovery by 90 days (**eTable 8**); with the inclusion of asymptomatic cases (**eTable 8**); and, with exclusion of non-definite infections (**eTable 9**). In models that included cohort as a potential correlate rather than a stratum term, there were significant associations observed for the two cohorts with the highest restricted mean recovery time: COPDGene (61.9 days) and SHS (52.1 days) (**eTable 10, eFigure 4**). Models excluding these two cohorts yielded similar results to the main model, as did models excluding the four lung disease-based cohorts (**eTable 11**).

In models that did not account for cohort effects by stratification or adjustment, American Indian ancestry was adversely associated with recovery (versus non-Hispanic white, HR: 0.59; 95%CI: 0.50, 0.69; **eTable 10**); this association was not significantly mediated by severity of infection, although the sample size was limited (N=371; 40 [10.8]% non-critical hospitalizations; 25 [6.7%] critical hospitalizations). In these models, significant associations were also observed for pre-pandemic smoking history (versus never-smoking, former smoking HR: 0.89; 95%CI: 0.82, 0.96); current smoking HR: 0.83; 95%CI: 0.74, 0.93), COPD (HR: 0.76; 95%CI: 0.65, 0.90), and CVD (HR: 0.83; 95%CI: 0.71, 0.98).

Discussion:

In a large, multi-ethnic, US population-based meta-cohort with standardized, prospective data collection during the pandemic, an estimated one in five participants did not recover from SARS-CoV-2 infection by 90 days post-infection.[30] In particular, self-reported recovery by 90 days was less likely in women than men and in participants with versus without pre-pandemic obesity . Recovery was favorably associated with vaccination prior to infection and infection during the Omicron wave, and these associations were partially mediated by reduced severity of acute infection.

Our findings are broadly consistent with a significant US population burden of Long COVID. The RECOVER initiative – a large prospective case-control study of PASC – defined a framework to classify PASC based upon self-report of 12 symptoms at 6 months post-infection[31] and found a 10% prevalence of symptom score-defined PASC at 6 months of follow-up, with fewer PASC-like symptoms following Omicron infections versus infections with pre-Omicron variants. C4R did not collect the same symptom data as RECOVER, hence we cannot cross-validate the RECOVER PASC score using the prospective population-based design of C4R. Nonetheless, our main findings are in general agreement. It will be important to examine the longer-term trajectories of recovery post-SARS-CoV-2 infection and to distinguish them from other non-COVID causes of PASC-like symptoms using data being collected by C4R via a third administration of questionnaires, which were harmonized with RECOVER questionnaires to support complementary research and opportunities for cross-validation in the future.

Our observations regarding reduction in the burden of PASC over time are consistent with other reports and may be due in large part to reductions in the risk of severe SARS-CoV-2 illness over the course of the pandemic.[10] In particular, infections during the Omicron wave showed substantially shorter recovery times. This is consistent with findings from RECOVER, in which the proportion of participants with PASC was lower in the post-acute Omicron wave versus the post-acute pre-Omicron sub-cohorts.[31] Our finding

of only 20% mediation of Omicron effects by severity may speak to significant differences in the pathogenicity of the Omicron variant versus prior variants.[32] Our models also suggest that that association of vaccination with reduced recovery time may be 41% mediated by the reduced severity of infections following vaccination. The finding of significant adverse associations of obesity in vaccinated cases is of unclear significance; whether it may be linked to differential effectiveness of standard vaccine dosing in the context of obesity merits further evaluation. However, results from these subgroup analyses could be due to type 1 or type 2 error and should be interpreted with caution. Overall, our findings consistently support the use of vaccines to reduce the risk of Long COVID, particularly in high-risk groups.

Sub-optimal pre-pandemic health status was generally associated with delayed recovery in unadjusted analyses. All pre-pandemic health conditions were associated with longer restricted mean recovery time except for chronic kidney disease and asthma, and the lack of association may be due to type 2 error in the setting of relatively small numbers of affected participants. Of note, the association with elevated depressive symptoms was only borderline significant. With simultaneous adjustment for all pre-pandemic health conditions, only pre-pandemic obesity was adversely associated with recovery by 90 days. However, it is possible that nonsignificant findings for other conditions could be related to collinearity in the setting of multi-morbidity as well as unexplained between-participant variability. Moreover, since some participants may have developed clinical conditions prior to SARS-CoV-2 infection, but subsequent to the most recent cohort exam, from which we obtained our measures, effect estimates for these conditions are likely to be conservative. It is also possible that differential response rates according to severity of acute infection and general health may have blunted associations; C4R participants who reported recovery data appeared healthier than those who were excluded due to missing outcome data, and C4R may be characterized by a healthy participant bias relative to the general population. For example, recovery time in the COPDGene cohort, which follows patients with COPD across the full spectrum of clinical severity, was substantially longer than in other cohorts. Sensitivity analyses that were not stratified by or adjusted for cohort suggested additional significant associations with smoking history, COPD, and CVD.

American Indian ancestry was associated with longer restricted mean time to recovery and, in multivariable-adjusted models that did not account for cohort, with lesser recovery by 90 days. Consistent with the disproportionately devastating impact that the COVID-19 pandemic has had on Indigenous communities in the United States, C4R participants reporting American Indian ancestry reported a higher burden of both non-severe and severe SARS-CoV-2 infection. Mediation of associations of American Indian ancestry with recovery by acute infection severity could not be confirmed, although this could be due in part to limited sample sizes. Additional research is urgently needed on the extent to which the pandemic may have exacerbated US social and health disparities, including among indigenous communities. Inclusion of multi-ethnic participants in Long COVID mechanistic research and clinical trials remains essential to identifying, targeting, and equitably distributing preventive and therapeutic interventions.

Paradoxically, although consistent with multiple prior reports from across the world,[33,34] women experienced substantially longer recovery times and less recovery by 90 days despite a lower rate of severe acute illness. This could be due to reporting biases differential by sex, although other possibilities must be considered. Sex differences in risk of PASC, and particularly PASC sub-phenotypes characterized by neurologic, musculoskeletal, and autoimmune conditions,[35] may be explained by multiple, likely co-

occurring mechanisms including differences in the immune response and higher risk of autoreactivity and thrombosis in women (relative to men),[36] a premise that warrants further study.

Limitations

Limitations of this work include the potential for measurement error, uncontrolled confounding, and selection bias. Infection history was primarily self-reported, with confirmation by medical records or serology in a subset; of note, results among participants with definite infections were similar. Further research is needed into recovery following reinfection,[37] which was not addressed in this report. Recovery time was self-reported and subjective, although it constitutes an important patient-reported outcome. Further research is needed using objective biological and physiologic measures, particularly with longitudinal comparisons to pre-pandemic measurements. The questionnaires did not specifically assess the development of new symptoms after initial recovery, which has been reported[30]; however, acutely asymptomatic cases with non-zero recovery time were observed and included in sensitivity analyses, suggesting that participants with post-acute emergence of symptoms may be represented in the sample. Questionnaires were administered at different time points over the pandemic, with more limited data in the third year (2022-2023); additional C4R follow-up will provide additional information on recovery time and post-acute symptoms in the setting of Omicron variant dominance. Missing data were not missing completely at random, which could introduce bias. Because of the potential for type 1 error due to multiple comparisons, findings for sensitivity analyses should be interpreted as exploratory. Pooling across cohorts may introduce heterogeneity, and methods of questionnaire administration and response rates differed across cohorts, yet COVID-related data were collected in a standardized manner using self-report, serology, and medical record review, pre-pandemic measures were harmonized systematically, and models were stratified by cohort. Sensitivity analyses that did not account for cohort were generally consistent with the main findings, although several additional significant associations were observed, suggesting our approach was conservative.[22,38]

Conclusions

One in five adults infected with SARS-CoV-2 did not fully recover by three months post-infection in a large, multi-ethnic US population-based sample. Recovery by 90 days was less likely in women and participants with pre-pandemic obesity. Vaccination prior to infection and infection during the Omicron variant wave were associated with greater recovery, and these associations were partially mediated by reduced severity of acute infection. The long-term prognostic significance of SARS-CoV-2 infection requires further investigation. Research on mechanisms of PASC, including comparisons of multi-organ structure and function before and after the incidence of PASC, is critical to inform treatments and targeted prevention efforts.

Figure legends

Figure 1. Trends in median time to recovery after SARS-CoV-2 infection in days, by vaccination status at time of infection. Median time to recovery was plotted according to infection during specific SARS-CoV-2 variant waves, defined by calendar date, stratified by vaccination status at time of infection.

Figure 2. Restricted Mean Recovery Time from SARS-CoV-2 Infection in Days, by (A) Socio-Demographic Characteristics; (B) Pre-pandemic Clinical Characteristics; and (C) Infection Characteristics. Restricted mean recovery time was calculated from the unadjusted Kaplan Meier curve for each characteristic, censored at 90 days post-infection. Characteristics of participants who had been observed to recover by 90 days were compared to those who had not recovered, or whose follow-up was censored prior to 90 days, by the log-rank test.

Table 1. Characteristics of C4R participants according to recovery status at 90 days post-infection.

Correlates	Categories	Observed recovery by 90 days (N=3,532, 75.0%)	Non-recovery censored at less than 90 days (N=334, 7.1%)	Non-recovery by 90 days (N=842, 17.9%)	Total (N=4,708, 100%)
Age in March 2020	< 50 years	707 (20.0%)	65 (19.4%)	172 (20.4%)	944 (20.1%)
	50 – 64 years	1564 (44.3%)	122 (36.5%)	356 (42.3%)	2042 (43.4%)
	65 – 79 years	980 (27.8%)	108 (32.2%)	252 (29.9%)	1340 (28.5%)
	80+ years	280 (7.9%)	40 (11.9%)	62 (7.4%)	382 (8.1%)
Sex	Male	1383 (39.2%)	113 (33.8%)	296 (35.2%)	1792 (38.1%)
	Female	2149 (60.8%)	221 (66.2%)	546 (64.8%)	2916 (61.9%)
Race and ethnicity	American Indian	202 (5.7%)	25 (7.5%)	144 (17.1%)	371 (7.9%)
	Asian	39 (1.1%)	7 (2.1%)	4 (0.5%)	50 (1.1%)
	Non-Hispanic Black	492 (13.9%)	43 (12.9%)	88 (10.5%)	623 (13.2%)
	Hispanic/Latino	1643 (46.5%)	112 (33.7%)	332 (39.5%)	2088 (44.4%)
	Non-Hispanic White	1156 (32.7%)	147 (43.9%)	273 (32.5%)	1576 (33.5%)
Educational attainment	< High School	538 (15.2%)	36 (10.7%)	115 (13.6%)	688 (14.6%)
	High School	860 (24.3%)	81 (24.2%)	222 (26.4%)	1163 (24.7%)
	Some College	635 (18.0%)	57 (17.2%)	185 (22.0%)	878 (18.6%)
	College	1499 (42.4%)	160 (47.9%)	320 (38.0%)	1979 (42.0%)
Smoking Status	Never	2066 (58.5%)	144 (43.2%)	410 (48.7%)	2620 (55.7%)
	Former	1050 (29.7%)	127 (38.0%)	275 (32.6%)	1451 (30.8%)
	Current	417 (11.8%)	63 (18.8%)	157 (18.7%)	637 (13.5%)
Body mass index	< 25 kg/m^2	719 (20.4%)	80 (23.9%)	161 (19.2%)	960 (20.4%)
	25-29 kg/m^2	1346 (38.1%)	106 (31.8%)	262 (31.1%)	1714 (36.4%)
	30-34 kg/m^2	855 (24.2%)	76 (22.7%)	231 (27.4%)	1162 (24.7%)
	35+ kg/m^2	612 (17.3%)	72 (21.6%)	188 (22.3%)	872 (18.5%)
Diabetes	Absent	2918 (82.6%)	265 (79.4%)	661 (78.4%)	3843 (81.6%)
	Present	614 (17.4%)	69 (20.6%)	182 (21.6%)	865 (18.4%)

