American Journal of Epidemiology
Published by Oxford University Press on behalf of the Johns Hopkins Bloomberg School of Public Health 2022.

Vol. 191, No. 7
https://doi.org/10.1093/aje/kwac032
Advance Access publication:
March 12, 2022

Study Design

Collaborative Cohort of Cohorts for COVID-19 Research (C4R) Study: Study Design

Elizabeth C. Oelsner*, Akshaya Krishnaswamy, Pallavi P. Balte, Norrina Bai Allen, Tauqeer Ali, Pramod Anugu, Howard F. Andrews, Komal Arora, Alyssa Asaro, R. Graham Barr, Alain G. Bertoni, Jessica Bon, Rebekah Boyle, Arunee A. Chang, Grace Chen, Sean Coady, Shelley A. Cole, Josef Coresh, Elaine Cornell, Adolfo Correa, David Couper, Mary Cushman, Ryan T. Demmer, Mitchell S. V. Elkind, Aaron R. Folsom, Amanda M. Fretts, Kelley P. Gabriel, Linda C. Gallo, Jose Gutierrez, Mei Lan K. Han, Joel M. Henderson, Virginia J. Howard, Carmen R. Isasi, David R. Jacobs, Jr., Suzanne E. Judd, Debora Kamin Mukaz, Alka M. Kanaya, Namratha R. Kandula, Robert C. Kaplan, Gregory L. Kinney, Anna Kucharska-Newton, Joyce S. Lee, Cora E. Lewis, Deborah A. Levine, Emily B. Levitan, Bruce D. Levy, Barry J. Make, Kimberly Malloy, Jennifer J. Manly, Carolina Mendoza-Puccini, Katie A. Meyer, Yuan-I Nancy Min, Matthew R. Moll, Wendy C. Moore, David Mauger, Victor E. Ortega, Priya Palta, Monica M. Parker, Wanda Phipatanakul, Wendy S. Post, Lisa Postow, Bruce M. Psaty, Elizabeth A. Regan, Kimberly Ring, Véronique L. Roger, Jerome I. Rotter, Tatjana Rundek, Ralph L. Sacco, Michael Schembri, David A. Schwartz, Sudha Seshadri, James M. Shikany, Mario Sims, Karen D. Hinckley Stukovsky, Gregory A. Talavera, Russell P. Tracy, Jason G. Umans, Ramachandran S. Vasan, Karol E. Watson, Sally E. Wenzel, Karen Winters, Prescott G. Woodruff, Vanessa Xanthakis, Ying Zhang, and Yiyi Zhang for the C4R Investigators

* Correspondence to Dr. Elizabeth C. Oelsner, Division of General Medicine, Columbia University Irving Medical Center, 622 West 168th Street, Room PH9-105K, New York, NY 10032 (e-mail: eco7@cumc.columbia.edu).

Initially submitted June 24, 2021; accepted for publication February 9, 2022.

The Collaborative Cohort of Cohorts for COVID-19 Research (C4R) is a national prospective study of adults comprising 14 established US prospective cohort studies. Starting as early as 1971, investigators in the C4R cohort studies have collected data on clinical and subclinical diseases and their risk factors, including behavior, cognition, biomarkers, and social determinants of health. C4R links this pre–coronavirus disease 2019 (COVID-19) phenotyping to information on severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infection and acute and postacute COVID-related illness. C4R is largely population-based, has an age range of 18–108 years, and reflects the racial, ethnic, socioeconomic, and geographic diversity of the United States. C4R ascertains SARS-CoV-2 infection and COVID-19 illness using standardized questionnaires, ascertainment of COVID-related hospitalizations and deaths, and a SARS-CoV-2 serosurvey conducted via dried blood spots. Master protocols leverage existing robust retention rates for telephone and in-person examinations and high-quality event surveillance. Extensive prepandemic data minimize referral, survival, and recall bias. Data are harmonized with research-quality phenotyping unmatched by clinical and survey-based studies; these data will be pooled and shared widely to expedite collaboration and scientific findings. This resource will allow evaluation of risk and resilience factors for COVID-19 severity and outcomes, including postacute sequelae, and assessment of the social and behavioral impact of the pandemic on long-term health trajectories.

cohort studies; coronavirus disease 2019; COVID-19; severe acute respiratory syndrome coronavirus 2

Abbreviations: ARIC, Atherosclerosis Risk in Communities; CARDIA, Coronary Artery Risk Development in Young Adults; CONNECTS, Collaborating Network of Networks for Evaluating COVID-19 and Therapeutic Strategies; COPD, chronic obstructive pulmonary disease; COPDGene, Genetic Epidemiology of COPD; COVID-19, coronavirus disease 2019; C4R, Collaborative Cohort of Cohorts for COVID-19 Research; HCHS/SOL, Hispanic Community Health Study/Study of Latinos; MASALA, Mediators of Atherosclerosis in South Asians Living in America; MESA, Multi-Ethnic Study of Atherosclerosis; NHLBI, National Heart, Lung, and Blood Institute; REGARDS, Reasons for Geographic and Racial Differences in Stroke; SARS-CoV-2, severe acute respiratory syndrome coronavirus 2; TOPMed, Trans-Omics for Precision Medicine.

Adverse effects of the coronavirus disease 2019 (COVID-19) pandemic on US society and on its health and economy are widespread (1). Eighteen months following the initial US outbreak in winter 2020, there have already been over 44 million cases and over 700,000 deaths, making COVID-19 the third-leading cause of death in the United States in 2020 and the second-leading cause of death among persons aged 85 years or more (2, 3). Prolonged symptoms and clinical abnormalities are observed in some COVID-19 survivors, raising concerns that postacute sequelae of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infection could pose an additional long-term health burden (4).

C4R includes 14 US prospective cohort studies that collectively constitute a large, well-characterized, population-based sample of Americans ranging in age from young adults to centenarians and that reflect the racial, ethnic, socioeconomic, and geographic diversity of the United States. C4R uses standardized protocols and active surveillance in an attempt to fully ascertain SARS-CoV-2 infection and COVID-19 illness across all cohorts.

C4R offers the additional major advantages of each component cohort study's long-standing practices of standardized data collection, including robust retention rates and high-quality clinical events surveillance dating back as far as 1971 in some studies. For decades, investigators in the C4R cohort studies have collected extensive longitudinal data on clinical and subclinical disease, behaviors, cognition, biomarkers, and social determinants of health. C4R links this "pre-COVID" phenotyping to information on SARS-CoV-2 infection and acute and postacute COVID-related illness. The integration of antecedent and illness-related data will not only define the consequences of COVID-19 infection reliably but also provide a unique opportunity to understand mechanisms and modifiers of risk and resilience for SARS-CoV-2 infection and adverse COVID-19 outcomes. C4R also supports comparisons of longitudinal changes in health measures over the course of the pandemic in persons with varying degrees of COVID-19 severity. Furthermore, the availability of well-characterized participants unaffected by COVID-19 allows assessment and differentiation of the effects of infection, illness, and pandemic-related social, economic, and behavioral changes.

Overall, C4R aims to provide a scientific resource to 1) evaluate risk and resilience factors for adverse COVID-19 outcomes, including severe illness and long-term complications; 2) assess the social and behavioral impact of the COVID-19 pandemic on long-term outcomes and trajectories of health; and 3) examine disparities in COVID-19 risk and outcomes according to race, ethnicity, geography, and other social determinants of health. This report summarizes the C4R study design and its progress in data collection in the first year of funding (October 1, 2020–September 30, 2021).

METHODS

Cohort of cohorts

Fourteen prospective cohort studies are included in the C4R (Table 1). Eight of the cohort studies were designed to study cardiovascular disease epidemiology: the Atherosclerosis Risk in Communities (ARIC) Study (5), the Coronary Artery Risk Development in Young Adults (CARDIA) Study (6), the Framingham Heart Study (7), the Hispanic Community Health Study/Study of Latinos (HCHS/SOL) (8–10), the Jackson Heart Study (11–13), the Mediators of Atherosclerosis in South Asians Living in America (MASALA) Study (14, 15), the Multi-Ethnic Study of Atherosclerosis (MESA) (16), and the Strong Heart Study (17, 18). Investigators in these studies generally recruited population-based samples, although only 4 studies (ARIC, CARDIA, the Framingham Heart Study, and HCHS/SOL) used representational sampling techniques at some or all sites. Four of the cardiovascular studies (ARIC, CARDIA, the Framingham Heart Study, and MESA) recruited multiracial participants, and 4 were designed to study primarily specific racial or ethnic groups (Hispanic/Latino participants in HCHS/SOL, Black participants in the Jackson Heart Study, South Asian participants in MASALA, and American Indian participants in the Strong Heart Study). Four multiethnic cohorts were established to study respiratory epidemiology: The Genetic Epidemiology of COPD (COPDGene) Study (19) and the Subpopulations and Intermediate Outcome Measures in COPD Study (SPIROMICS) (20) were established as longitudinal case-control studies of cigarette smokers with and without chronic obstructive pulmonary disease (COPD); Prevent Pulmonary Fibrosis (21) is a study of early and established interstitial lung disease; and the Severe Asthma Research Program (22) is a study of the entire range of mild to severe asthma, enriched for severe disease. Two studies—the Northern Manhattan Study (NOMAS) (23) and the Reasons for Geographic and Racial Differences in Stroke (REGARDS) (24)—were established to study primarily neurological outcomes, including stroke and cognition. The Northern Manhattan Study is a multiethnic community study (23),