Hypertension	Absent	2237 (63.3%)	190 (56.8%)	466 (55.5%)	2895 (61.5%)
	Present	1295 (36.7%)	144 (43.2%)	374 (44.5%)	1814 (38.5%)
Cardiovascular disease	Absent	3328 (94.2%)	301 (90.2%)	768 (91.2%)	4397 (93.4%)
	Present	205 (5.8%)	33 (9.8%)	74 (8.8%)	311 (6.6%)
eGFR < 45 ml/min/1.73m^2	Absent	3419 (96.8%)	321 (96.1%)	811 (96.3%)	4551 (96.7%)
	Present	113 (3.2%)	13 (3.9%)	31 (3.7%)	157 (3.3%)
Asthma	Absent	3010 (85.2%)	272 (81.6%)	694 (82.4%)	3977 (84.5%)
	Present	522 (14.8%)	62 (18.4%)	148 (17.6%)	731 (15.5%)
COPD	Absent	3287 (93.1%)	287 (85.8%)	738 (87.7%)	4312 (91.6%)
	Present	245 (6.9%)	47 (14.2%)	104 (12.3%)	396 (8.4%)
Elevated depressive symptoms	Absent	2986 (84.5%)	274 (82.1%)	679 (80.7%)	3939 (83.7%)
	Present	546 (15.5%)	60 (17.9%)	163 (19.3%)	769 (16.3%)
Vaccination prior to infection	No (unvaccinated)	2826 (80.0%)	213 (63.7%)	763 (90.7%)	3802 (80.8%)
	Yes (vaccinated)	706 (20.0%)	121 (36.3%)	79 (9.3%)	906 (19.2%)
Infection wave	First (Wild type)	791 (22.4%)	68 (20.4%)	180 (21.3%)	1039 (22.1%)
	Second (Wild type)	470 (13.3%)	32 (9.6%)	109 (12.9%)	610 (13.0%)
	Third (Alpha)	1170 (33.1%)	57 (17.1%)	377 (44.8%)	1604 (34.1%)
	Fourth (Alpha)	213 (6.0%)	15 (4.5%)	53 (6.3%)	281 (6.0%)
	Fifth (Delta)	425 (12.0%)	67 (20.1%)	82 (9.7%)	574 (12.2%)
	Sixth (Omicron)	463 (13.1%)	95 (28.4%)	43 (5.1%)	600 (12.8%)
Acute infection severity	Outpatient	3173 (88.3%)	295 (88.3%)	643 (76.4%)	4111 (87.3%)
	Non-critical hospitalized	279 (7.9%)	28 (8.4%)	142 (16.9%)	449 (9.5%)
	Critical hospitalized	80 (2.3%)	11 (3.3%)	57 (6.8%)	148 (3.1%)
Study	ARIC	201 (5.8%)	27 (8.4%)	29 (3.9%)	257 (5.7%)
	CARDIA	274 (7.9%)	8 (2.5%)	29 (3.9%)	311 (6.9%)
	COPDGene	116 (3.4%)	84 (26.2%)	117 (15.6%)	317 (7.0%)

	FHS	351 (9.9%)	38 (11.4%)	58 (6.9%)	447 (9.5%)
	HCHS/SOL	1498 (43.4%)	101 (31.5%)	302 (40.2%)	1901 (42.0%)
	JHS	165 (4.8%)	6 (1.9%)	14 (1.9%)	185 (4.1%)
	MASALA	23 (0.7%)	5 (1.6%)	3 (0.4%)	31 (0.7%)
	MESA	164 (4.8%)	12 (3.7%)	37 (4.9%)	213 (4.7%)
	NOMAS	97 (2.8%)	9 (2.8%)	24 (3.2%)	130 (2.9%)
	PrePF	75 (2.2%)	3 (0.9%)	12 (1.6%)	90 (2.0%)
	REGARDS	270 (7.8%)	6 (1.9%)	61 (8.1%)	337 (7.5%)
	SARP	31 (0.9%)	3 (0.9%)	8 (1.1%)	42 (0.9%)
	SHS	195 (5.5%)	25 (7.5%)	141 (16.7%)	361 (7.7%)
	SPIROMICS	72 (2.1%)	7 (2.2%)	7 (0.9%)	86 (1.9%)

This table presents data based on the average of 10 multiply imputed datasets. Numbers may not exactly sum to totals due to rounding. Column percents reported.

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

Table 2. Correlates of recovery by 90 days after SARS-CoV-2 infection after multivariable adjustment.

Correlate	Categories	Hazard ratio for recovery	95% Confidence Interval	p-value
Age (ref: <50 years)	50 – 64 years	1.03	0.94, 1.14	0.52
	65 – 79 years	1.03	0.92, 1.16	0.62
	80+ years	1.02	0.85, 1.23	0.84
Sex (ref: male)	Female	0.87	0.82, 0.94	<0.001
Educational attainment (ref: College)	< High School	1.05	0.94, 1.17	0.41
	High School	1.04	0.95, 1.14	0.36
	Some College	1.02	0.92, 1.12	0.73
Smoking Status (ref: Never)	Former	0.95	0.88, 1.03	0.22
	Current	0.93	0.83, 1.04	0.21
BMI (ref: <25 kg/m^2)	25-29 kg/m^2	1.02	0.93, 1.13	0.66
	30+ kg/m^2	0.91	0.82, 1.00	0.046
Diabetes (ref: absent)	Present	0.96	0.87, 1.05	0.35
Hypertension (ref: Absent)	Present	0.96	0.89, 1.04	0.30
Cardiovascular disease (ref: absent)	Present	0.86	0.73, 1.02	0.080
Chronic kidney disease (ref: absent)	Present	0.92	0.75, 1.12	0.40
Asthma (ref: absent)	Present	0.98	0.87, 1.11	0.79
COPD (ref: absent)	Present	0.91	0.78, 1.08	0.28
Elevated depressive symptoms (ref: absent)	Present	0.92	0.83, 1.03	0.14
Vaccination prior to infection (ref: no)	Yes	1.23	1.06, 1.42	0.006

Infection wave (ref: 1st, WT, Spring 2020)	Second (WT, Summer/Fall 2020)	1.05	0.93, 1.18	0.46
	Third (Alpha, Winter 2020-21)	0.98	0.88, 1.08	0.61
	Fourth (Spring 2021)	0.94	0.80, 1.10	0.41
	Fifth (Delta, Summer 2021)	0.99	0.85, 1.15	0.86
	Sixth (Omicron, Fall 2021—present)	1.24	1.05, 1.46	0.013

Ref = referent. BMI = Body Mass Index. WT = Wild Type. Cox proportional hazards models were estimated to assess associations of time-to-recovery with the correlates of interest. Estimates were generated from models adjusted for all the correlates listed in the table. Hazards ratios (HRs) greater than one indicate faster recovery, whereas HRs less than one indicate slower recovery.

Table 3. Mediation of associations with time to recovery via severity of acute SARS-CoV-2 infection

Correlate	Severity of acute SARS-CoV-2 infection			Mediation analysis			
	Outpatient cases, N (%)	Non-critical hospitalized cases, N (%)	Critical hospitalized cases, N (%)	Average mediated effect, days to recovery	Average direct effect, days to recovery	Total effect, days to recovery	Proportion of effect mediated by infection severity (p-value)
Vaccination status at time of infection							
Unvaccinated (ref)	3249 (85.5%)	422 (11.1%)	132 (3.5%)				
Vaccinated	862 (95.2%)	27 (3.0%)	16 (1.8%)	-5.6 (-8.4, -2.8)	-8.2 (-17.9, 1.5)	-13.8 (-23.4, -4.3)	41.2% (p = .008)
Variant wave of infection							
Pre-Omicron (ref)	3532 (86.0%)	440 (10.7%)	135 (3.3%)				
Omicron	579 (96.4%)	9 (1.4%)	13 (2.2%)	-3.2 (-6.0, -0.4)	-13.0 (-21.0, -5.0)	-16.2 (-24.0, -8.3)	19.7% (p = .018)
Sex							
Male (ref)	1531 (85.4%)	186 (10.4%)	75 (4.2%)				
Female	2580 (88.5%)	263 (9.0%)	73 (2.5%)	-2.3 (-4.5, 0.1)	10.8 (5.6, 16.0)	8.6 (3.0, 14.1)	-23.2% (p = .001)
Obesity							
Non-obese (ref)	2392 (89.4%)	221 (8.3%)	61 (2.3%)				
Obese	1720 (84.5%)	228 (11.2%)	87 (4.3%)	2.2 (-0.1, 4.5)	6.7 (1.1, 12.4)	8.9 (2.8, 15.1)	24.4% (p = .030)

Mediation of associations by severity of acute infection was tested with respect to significant correlates identified in the multivariable-adjusted Cox proportional hazards model. Since there are no standard methods to assess mediation within the Cox proportional hazards model framework, we applied a parametric model-based mediation analysis to estimate the average mediation effects, the average direct effects, and the average percent mediated. Recovery by 90 days was modeled using an accelerated failure time model with a Weibull distribution, and ordinal logistic regression was used to model infection severity.

Figure 1. Trends in median time to recovery after SARS-CoV-2 infection in days, by vaccination status at time of infection.