Table 1. Characteristics of Participants in the C4R Study Cohorts, United States, March 1, 2020

Cohort	No. of Participants	Current Age Range, years	% Female	Race/Ethnicity, %						Original Research Focus
				Non-Hispanic White	Black	Hispanic/ Latinx	Asian-American	American Indian	Other	
ARIC Study	6,690	75–97	63	77	23	0[a]	0	0	0	Cardiovascular
CARDIA Study	4,590	53–66	56	50	50	0	0	0	0	Cardiovascular
COPDGene Study	7,731	50–90	48	65	35	0	0	0	0	Pulmonary
FHS	7,339	26–108	56	86	3	4	0	0	7	Cardiovascular
HCHS/SOL	13,142	30–87	60	0	0	100	0	0	0	Cardiovascular
JHS	2,444	38–102	63	0	100	0	0	0	0	Cardiovascular
MASALA Study	1,132	50–94	47	0	0	0	100	0	0	Cardiovascular
MESA	4,683	65–103	56	38	27	24	12	0	0	Cardiovascular
NOMAS	1,256	62–106	65	12	14	72	0	1	0	Neurological
PrePF Study	5,000	40–80	55	92	3	3	0	0	0	Pulmonary
REGARDS Study	12,766	57–105	58	62	38	0	0	0	0	Neurological
SARP	397	18–80	65	75	25	0	0	0	0	Pulmonary
SPIROMICS	2,273	47–87	48	82	4	4	0	0	0	Pulmonary
SHS	2,915	31–105	62	0	0	0	0	100	0	Cardiovascular

Abbreviations: ARIC, Atherosclerosis Risk in Communities; CARDIA, Coronary Artery Risk Development in Young Adults; COPD, chronic obstructive pulmonary disease; COPDGene, Genetic Epidemiology of COPD; COVID-19, coronavirus disease 2019; C4R, Collaborative Cohort of Cohorts for COVID-19 Research; 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; SHS, Strong Heart Study; SPIROMICS, Subpopulations and Intermediate Outcome Measures in COPD Study.

[a] The ARIC investigators did not inquire about Hispanic/Latino ethnicity; hence, White participants in the ARIC Study cannot be definitely defined as non-Hispanic.

and REGARDS is a biracial (non-Hispanic Black, White) national sample of the continental United States that oversampled Black people and residents of the Southeast (24).

These cohort studies have collected detailed data on participants' health and behavior for as long as 50 years (Figures 1 and 2). Investigators in the C4R cohorts have performed extensive longitudinal (repeated) phenotyping of subclinical and clinical disease, as well as assessments of laboratory biomarkers, -omics, imaging, diet, behavior, and social determinants of health, and have extensive biorepositories of stored specimens (see Web Table 1, available at https://doi.org/10.1093/aje/kwac032). Twelve cohorts have geocoding available, supporting participant-level assessment of neighborhood socioeconomic status, exposures to systemic racism, and environmental exposures such as air pollution. All of these cohort studies use similar or identical adjudication protocols to ascertain all-cause mortality. Ten of them ascertain cardiovascular events, including myocardial infarction, stroke, and heart failure. Eight of them ascertain respiratory events such as COPD and asthma exacerbations. Seven ascertain incident cognitive impairment and/or dementia.

Collaboration and governance

Most cohort studies in the C4R have a history of collaboration with the genomics-oriented Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium (25); the National Heart, Lung, and Blood Institute (NHLBI) Pooled Cohorts Study, focusing on respiratory epidemiology (26); the Cross-Cohort Collaboration Consortium, for cardiovascular epidemiology (27); the Blood Pressure and Cognition Study (28); and the genetic sequencing– and multiomics-focused Trans-Omics for Precision Medicine (TOPMed) Project (29). C4R builds and expands upon these successes to advance COVID-19 research.

Planning for C4R began in March 2020, when the need for a coordinated, cross-cohort response to the knowledge gaps posed by the COVID-19 pandemic was self-evident and urgent. Cohort investigators initiated discussions regarding approaches to ascertain SARS-CoV-2 infections and COVID-related illnesses within the context of unprecedented cohort operational challenges associated with the outbreak. The NHLBI funded C4R via an Other Transactional Authority mechanism in October 2020. Additional funding for inclusion of the neurology-focused cohorts was provided via the Other Transactional Authority by the National Institute of Neurological Disorders and Stroke and the National Institute on Aging.

The collaborative governance structure of the C4R is shown in Figure 3. The administrative coordinating center is the NHBLI Collaborating Network of Networks for Evaluating COVID-19 and Therapeutic Strategies (CONNECTS) Program, which reviews C4R operational milestones biweekly. Central C4R functions are overseen by an observational studies monitoring board convened by CONNECTS.

Leadership for C4R is provided by an organizing committee that includes leading—and often, founding—principal investigators from all C4R cohorts, principal investigators from the C4R Data Coordination and Harmonization Center at Columbia University Irving Medical Center (New York, New York), principal investigators from the C4R Biorepository and Central Laboratory at the University of Vermont (Burlington, Vermont), and program officers from the NHLBI, the National Institute of Neurological Disorders and Stroke, and the National Institute on Aging. Master C4R protocols for COVID-19 data collection were drafted, piloted, and refined using a consensus-driven approach by dedicated subcommittees that included content-area experts from each cohort study. Draft protocols were reviewed, refined, and approved by the organizing committee. Protocols were designed to provide flexibility for multimodal data collection, including remote options that could be used by both on-site and off-site study personnel, and are available on the study's website (https://c4r-nih.org).

Consistent with an ancillary studies model, researchers in each cohort study are directly responsible for accomplishing their own data collection in accordance with the master protocols and under the supervision of their own observational studies monitoring board, steering committee, and any other applicable regulatory authorities. To maintain full participation, cohort investigators are permitted to adapt protocols to their cohort-specific needs.

The C4R Biorepository and Central Laboratory is responsible for establishing a C4R biorepository of dried blood spots, plus other biospecimens that may be collected in the future, and for performing and/or coordinating the performance of centralized clinical and biomarker assays and serological assays.

Data collection is coordinated centrally at the Data Coordination and Harmonization Center. Electronic data collection forms are programmed into REDCap (30, 31) for use or adaptation by the cohort data coordinating centers. Metadata on completion of questionnaires, events ascertainment, and dried blood spots are reported to the Data Coordination and Harmonization Center biweekly. The Data Coordination and Harmonization Center maintains the C4R website (https://c4r-nih.org), which includes a password-protected investigator section with regular status updates and study materials.

To promote and sustain this broad collaborative effort, C4R principal investigators have invited additional investigators and cohort personnel to participate in subcommittees and working groups. In this manner, C4R has engaged over 180 investigators to date.

Participants

Cohort participants previously consented to in-person, telephone, and/or e-mail contact and for abstraction of medical records. Additional consent for ascertainment of COVID-19 data, including the serosurvey, is obtained according to cohort-specific procedures, including verbal, remote, and traditional written informed consent.

All cohort participants who were alive on March 1, 2020, and had not withdrawn consent for cohort participation were considered eligible for enrollment in the C4R. Of 72,358 participants who were believed to meet these inclusion criteria, the C4R principal investigators estimated that