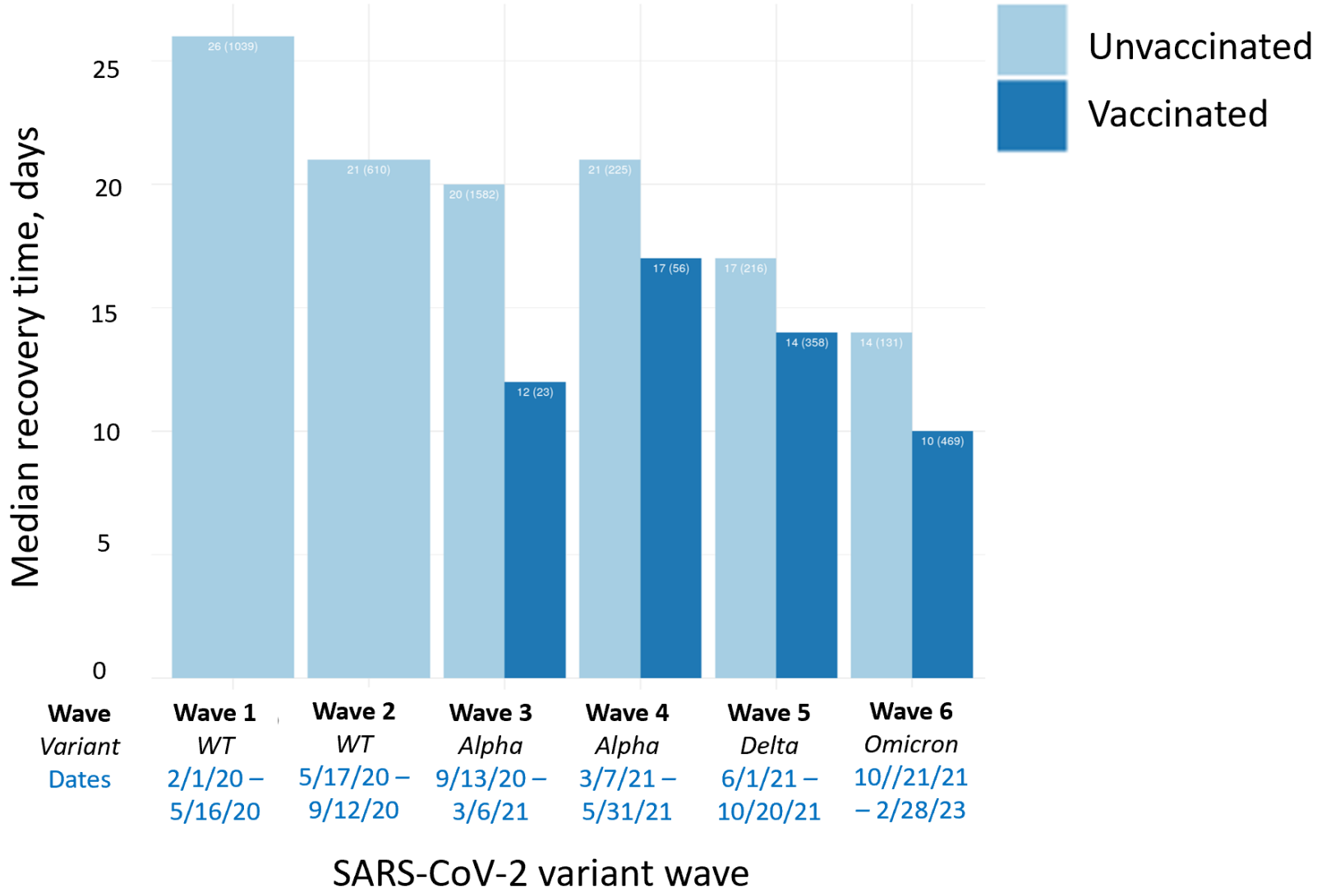


Figure 2A. Restricted Mean Recovery Time from SARS-CoV-2 Infection in Days, by Socio-Demographic Characteristics

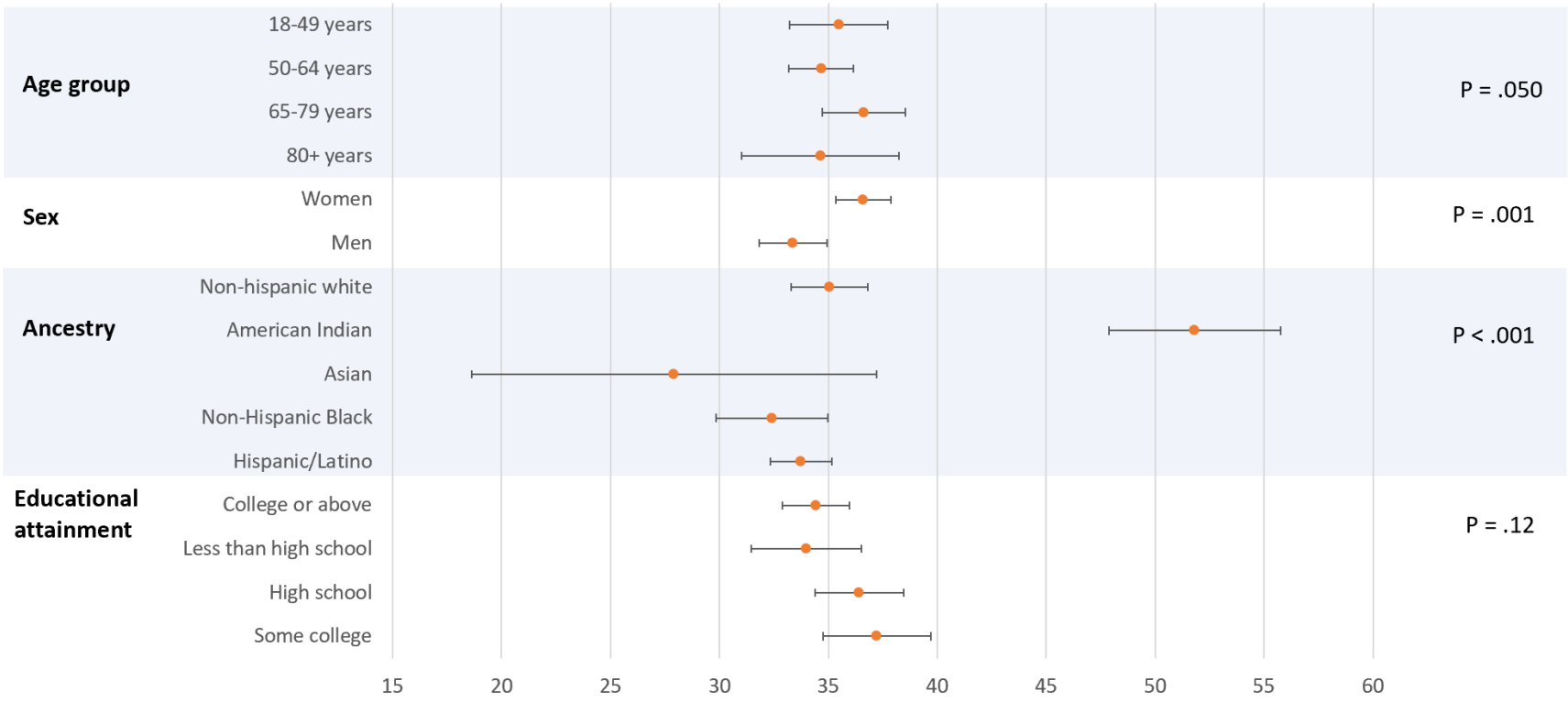


Figure 2B. Restricted Mean Recovery Time from SARS-CoV-2 Infection in Days, by Pre-Pandemic Clinical Characteristics

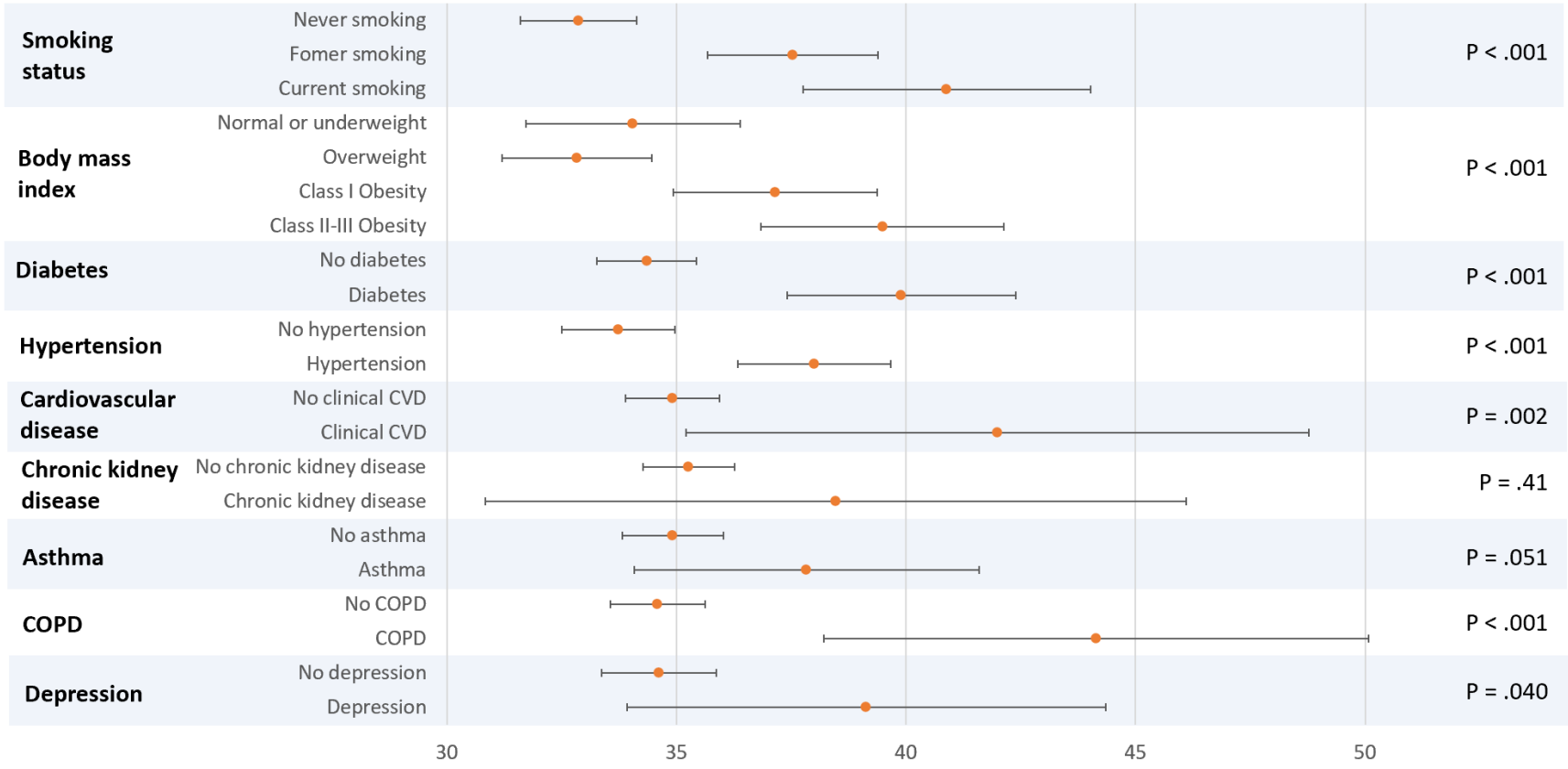
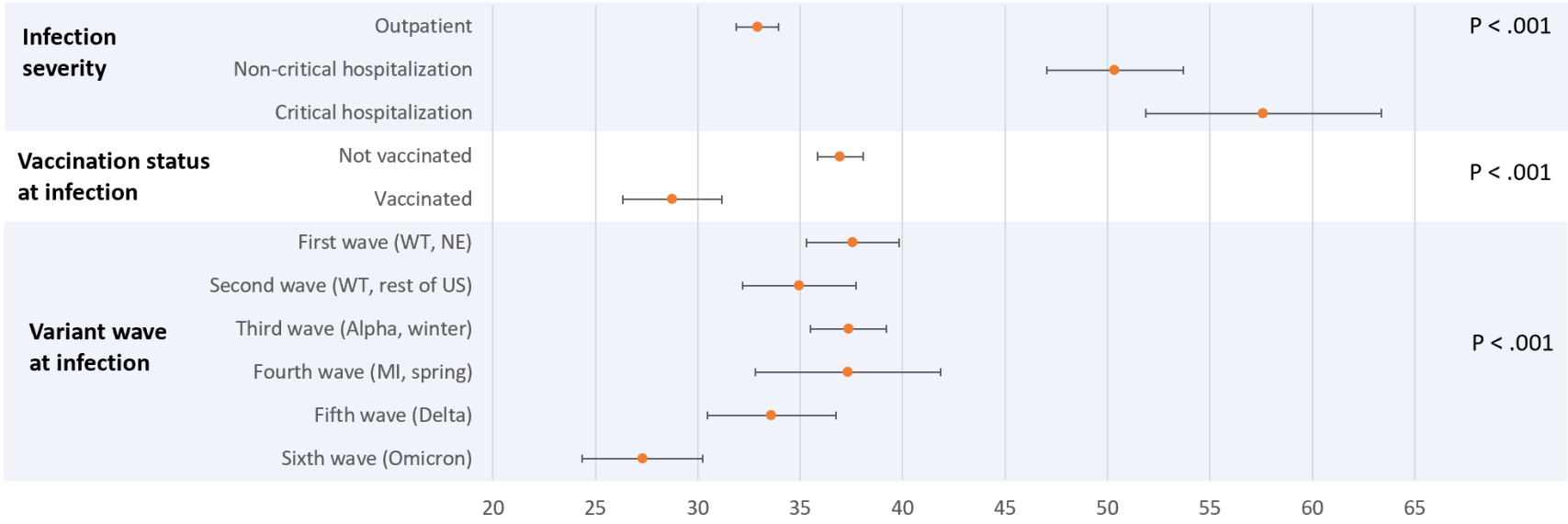