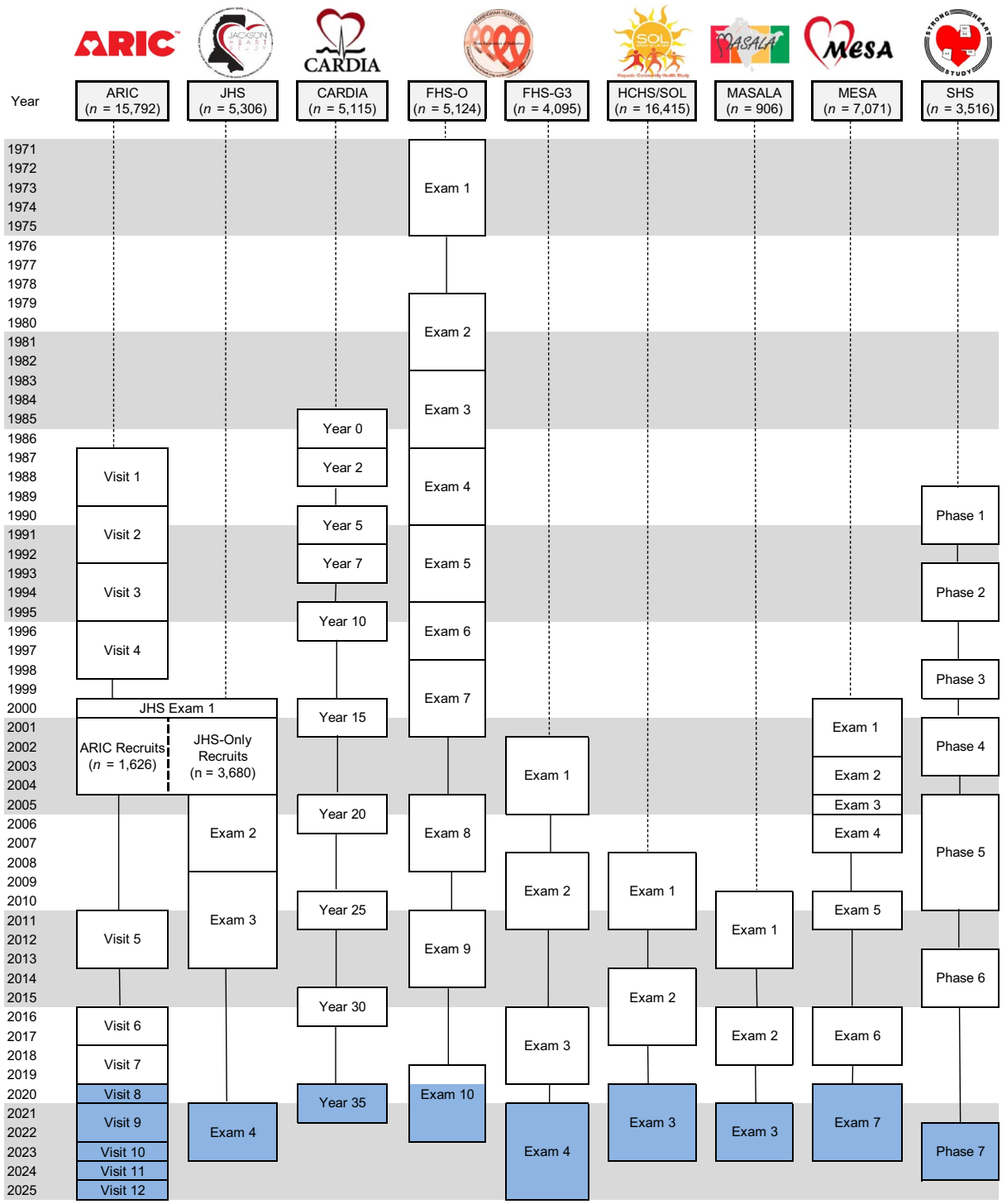

Figure 1. Longitudinal pre–coronavirus disease 2019 (COVID-19) follow-up and planned follow-up for cardiovascular cohort studies participating in the Collaborative Cohort of Cohorts for COVID-19 Research (C4R) Study, by cohort, United States, 1971–2025. Some study visits were overlapping, which is not shown; instead, midpoints of the visits are indicated. COVID-era examinations (exams) are shaded in blue. Solid lines indicate cohort follow-up, which typically includes regular contact by telephone and mail and ongoing ascertainment of health events. In the Atherosclerosis Risk in Communities (ARIC) Study, 424 participants gave restricted consent. The Jackson Heart Study (JHS) included 1,626 participants recruited from ARIC; it also included participants not recruited from ARIC (JHS-Only). One participant in the Coronary Artery Risk Development in Young Adults (CARDIA) Study withdrew consent. The Multi-Ethnic Study of Atherosclerosis (MESA) cohort included the original MESA participants plus 257 new participants recruited into the MESA Air Pollution Study. FHS-G3, Framingham Heart Study Third Generation; FHS-O, Framingham Heart Study Offspring Study; HCHS/SOL, Hispanic Community Health Study/Study of Latinos; MASALA, Mediators of Atherosclerosis in South Asians Living in America; SHS, Strong Heart Study.