Figure 2C. Restricted Mean Recovery Time from SARS-CoV-2 Infection in Days, by Infection Characteristics



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 co-funding 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 (CARDIA) Study is conducted and supported by the NHLBI in collaboration with the University of Alabama at Birmingham (contracts HHSN268201800005I and HHSN268201800007I), Northwestern University (contract HHSN268201800003I), the University of Minnesota (contract HHSN268201800006I), and the Kaiser Foundation Research Institute (contract HSN268201800004I).

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, co-funded 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.

SUPPLEMENT

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 exercise patterns, substance use (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.

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.[39] 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),[40] 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.[41]

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[42,43]: 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.[43] 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.[42] 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 asymptomatic men and women aged 45-84. Thirty-eight percent of the recruited participants are white, 28 percent African-American, 22 percent Hispanic, and 12 percent Asian, predominantly 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. Each participant received an extensive physical exam and determination of coronary 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 risk factors and stored for case-control studies. DNA has been extracted and lymphocytes cryopreserved (for possible immortalization) for study of candidate genes and possibly, genome-wide scanning, expression, and other genetic techniques. 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 congestive heart failure; for CVD interventions; and for mortality. In addition to the six Field Centers, MESA involves a Coordinating Center, a Central Laboratory, and Central Reading Centers for Computed Tomography (CT), Magnetic Resonance Imaging (MRI), Ultrasound, and Electrocardiography (ECG). Protocol development, staff training, and pilot testing were performed in the first 18 months of the study. The first examination took place over two years, from July 2000 – July 2002. It was followed by five 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. The MESA Lung Study enrolled 3,965 MESA participants in 2004-06 and has performed spirometry 3 times and full-lung CT scans twice, most recently in 2016-18.

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.

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.

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

Cohort	N	Original enrollment	Current age range, years	Sex, %	Race and 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. Definition of infection, infection severity, and time-to-recovery via C4R Wave 1 Questionnaires.

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]
Following your COVID-19 infection, would you say that you are completely recovered from COVID-19 now?	13[a]
How many days did it take you to recover from COVID-19?	11[c]

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

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

[c]Not included in Wave 1 Questionnaire in COPDGene, REGARDS or SHS.

Supplementary Table 3. Definition of infection, infection severity, and time-to-recovery via C4R Wave 2 Questionnaires.

Domain	Question	N, Cohorts
COVID-19 TESTING	Have you ever had any kind of test for COVID-19?	14
	Have you ever had a test that showed you had COVID-19? Please include all types of tests.	14
	When was it that you first had a test that showed you had COVID-19?	14
COVID-19 SELF-REPORT	Do you think that you have had COVID-19?	14
	When do you think you FIRST had COVID-19?	14
	Were you tested at that time?	11
COVID-19 HOSPITALIZATION	Since March 2020, have you had an overnight stay in a hospital for any illness related to COVID-19?	14
COVID-19 SYMPTOMS	When you knew or thought that you had COVID-19, did you have any symptoms?	14
COVID-19 RECOVERY	Following your COVID-19 infection, would you say that you are completely recovered from COVID-19 now?	14

Supplementary Table 4. Classification of selected COVID-19 outcomes available in C4R as definite versus probable, and number (percent) of cases in the analysis sample (N=4,708).

Outcome	Definite cases		Probable cases	
	Criteria for definite case	N (%)	**Criteria for probable case**	N (%)
Infection	Adjudicated definite COVID-19 infection Proxy/self-report of positive test for SARS-CoV-2 Anti-nucleocapsid protein IgG on C4R serology, regardless of COVID vaccination Anti-spike protein IgG on C4R serology without COVID vaccination	3825 (81.2%)	Adjudicated probable COVID-19 infection Proxy/self-report of self- or healthcare provider-diagnosed COVID-19 without confirmatory testing	883 (18.8%)
Recovery to usual state of health	Self-report of recovery to usual state of health in the context of definite infection	2997 (63.7%)	Self-report of recovery to usual state of health in the context of probable infection	659 (14.0%)
Symptomatic infection	Self-report of symptoms in the context of definite infection Adjudication of symptoms definitely caused by COVID-19 recorded in medical records Adjudicated definite COVID-19 hospitalization, critical illness, and/or death	3320 (70.5%)	Self-report of symptoms in the context of probable infection Adjudication of symptoms probably caused by COVID-19 recorded in medical records Adjudicated probable COVID-19 hospitalization, critical illness, and/or death	790 (16.8%)
Hospitalization	Adjudicated hospitalization definitely caused by COVID-19	292 (6.2%)	Adjudicated hospitalization probably caused by COVID-19 Proxy/self-report of hospitalization for COVID-19	305 (6.5%)
Critical illness	Adjudicated definite COVID-19 critical illness	45 (1.0%)	Adjudicated probable COVID-19 critical illness Proxy/self-report of mechanical ventilation or ICU-level care for COVID-19	103 (2.2%)

Percentages based on main analysis sample (N=4,708).

Supplementary Table 5. Questionnaire administration, including response rates and availability of complete data for analysis, by cohort.

Cohort	Eligible for C4R	Responded to questionnaire, of estimated eligible	Infection self reported, of respondents	Valid data on recovery by 90 days, of infections	Average date of questionnaire completion (IQR)
ARIC	5046	5504 (109.1%)	450 (8.2%)	257 (57.1%)	July 25, 2021 (January 25, 2021; November 30, 2021)
CARDIA	4221	2815 (66.6%)	429 (15.2%)	311 (72.5%)	July 31, 2021 (March 14, 2021; November 2, 2021)
COPDGene	4000	4164 (104.1%)	432 (10.4%)	317 (73.4%)	May 31, 2021 (June 30, 2020; February 4, 2022)
FHS	7339	3206 (43.7%)	514 (16.0%)	447 (87.0%)	July 23, 2021 (March 10, 2021; December 16, 2021)
HCHS/SOL	8400	11304 (134.6%)	2296 (20.3%)	1901 (82.8%)	September 2, 2021 (January 17, 2021; March 24, 2022)
JHS	2317	2314 (99.9%)	249 (10.8%)	185 (74.3%)	August 2, 2021 (March 8, 2021; January 15, 2022)
MASALA	500	571 (114.2%)	38 (6.7%)	31 (81.6%)	July 5, 2021 (February 21, 2021; October 7, 2021)
MESA	4683	3478 (74.3%)	292 (8.4%)	213 (72.9%)	July 30, 2021 (August 20, 2021; December 10, 2021)
NOMAS	1256	849 (67.6%)	169 (19.9%)	130 (76.9%)	July 27, 2021 (May 15, 2021; August 2, 2021)
PrePF	2500	628 (25.1%)	115 (18.3%)	90 (78.3%)	July 12, 2021 (March 31, 2021; October 14, 2021)
REGARDS	8000	10268 (128.4%)	1169 (11.4%)	337 (28.8%)	August 29, 2021 (July 19, 2021; September 14, 2021)
SARP	380	387 (101.8%)	62 (16.0%)	42 (67.7%)	July 28, 2021 (March 30, 2021; January 18, 2022)
SHS	2701	1967 (72.8%)	630 (32.0%)	361 (57.3%)	November 14, 2021 (August 23, 2021; February 23, 2022)
SPIROMICS	1800	1539 (85.5%)	146 (9.5%)	86 (58.9%)	May 21, 2021 (October 29, 2020; December 9, 2021)

Row percents reported. 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

Supplementary Table 6. Comparison of characteristics of infected C4R participants included versus not included in the analysis sample.

		Not in analysis sample due to missing recovery data (N=3,127)	Analysis sample (N=4,708)
Age	< 50 years	297 (9.7%)	932 (20.0%)
	50 – 64 years	847 (27.6%)	2019 (43.3%)
	65 – 79 years	1318 (42.9%)	1330 (28.6%)
	80+ years	611 (19.9%)	377 (8.1%)
Sex	Male	1333 (42.6%)	1791 (38.1%)
	Female	1793 (57.4%)	2915 (61.9%)
Race and ethnicity	American Indian	410 (13.1%)	371 (7.9%)
	Asian	29 (0.9%)	50 (1.1%)
	Non-Hispanic Black	809 (25.9%)	622 (13.2%)
	Hispanic	540 (17.3%)	2086 (44.3%)
	Non-Hispanic White	1332 (42.6%)	1569 (33.3%)
Educational attainment	< High School	374 (12.5%)	676 (15.0%)
	High School	825 (27.7%)	1110 (24.6%)
	Some College	656 (22.0%)	833 (18.4%)
	College	1126 (37.8%)	1898 (42.0%)
Smoking Status	Never	1427 (47.3%)	2584 (55.9%)
	Former	1124 (37.3%)	1428 (30.9%)
	Current	463 (15.4%)	608 (13.2%)
BMI	< 25 kg/m^2	583 (19.7%)	899 (20.3%)
	25-29 kg/m^2	1027 (34.7%)	1610 (36.4%)
	30+ kg/m^2	1352 (45.6%)	1909 (43.2%)
Diabetes	Absent	2252 (75.6%)	3666 (81.8%)
	Present	725 (24.4%)	818 (18.2%)
Hypertension	Absent	1336 (44.7%)	2762 (61.4%)
	Present	1655 (55.3%)	1739 (38.6%)

Cardiovascular disease	Absent	1672 (90.0%)	3652 (94.0%)
	Present	186 (10.0%)	232 (6.0%)
eGFR < 45	Absent	2682 (97.4%)	4176 (96.8%)
	Present	72 (2.6%)	138 (3.2%)
Asthma	Absent	1554 (85.1%)	3202 (84.7%)
	Present	272 (14.9%)	580 (15.3%)
COPD, emphysema, chronic bronchitis	Absent	1686 (89.8%)	3560 (91.9%)
	Present	192 (10.2%)	314 (8.1%)
Depressive symptoms	Absent		
	Present		
Vaccination prior to infection	No	1178 (66.2%)	3214 (81.9%)
	Yes	602 (33.8%)	709 (18.1%)
Infection wave	First (Wild type)	262 (12.9%)	938 (22.3%)
	Second (Wild type)	224 (11.0%)	543 (12.9%)
	Third (Alpha)	610 (30.0%)	1425 (33.9%)
	Fourth (Alpha)	208 (10.2%)	251 (6.0%)
	Fifth (Delta)	413 (20.3%)	508 (12.1%)
	Sixth (Omicron)	318 (15.6%)	534 (12.7%)
COVID severity	Asymptomatic	362 (11.6%)	0 (0.0%)
	Symptomatic, non-hospitalized	893 (28.6%)	3513 (74.6%)
	Symptoms unknown, non-hospitalized	1358 (43.5%)	598 (12.7%)
	Hospitalized, non-critical	281 (9.0%)	449 (9.5%)
	Hospitalized, critical	53 (1.7%)	148 (3.1%)
	Fatal	174 (5.6%)	0 (0.0%)

Study	ARIC	365 (11.7%)	257 (5.5%)
	CARDIA	159 (5.1%)	311 (6.6%)
	COPDGene	213 (6.8%)	317 (6.7%)
	FHS	102 (3.3%)	447 (9.5%)
	HCHS/SOL	463 (14.8%)	1901 (40.4%)
	JHS	85 (2.7%)	185 (3.9%)
	MASALA	12 (0.4%)	31 (0.7%)
	MESA	177 (5.7%)	213 (4.5%)
	NOMAS	41 (1.3%)	130 (2.8%)
	PrePF	30 (1.0%)	90 (1.9%)
	REGARDS	972 (31.1%)	337 (7.2%)
	SARP	29 (0.9%)	42 (0.9%)
	SHS	406 (13.0%)	361 (7.7%)
	SPIROMICS	73 (2.3%)	86 (1.8%)

Supplementary Table 7. Correlates of recovery by 90 days after SARS-CoV-2 infection after multivariable adjustment, adjusted for disease severity.

Correlate	Categories	Hazard ratio for recovery	95% Confidence Interval	p-value
Age (ref: <50 years)	50 – 64 years	1.05	0.95, 1.16	.31
	65 – 79 years	1.10	0.98, 1.24	.11
	80+ years	1.14	0.95, 1.37	.17
Sex (ref: male)	Female	0.85	0.79, 0.91	<.001
Educational attainment (ref: College)	< High School	1.05	0.95, 1.18	.35
	High School	1.05	0.96, 1.15	.30
	Some College	1.02	0.93, 1.13	.68
Smoking Status (ref: Never)	Former	0.96	0.88, 1.04	.28
	Current	0.93	0.83, 1.05	.23
BMI (ref: <25 kg/m^2)	25-29 kg/m^2	1.04	0.94, 1.14	.49
	30+ kg/m^2	0.93	0.84, 1.02	.13
Diabetes (ref: absent)	Present	1.00	0.91, 1.10	.99
Hypertension (ref: Absent)	Present	0.97	0.90, 1.05	.46
Cardiovascular disease (ref: absent)	Present	0.87	0.74, 1.04	.12
Chronic kidney disease (ref: absent)	Present	0.89	0.73, 1.04	.26
Asthma (ref: absent)	Present	0.99	0.88, 1.10	.79
COPD (ref: absent)	Present	0.95	0.81, 1.12	.53
Elevated depressive symptoms (ref: absent)	Present	0.94	0.85, 1.05	.25
Vaccination prior to infection (ref: no)	Yes	1.15	1.00, 1.33	.054

Infection wave (ref: 1st, WT, Spring 2020)	Second (WT, Summer/Fall 2020)	1.06	0.94, 1.20	.31
	Third (Alpha, Winter 2020-21)	0.98	0.89, 1.08	.65
	Fourth (Spring 2021)	0.93	0.79, 1.09	.35
	Fifth (Delta, Summer 2021)	1.01	0.86, 1.18	.93
	Sixth (Omicron, Winter 2021-22)	1.23	1.04, 1.46	.014
Acute infection severity (ref: outpatient	Non-critical hospitalization	0.58	0.51, 0.66	<.001
	Critical hospitalization	0.45	0.36, 0.57	<.001

Ref = referent. BMI = Body Mass Index. WT = Wild Type. Cox proportional hazards models were estimated to assess associations of time-to-recovery with the correlates of interest. Estimates were generated from models adjusted for all the correlates listed in the table. Hazards ratios (HRs) greater than one indicate faster recovery, whereas HRs less than one indicate slower recovery.

Supplementary Table 8. Main correlates of recovery in time-to-event models, including cases of asymptomatic or fatal SARS-CoV-2 infection.

Correlate	Categories	Hazard ratio (95% CI) for recovery, including asymptomatic cases (N = 4,959)	Hazard ratio (95% CI) for recovery, including fatal cases (N=4,974)
Age (ref: <50 years)	50 – 64 years	1.03 (0.94, 1.14)	1.01 (0.92, 1.12)
	65 – 79 years	1.06 (0.94, 1.19)	0.95 (0.84, 1.07)
	80+ years	1.09 (0.91, 1.30)	0.69 (0.58, 0.83)
Sex (ref: male)	Female	0.87 (0.82, 0.93)	0.89 (0.83, 0.96)
Educational attainment (ref: College)	< High School	1.04 (0.94, 1.16)	1.00 (0.90, 1.12)
	High School	1.05 (0.96, 1.15)	1.01 (0.92, 1.11)
	Some College	1.02 (0.93, 1.12)	1.04 (0.94, 1.15)
Smoking Status (ref: Never)	Former	0.96 (0.89, 1.04)	0.92 (0.85, 1.00)
	Current	0.93 (0.83, 1.03)	0.88 (0.79, 0.99)
BMI (ref: <25 kg/m^2)	25-29 kg/m^2	1.04 (0.94, 1.14)	1.04 (0.94, 1.15)
	30+ kg/m^2	0.92 (0.84, 1.01)	0.93 (0.84, 1.03)
Diabetes (ref: absent)	Present	0.95 (0.87, 1.04)	0.91 (0.83, 1.00)
Hypertension (ref: Absent)	Present	0.94 (0.87, 1.02)	0.92 (0.85, 1.00)
Cardiovascular disease (ref: absent)	Present	0.87 (0.74, 1.03)	0.85 (0.72, 1.00)
Chronic kidney disease (ref: absent)	Present	0.89 (0.73, 1.09)	0.92 (0.76, 1.12)
Asthma (ref: absent)	Present	0.99 (0.89, 1.10)	0.99 (0.90, 1.10)
COPD (ref: absent)	Present	0.90 (0.77, 1.05)	0.93 (0.80, 1.08)
Elevated depressive symptoms (ref: absent)	Present	0.92 (0.83, 1.01)	0.93 (0.83, 1.03)

Vaccination prior to infection (ref: no)	Yes	1.25 (1.08, 1.46)	1.48 (1.28, 1.71)
Infection wave (ref: 1st, WT, Spring 2020)	Second (WT, Summer/Fall 2020)	1.08 (0.96, 1.21)	1.08 (0.95, 1.22)
	Third (Alpha, Winter 2020-21)	1.00 (0.91, 1.10)	1.02 (0.93, 1.13)
	Fourth (Spring 2021)	1.00 (0.86, 1.16)	0.93 (0.79, 1.09)
	Fifth (Delta, Summer 2021)	1.00 (0.86, 1.16)	0.95 (0.82, 1.11)
	Sixth (Omicron, Winter 2021 – present)	1.24 (1.05, 1.46)	1.17 (0.98, 1.39)

Cox proportional hazards models were performed to assess associations of time-to-recovery with the correlates of interest. Estimates were generated from models adjusted for all of the covariates listed in the table. Hazards ratios (HRs) greater than one indicate faster recovery, whereas HRs less than one indicate slower recovery. Estimates with a p-value of <0.05 are indicated in bold text.

Fatal cases include participants who reported recovery information but died within 90 days (N=15, shown in Supplementary Figure 1) plus participants with fatal infections who did not report recovery time (N=236).

Supplementary Table 9. Main correlates of recovery in time-to-event models, after exclusion of probable (non-definite) cases.

Correlate	Categories	Hazard ratio for recovery	95% Confidence Interval	p-value
Age (ref: <50 years)	50 – 64 years	1.04	0.93, 1.15	.52
	65 – 79 years	1.07	0.94, 1.22	.30
	80+ years	1.07	0.87, 1.31	.52
Sex (ref: male)	Female	0.88	0.81, 0.95	<.001
Educational attainment (ref: College)	< High School	1.03	0.91, 1.15	.66
	High School	1.05	0.96, 1.16	.30
	Some College	0.98	0.88, 1.10	.76
Smoking Status (ref: Never)	Former	0.96	0.88, 1.05	.37
	Current	0.88	0.78, 1.00	.057
BMI (ref: <25 kg/m^2)	25-29 kg/m^2	0.98	0.88, 1.09	.71
	30+ kg/m^2	0.88	0.79, 0.99	.028
Diabetes (ref: absent)	Present	0.93	0.84, 1.04	.20
Hypertension (ref: Absent)	Present	0.95	0.87, 1.04	.24
Cardiovascular disease (ref: absent)	Present	0.86	0.72, 1.02	.087
Chronic kidney disease (ref: absent)	Present	0.97	0.77, 1.20	.75
Asthma (ref: absent)	Present	0.96	0.84, 1.09	.52
COPD (ref: absent)	Present	0.93	0.78, 1.10	.39
Elevated depressive symptoms (ref: absent)	Present	0.90	0.80, 1.01	.083
Vaccination prior to infection (ref: no)	Yes	1.24	1.05, 1.45	.012

Infection wave (ref: 1st, WT, Spring 2020)	Second (WT, Summer/Fall 2020)	1.03	0.90, 1.18	.67
	Third (Alpha, Winter 2020-21)	0.93	0.83, 1.03	.17
	Fourth (Spring 2021)	0.90	0.75, 1.08	.27
	Fifth (Delta, Summer 2021)	0.93	0.78, 1.11	.42
	Sixth (Omicron, Winter 2021-present)	1.21	1.00, 1.45	.049

Excludes 862 probable (non-definite) infections for a sample size of 3659. Cox proportional hazards models were performed to assess associations of time-to-recovery with the correlates of interest. Estimates were generated from models adjusted for all of the covariates listed in the table. Hazards ratios (HRs) greater than one indicate faster recovery, whereas HRs less than one indicate slower recovery.

Supplementary Table 10. Main correlates of recovery in time-to-event models, without stratification by cohort, adjusting for race, ethnicity, and cohort.