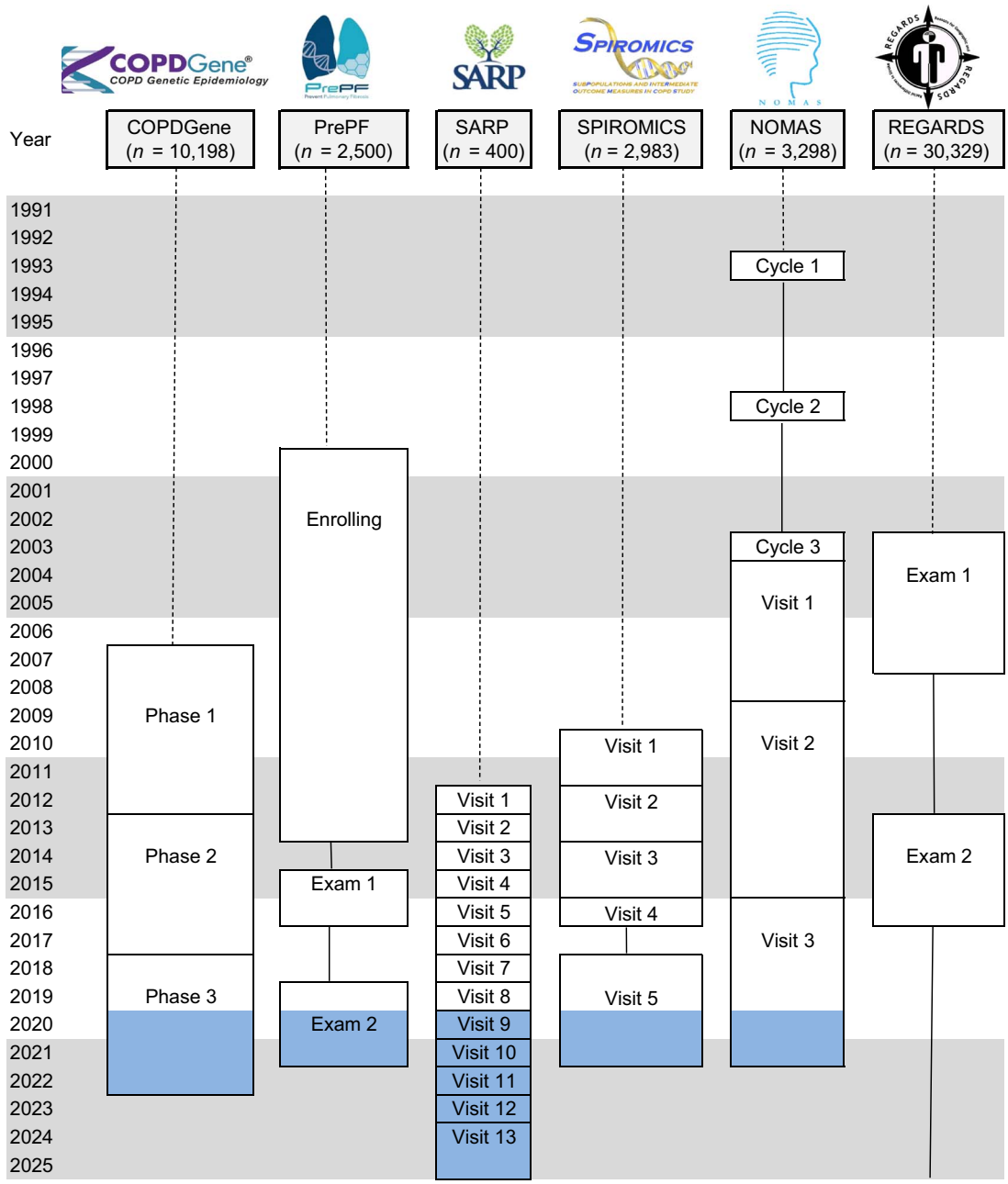

Figure 2. Longitudinal pre–coronavirus disease 2019 (COVID-19) follow-up and planned follow-up for pulmonary and neurological cohorts in the Collaborative Cohort of Cohorts for COVID-19 Research (C4R) Study, by cohort, United States, 1991–2025. Some visits were overlapping, which is not shown; instead, midpoints of the visits are indicated. COVID-era examinations (exams) are shaded in blue. Solid lines indicate cohort follow-up, which typically includes regular contact by telephone and mail and ongoing ascertainment of health events. COPD, chronic obstructive pulmonary disease; COPDGene, Genetic Epidemiology of COPD; 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.