Correlate	Categories	Hazard ratio (95% CI) for recovery No stratification by cohort	Hazard ratio (95% CI) for recovery No stratification by cohort, adjusted for cohort	Hazard ratio (95% CI) for recovery No stratification by cohort, adjusted for race and ethnicity
Age (ref: <50 years)	50 – 64 years	1.08 (0.98, 1.19)	1.03 (0.94, 1.14)	1.04 (0.95, 1.15)
	65 – 79 years	1.06 (0.95, 1.18)	1.03 (0.91, 1.16)	1.01 (0.91, 1.13)
	80+ years	1.14 (0.97, 1.32)	1.01 (0.84, 1.22)	1.06 (0.91, 1.24)
Sex (ref: male)	Female	0.89 (0.83, 0.95)	0.88 (0.82, 0.95)	0.88 (0.82, 0.95)
Educational attainment (ref: College)	< High School	1.04 (0.94, 1.16)	1.05 (0.94, 1.17)	1.06 (0.95, 1.18)
	High School	0.98 (0.90, 1.07)	1.04 (0.95, 1.14)	1.03 (0.94, 1.12)
	Some College	0.96 (0.87, 1.06)	1.01 (0.92, 1.12)	1.00 (0.91, 1.10)
Smoking Status (ref: Never)	Former	0.86 (0.79, 0.93)	0.95 (0.88, 1.03)	0.89 (0.82, 0.96)
	Current	0.77 (0.68, 0.86)	0.93 (0.83, 1.04)	0.83 (0.74, 0.93)
BMI (ref: <25 kg/m^2)	25-29 kg/m^2	1.03 (0.93, 1.13)	1.02 (0.93, 1.13)	1.03 (0.93, 1.13)
	30+ kg/m^2	0.91 (0.83, 1.01)	0.91 (0.82, 1.00)	0.92 (0.83, 1.01)
Diabetes (ref: absent)	Present	0.92 (0.83, 1.01)	0.96 (0.87, 1.05)	0.96 (0.87, 1.05)
Hypertension (ref: Absent)	Present	0.92 (0.85, 1.00)	0.96 (0.89, 1.04)	0.94 (0.87, 1.02)
Cardiovascular disease (ref: absent)	Present	0.84 (0.72, 1.00)	0.86 (0.73, 1.02)	0.83 (0.71, 0.98)
Chronic kidney disease (ref: absent)	Present	1.00 (0.81, 1.23)	0.93 (0.76, 1.13)	0.97 (0.80, 1.19)
Asthma (ref: absent)	Present	0.98 (0.87, 1.11)	0.98 (0.87, 1.11)	1.01 (0.90, 1.14)
COPD (ref: absent)	Present	0.77 (0.66, 0.91)	0.91 (0.77, 1.07)	0.76 (0.65, 0.90)

Elevated depressive symptoms (ref: absent)	Present	0.91 (0.81, 1.03)	0.93 (0.83, 1.03)	0.92 (0.82, 1.02)
Vaccination status at time of infection (ref: unvaccinated)	Vaccinated	1.24 (1.08, 1.44)	1.23 (1.07, 1.42)	1.25 (1.09, 1.44)
Infection wave (ref: First, WT, Spring 2020)	Second (WT, Summer/Fall 2020)	1.12 (0.99, 1.26)	1.05 (0.94, 1.19)	1.13 (1.00, 1.27)
	Third (Alpha, Winter 2020-21)	1.01 (0.92, 1.12)	0.99 (0.90, 1.09)	1.07 (0.97, 1.18)
	Fourth (Spring 2021)	1.02 (0.87, 1.20)	0.95 (0.81, 1.11)	1.02 (0.87, 1.20)
	Fifth (Delta, Summer 2021)	1.06 (0.90, 1.24)	1.00 (0.86, 1.16)	1.07 (0.92, 1.25)
	Sixth (Omicron, Winter 2021-present)	1.26 (1.06, 1.49)	1.24 (1.05, 1.47)	1.29 (1.09, 1.52)
Race and ethnicity (ref: non-Hispanic white)	AIAN			0.59 (0.50, 0.69)
	Asian			1.16 (0.83, 1.62)
	Non-Hispanic Black			1.11 (1.00, 1.24)
	Hispanic/Latino			0.99 (0.91, 1.09)
Study (ref: ARIC)	CARDIA		1.14 (0.91, 1.42)	
	COPDGene		0.32 (0.25, 0.41)	
	FHS		0.94 (0.77, 1.16)	
	HCHS/SOL		0.86 (0.71, 1.04)	
	JHS		1.08 (0.86, 1.36)	
	MASALA		1.04 (0.66, 1.65)	
	MESA		0.86 (0.69, 1.07)	
	NOMAS		0.91 (0.70, 1.17)	
	PrePF		1.23 (0.92, 1.64)	
	REGARDS		0.98 (0.80, 1.20)	
	SARP		0.91 (0.61, 1.36)	
	SHS		0.49 (0.39, 0.62)	
	SPIROMICS		1.10 (0.82, 1.47)	

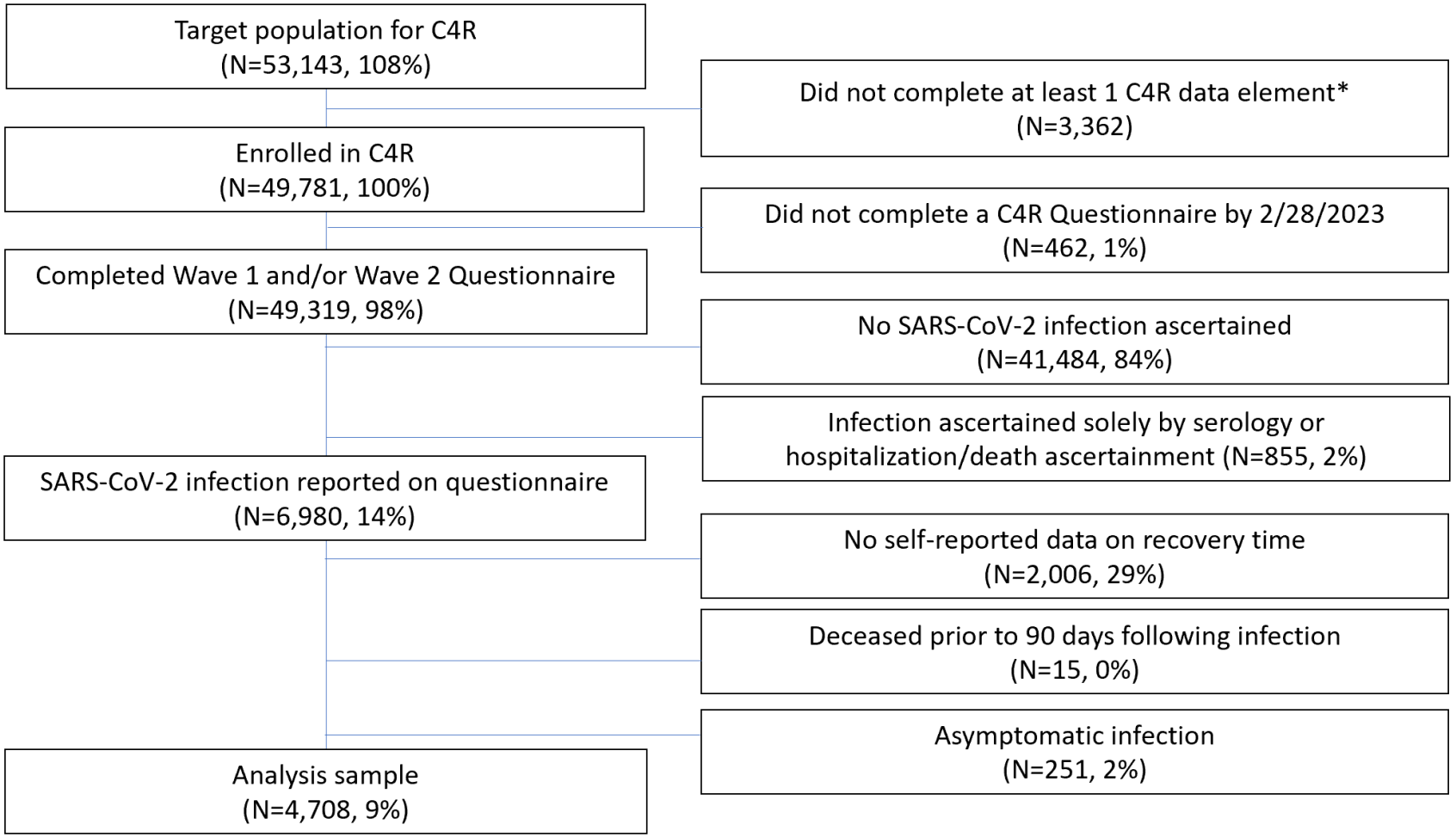
Supplementary Table 11. Main correlates of recovery in time-to-event models, after exclusion of selected cohorts.

N = 3860 (3027 recovered by 90 days)	Categories	Hazard ratio (95% CI) for recovery Stratification by cohort Excluding COPDGene, SHS	Hazard ratio (95% CI) for recovery Stratification by cohort Excluding COPDGene, PrePF, SARP, SPIROMICS
Age (ref: <50 years)	50 – 64 years	1.03 (0.93, 1.14)	1.01 (0.91, 1.11)
	65 – 79 years	1.02 (0.90, 1.16)	1.03 (0.91, 1.16)
	80+ years	1.04 (0.86, 1.27)	1.01 (0.83, 1.23)
Sex (ref: male)	Female	0.84 (0.78, 0.91)	0.86 (0.80, 0.93)
Educational attainment (ref: College)	< High School	1.04 (0.93, 1.16)	1.03 (0.93, 1.15)
	High School	1.03 (0.93, 1.13)	1.02 (0.93, 1.12)
	Some College	1.01 (0.91, 1.12)	1.03 (0.93, 1.15)
Smoking Status (ref: Never)	Former	0.97 (0.89, 1.05)	0.96 (0.88, 1.04)
	Current	0.96 (0.85, 1.08)	0.92 (0.82, 1.04)
BMI (ref: <25 kg/m^2)	25-29 kg/m^2	1.01 (0.92, 1.12)	1.02 (0.92, 1.13)
	30+ kg/m^2	0.90 (0.82, 1.00)	0.90 (0.82, 1.00)
Diabetes (ref: absent)	Present	0.94 (0.85, 1.04)	0.97 (0.88, 1.07)
Hypertension (ref: Absent)	Present	0.95 (0.87, 1.03)	0.96 (0.88, 1.04)
Cardiovascular disease (ref: absent)	Present	0.87 (0.73, 1.03)	0.87 (0.73, 1.03)
Chronic kidney disease (ref: absent)	Present	0.93 (0.76, 1.14)	0.91 (0.74, 1.12)
Asthma (ref: absent)	Present	0.99 (0.88, 1.11)	1.00 (0.88, 1.12)
COPD (ref: absent)	Present	0.95 (0.79, 1.15)	0.90 (0.76, 1.08)
Elevated depressive symptoms (ref: absent)	Present	0.92 (0.82, 1.03)	0.92 (0.82, 1.03)