53,143 participants, hereafter described as the C4R target population, were readily available for recruitment into C4R based on recent participation in cohort follow-up calls. The sociodemographic characteristics of participants eligible for C4R (Web Table 2) are similar to those of the C4R target population (Table 2). Fifty-eight percent of participants in the target population are aged 65 years or older, and thus at high risk for severe COVID-19. The anticipated sample is racially and ethnically diverse, based on self-reporting (32), with approximately 6% American Indian participants,

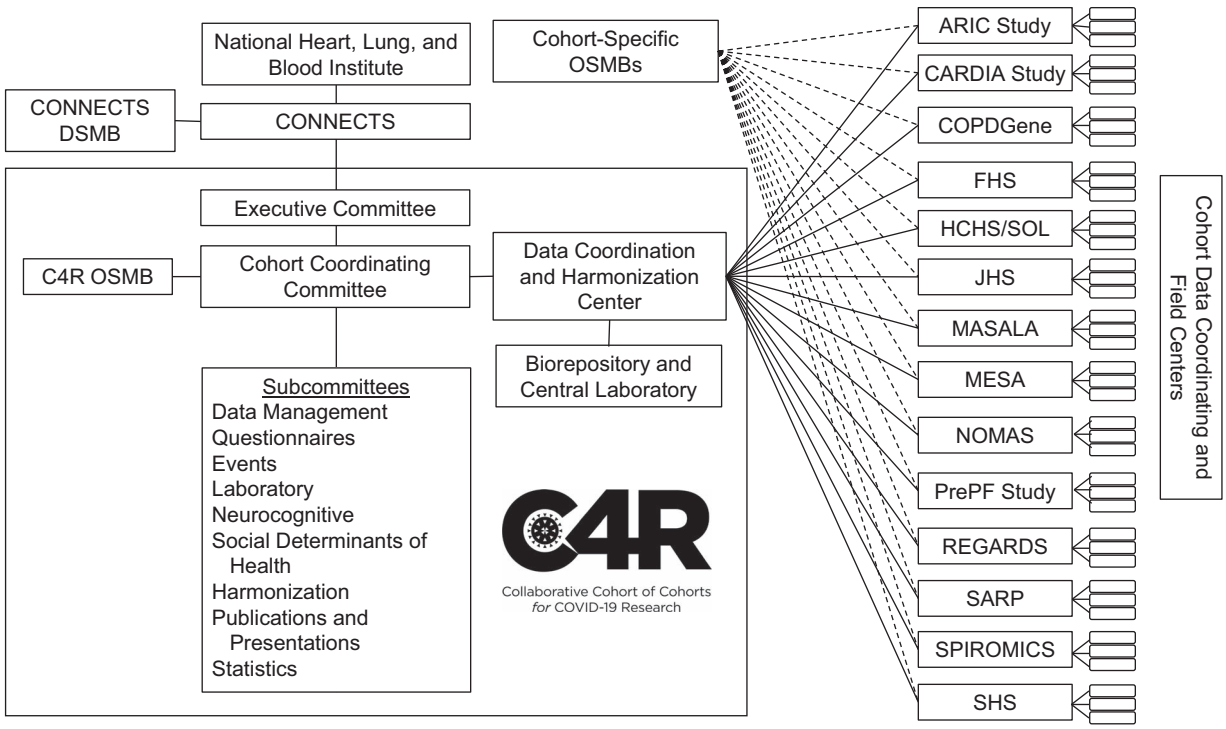

Figure 3. Organization of the Collaborative Cohort of Cohorts for COVID-19 Research (C4R) Study, 1991–2025. ARIC, Atherosclerosis Risk in Communities; CARDIA, Coronary Artery Risk Development in Young Adults; CONNECTS, Collaborating Network of Networks for Evaluating COVID-19 and Therapeutic Strategies; COPD, chronic obstructive pulmonary disease; COPDGene, Genetic Epidemiology of COPD; COVID-19, coronavirus disease 2019; DSMB, Data Safety Monitoring Board; 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; OSMB, Observational Studies Monitoring Board; PrePF, Prevent 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.

2% Asian participants, 26% Black participants, and 20% Hispanic/Latino participants.