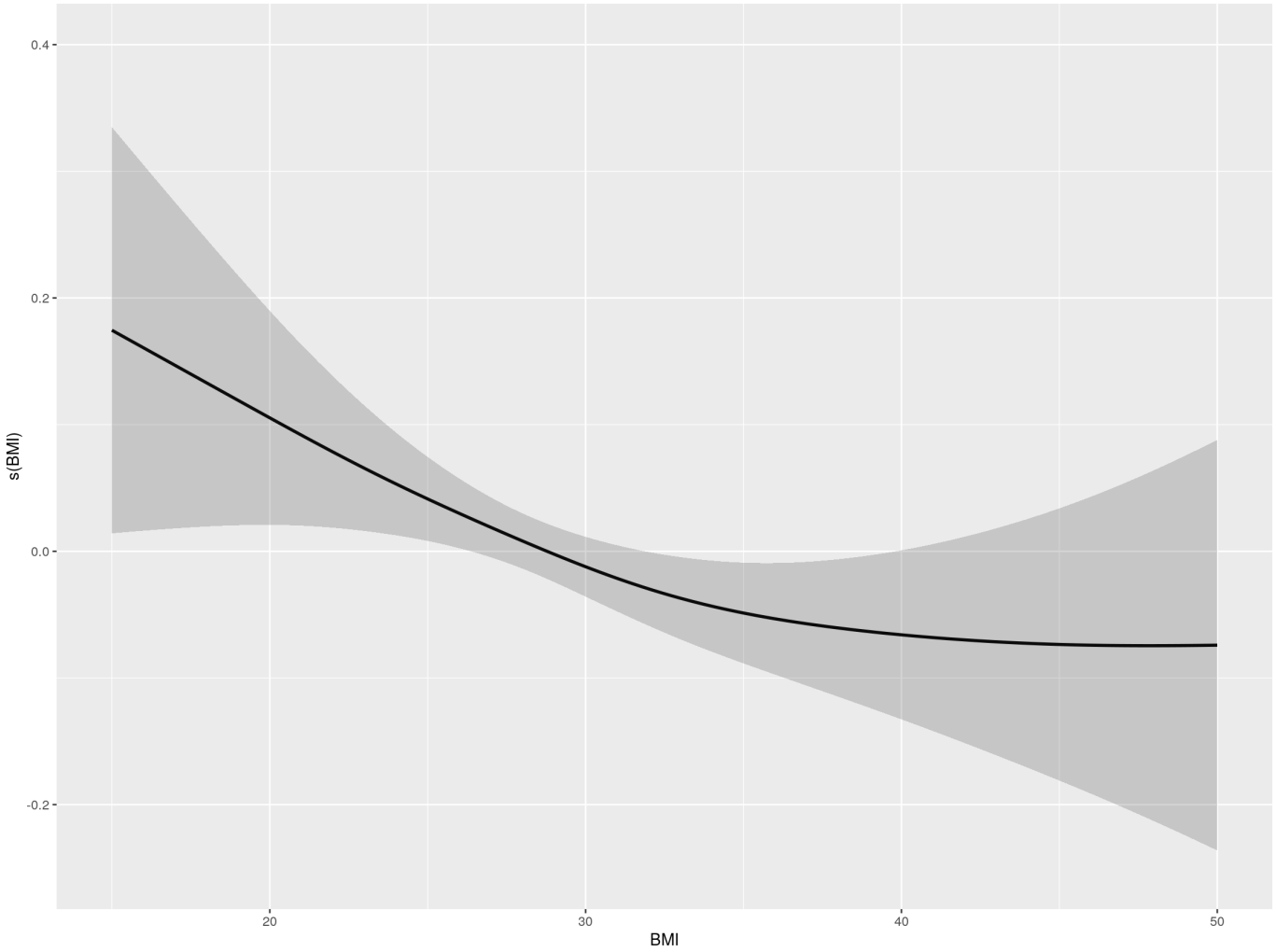
Vaccination prior to infection (ref: no)	Yes	1.19 (1.02, 1.39)	1.16 (1.00, 1.35)
Infection wave (ref: 1st, WT, Spring 2020)	Second (WT, Summer/Fall 2020)	1.00 (0.89, 1.13)	1.00 (0.88, 1.12)
	Third (Alpha, Winter 2020-21)	0.93 (0.84, 1.03)	0.91 (0.82, 1.01)
	Fourth (Spring 2021)	0.90 (0.76, 1.06)	0.88 (0.74, 1.04)
	Fifth (Delta, Summer 2021)	0.96 (0.82, 1.12)	0.93 (0.80, 1.10)
	Sixth (Omicron, Winter 2021-present)	1.20 (1.01, 1.42)	1.18 (0.99, 1.40)

Cox proportional hazards models were performed to assess associations of time-to-recovery with the correlates of interest. Models treated cohort as a stratum term, allowing each cohort to have its own baseline hazard function. Estimates were generated from models adjusted for all of the covariates listed in the table. Hazards ratios (HRs) greater than one indicate faster recovery, whereas HRs less than one indicate slower recovery.

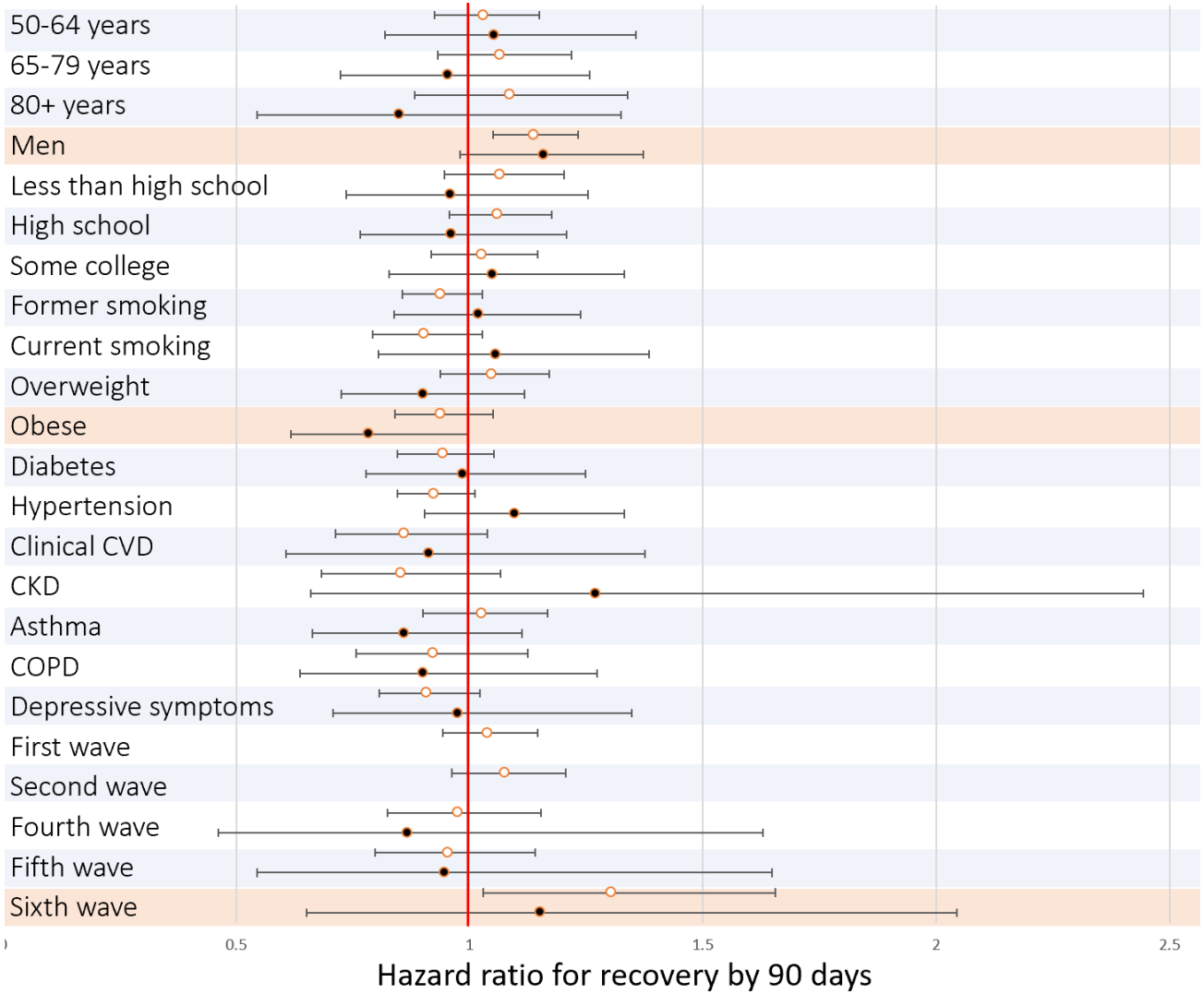
Supplementary Figure 1. CONSORT Diagram of participants included in analyses. *Selected participants did complete a C4R data element but not consent to data sharing on the C4R Analysis Commons by the time of manuscript preparation; these participants are included in N=3,362.



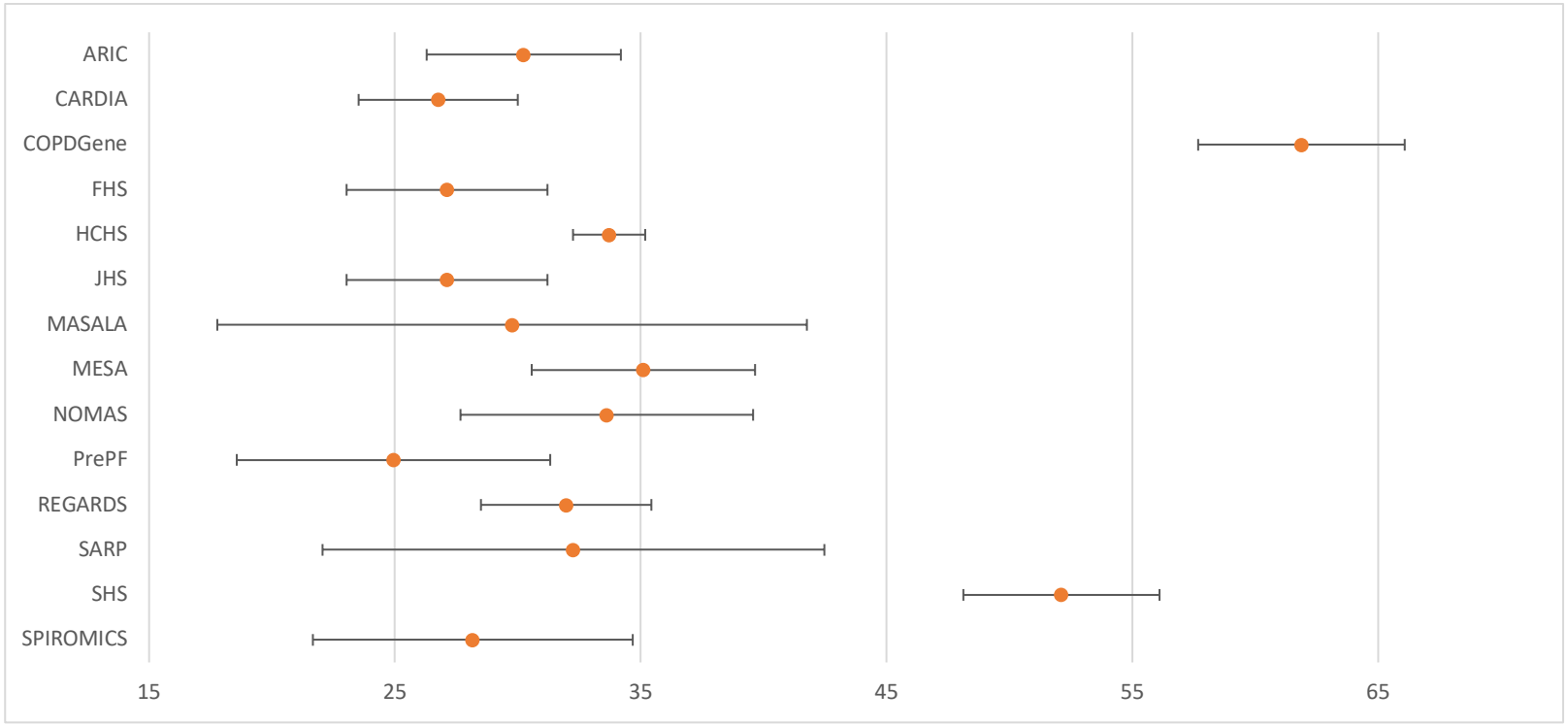
Supplementary Figure 2. Associations of body mass index (BMI, kg/m^2) with recovery by 90 days after SARS-CoV-2 infection. Generalized additive model adjusted for age, sex, race and ethnicity, educational attainment, smoking status, hypertension, diabetes, cardiovascular disease history, chronic kidney disease, asthma, COPD, elevated depressive symptoms, vaccination for SARS-CoV-2, and variant wave, stratified by cohort.



Supplementary Figure 3. Correlates of recovery by 90 days after SARS-CoV-2 infection, after multivariable adjustment, stratified by vaccination status at time of infection. Effect estimates for unvaccinated cases indicated by empty dots, and effect estimates for vaccinated cases indicated by black dots. Comparisons for which the multiplicative interaction term was statistically significant (P-interaction < 0.05: sex, obesity, Omicron variant wave) are highlighted in orange. Null effect (HR = 1.0) indicated by red line.



Supplementary Figure 4. Restricted mean recovery time in days following SARS-CoV-2 infection, by cohort.



REFERENCES

1. Carfì A, Bernabei R, Landi F. Persistent Symptoms in Patients After Acute COVID-19. *Jama*. Aug 11 2020;324(6):603-605. doi:10.1001/jama.2020.12603
2. Rubin R. As Their Numbers Grow, COVID-19 "Long Haulers" Stump Experts. *Jama*. Oct 13 2020;324(14):1381-1383. doi:10.1001/jama.2020.17709
3. Silverberg JI, Zyskind I, Naiditch H, et al. Predictors of chronic COVID-19 symptoms in a community-based cohort of adults. *PLoS One*. 2022;17(8):e0271310. doi:10.1371/journal.pone.0271310
4. Raman B, Cassar MP, Tunnicliffe EM, et al. Medium-term effects of SARS-CoV-2 infection on multiple vital organs, exercise capacity, cognition, quality of life and mental health, post-hospital discharge. *EClinicalMedicine*. Jan 2021;31:100683. doi:10.1016/j.eclinm.2020.100683
5. Mandal S, Barnett J, Brill SE, et al. 'Long-COVID': a cross-sectional study of persisting symptoms, biomarker and imaging abnormalities following hospitalisation for COVID-19. *Thorax*. Apr 2021;76(4):396-398. doi:10.1136/thoraxjnl-2020-215818
6. Peluso MJ, Deeks SG. Early clues regarding the pathogenesis of long-COVID. *Trends Immunol*. Apr 2022;43(4):268-270. doi:10.1016/j.it.2022.02.008
7. Su Y, Yuan D, Chen DG, et al. Multiple early factors anticipate post-acute COVID-19 sequelae. *Cell*. Mar 3 2022;185(5):881-895.e20. doi:10.1016/j.cell.2022.01.014
8. Peluso MJ, Lu S, Tang AF, et al. Markers of Immune Activation and Inflammation in Individuals With Postacute Sequelae of Severe Acute Respiratory Syndrome Coronavirus 2 Infection. *J Infect Dis*. Dec 1 2021;224(11):1839-1848. doi:10.1093/infdis/jiab490
9. Patterson BK, Francisco EB, Yogendra R, et al. Persistence of SARS CoV-2 S1 Protein in CD16+ Monocytes in Post-Acute Sequelae of COVID-19 (PASC) up to 15 Months Post-Infection. *Front Immunol*. 2021;12:746021. doi:10.3389/fimmu.2021.746021
10. Huang C, Huang L, Wang Y, et al. 6-month consequences of COVID-19 in patients discharged from hospital: a cohort study. *Lancet*. Jan 16 2021;397(10270):220-232. doi:10.1016/S0140-6736(20)32656-8
11. Chen C, Haupert SR, Zimmermann L, Shi X, Fritsche LG, Mukherjee B. Global Prevalence of Post-Acute Sequelae of COVID-19 (PASC) or Long COVID: A Meta-Analysis and Systematic Review. *medRxiv*. 2021:2021.11.15.21266377. doi:10.1101/2021.11.15.21266377
12. Hirschtick JL, Titus AR, Slocum E, et al. Population-Based Estimates of Post-acute Sequelae of Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) Infection (PASC) Prevalence and Characteristics. *Clin Infect Dis*. Dec 6 2021;73(11):2055-2064. doi:10.1093/cid/ciab408
13. Yoo SM, Liu TC, Motwani Y, et al. Factors Associated with Post-Acute Sequelae of SARS-CoV-2 (PASC) After Diagnosis of Symptomatic COVID-19 in the Inpatient and Outpatient Setting in a Diverse Cohort. *J Gen Intern Med*. Jun 2022;37(8):1988-1995. doi:10.1007/s11606-022-07523-3
14. Fung KW, Baye F, Baik SH, Zheng Z, McDonald CJ. Prevalence and characteristics of long COVID in elderly patients: An observational cohort study of over 2 million adults in the US. *PLoS Med*. Apr 2023;20(4):e1004194. doi:10.1371/journal.pmed.1004194
15. Horberg MA, Watson E, Bhatia M, et al. Post-acute sequelae of SARS-CoV-2 with clinical condition definitions and comparison in a matched cohort. *Nat Commun*. Oct 12 2022;13(1):5822. doi:10.1038/s41467-022-33573-6
16. Al-Aly Z, Xie Y, Bowe B. High-dimensional characterization of post-acute sequelae of COVID-19. *Nature*. Jun 2021;594(7862):259-264. doi:10.1038/s41586-021-03553-9
17. CDC. Self-Testing At Home or Anywhere. July 18, 2023. https://www.cdc.gov/coronavirus/2019-ncov/testing/self-testing.html

18. Brannock MD, Chew RF, Preiss AJ, et al. Long COVID risk and pre-COVID vaccination in an EHR-based cohort study from the RECOVER program. *Nat Commun*. May 22 2023;14(1):2914. doi:10.1038/s41467-023-38388-7
19. Al-Aly Z, Bowe B, Xie Y. Long COVID after breakthrough SARS-CoV-2 infection. *Nat Med*. Jul 2022;28(7):1461-1467. doi:10.1038/s41591-022-01840-0
20. Arjun MC, Singh AK, Pal D, et al. Characteristics and predictors of Long COVID among diagnosed cases of COVID-19. *PLoS One*. 2022;17(12):e0278825. doi:10.1371/journal.pone.0278825
21. Oelsner EC, Krishnaswamy A, Balte PP, et al. Collaborative Cohort of Cohorts for COVID-19 Research (C4R) Study: Study Design. *Am J Epidemiol*. Jun 27 2022;191(7):1153-1173. doi:10.1093/aje/kwac032
22. Oelsner EC, Krishnaswamy A, Balte PP, et al. Collaborative Cohort of Cohorts for COVID-19 Research (C4R) Study: Study Design. *Am J Epidemiol*. Jun 27 2022;191(7):1153-1173. doi:10.1093/aje/kwac032
23. COVID-19 Treatment Guidelines Panel. Coronavirus Disease 2019 (COVID-19) Treatment Guidelines. National Institutes of Health. Available at https://www.covid19treatmentguidelines.nih.gov/. Accessed May 31, 2023.
24. Times NY. What Previous Covid-19 Waves Tell Us About the Virus Now. Accessed July 21, 2023. https://www.nytimes.com/interactive/2021/10/23/us/covid-surges.html
25. Mackey K, Ayers CK, Kondo KK, et al. Racial and Ethnic Disparities in COVID-19-Related Infections, Hospitalizations, and Deaths : A Systematic Review. *Ann Intern Med*. Mar 2021;174(3):362-373. doi:10.7326/M20-6306
26. Delanaye P, Glassock RJ, Pottel H, Rule AD. An Age-Calibrated Definition of Chronic Kidney Disease: Rationale and Benefits. *Clin Biochem Rev*. Feb 2016;37(1):17-26.
27. Tingley D, Yamamoto T, Hirose K, Keele L, Imai K. mediation: R Package for Causal Mediation Analysis. *Journal of Statistical Software*. 2014;59(5):1-38.
28. Tingley D, Yamamoto T, Hirose K, Keele L, Imai K. mediation: R Package for Causal Mediation Analysis. *Journal of Statistical Software*. 2014;59(5):1-38.
29. *A language and environment for statistical computing.* . R Foundation for Statistical Computing, Vienna, Austria; 2022. https://www.R-project.org/
30. World Health Organization. (2021). A clinical case definition of post COVID-19 condition by a Delphi consensus, 6 October 2021. World Health Organization. https://apps.who.int/iris/handle/10665/345824. .
31. Thaweethai T, Jolley SE, Karlson EW, et al. Development of a Definition of Postacute Sequelae of SARS-CoV-2 Infection. *JAMA*. 2023;doi:10.1001/jama.2023.8823
32. Sigal A. Milder disease with Omicron: is it the virus or the pre-existing immunity? *Nat Rev Immunol*. Feb 2022;22(2):69-71. doi:10.1038/s41577-022-00678-4
33. Nguyen NT, Chinn J, De Ferrante M, Kirby KA, Hohmann SF, Amin A. Male gender is a predictor of higher mortality in hospitalized adults with COVID-19. *PLoS One*. 2021;16(7):e0254066. doi:10.1371/journal.pone.0254066
34. Peckham H, de Gruijter NM, Raine C, et al. Male sex identified by global COVID-19 meta-analysis as a risk factor for death and ITU admission. *Nat Commun*. Dec 9 2020;11(1):6317. doi:10.1038/s41467-020-19741-6
35. Zhang H, Zang C, Xu Z, et al. Data-driven identification of post-acute SARS-CoV-2 infection subphenotypes. *Nat Med*. Jan 2023;29(1):226-235. doi:10.1038/s41591-022-02116-3
36. Klein SL, Flanagan KL. Sex differences in immune responses. *Nat Rev Immunol*. Oct 2016;16(10):626-38. doi:10.1038/nri.2016.90
37. Bowe B, Xie Y, Al-Aly Z. Acute and postacute sequelae associated with SARS-CoV-2 reinfection. *Nat Med*. Nov 2022;28(11):2398-2405. doi:10.1038/s41591-022-02051-3

38. Oelsner EC, Balte PP, Cassano PA, et al. Harmonization of Respiratory Data From 9 US Population-Based Cohorts: The NHLBI Pooled Cohorts Study. *Am J Epidemiol*. Nov 1 2018;187(11):2265-2278. doi:10.1093/aje/kwy139
39. 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
40. 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
41. 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
42. 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
43. 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