All 48 continental states are represented among C4R participants, including rural, suburban, and urban communities (Figure 4). C4R is being conducted across 40 field/clinical centers, many of which are associated with more than 1 C4R cohort; one cohort with extensive geographic reach, REGARDS, operates via telephone and in-home examinations only (24).

COVID-19 questionnaires

Each cohort study was funded to deploy COVID-19 questionnaires twice within 18 months following the initial outbreak in March 2020 via telephone, mail, the Internet, e-mail, or smartphone applications. Both the first questionnaire administration, called wave 1, and the second questionnaire administration, called wave 2, were attempted across the entire target population.

Various COVID-19 questionnaires were developed as early as March 2020 in certain cohorts (33) and administered in spring and summer 2020. Although these efforts predated funding of C4R, early informal cross-cohort collaborations ensured that many cohorts used identical questionnaires, and all of them generated common data elements regarding infection, testing, hospitalization, and recovery. Following funding, the C4R questionnaire was developed to include domains on COVID-19 infection, testing, hospitalization, symptoms, recovery, reinfection, contacts, vaccination, behavioral changes, sleep, memory loss, depression, anxiety, fatigue, and resilience. This C4R questionnaire includes validated and PhenX toolkit instruments (34–43) in order to optimize comparability with prepandemic assessments and across C4R and other cohort studies. The C4R questionnaire, including translations into Spanish and Mandarin, is available on the PhenX (https://www.phenx.org/) and C4R (https://c4r-nih.org) websites; REDCap (30, 31) programming is available upon request. The C4R questionnaire was used by 3 cohorts in wave 1 and was adapted for use by all 14 cohorts in wave 2. Comparisons of cohort-specific instruments are provided in Web Tables 3 and 4.

As of September 30, 2021, wave 1 was completed by 13 of the 14 cohorts and wave 2 was initiated in all cohorts (Web Figure 1). Characteristics of participants completing the wave 1 questionnaires as of this date were similar to those of the target population (Table 2).

Table 2. Characteristics of the C4R Target Population and Participants Completing the C4R Wave 1 Questionnaire, March 1, 2020–September 30, 2021

Cohort	Target Population (n = 53,143)		Participants Completing Wave 1 Questionnaire[a] (n = 45,262)	
	No.	%	No.	%
Sex				
Female	30,557	57.50	26,814	59.24
Male	22,586	42.50	18,448	40.76
Race				
Asian	1,238	2.33	1,044	2.31
American Indian/Alaska Native	2,971	5.59	1,923	4.25
Black	13,722	25.82	9,860	21.78
Native Hawaiian/Pacific Islander	32	0.06	34	0.08
White	28,655	53.92	25,070	55.39
Other	3,109	5.85	4,544	10.04
Multiracial	2,003	3.77	1,954	4.32
Unknown	1,413	2.66	833	1.84
Ethnicity				
Hispanic	10,698	20.13	12,677	28.01
Non-Hispanic	34,362	64.66	25,549	56.45
Unknown	8,083	15.21	7,036	15.55
Age group, years				
18–29	654	1.23	388	0.86
30–64	21,581	40.61	15,697	34.68
≥65	30,908	58.16	29,177	64.46
Cohort				
ARIC Study	5,046	9.50	5,466	12.08
CARDIA Study	4,221	7.94	2,530	5.59
COPDGene Study	4,000	7.53	3,764	8.32
FHS	7,339	13.81	3,173	7.01
HCHS/SOL	8,400	15.81	11,152	26.64
JHS	2,317	4.36	1,697	3.75
MASALA Study	500	0.94	460	1.02
MESA	4,683	8.81	3,450	7.62
NOMAS	1,256	2.36	887	1.96
PrePF Study	2,500	4.70	614	1.36
REGARDS Study	8,000	15.05	8,750	19.33
SARP	380	0.72	326	0.72
SPIROMICS	1,800	3.39	1,483	3.28
SHS	2,701	5.08	1,510	3.34

Abbreviations: ARIC, Atherosclerosis Risk in Communities; CARDIA, Coronary Artery Risk Development in Young Adults; COPD, chronic obstructive pulmonary disease; COPDGene, Genetic Epidemiology of COPD; COVID-19, coronavirus disease 2019; C4R, Collaborative Cohort of Cohorts for COVID-19 Research; 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; SHS, Strong Heart Study; SPIROMICS, Subpopulations and Intermediate Outcome Measures in COPD Study.

[a] Participants who had completed the C4R wave 1 questionnaire as of September 30, 2021.

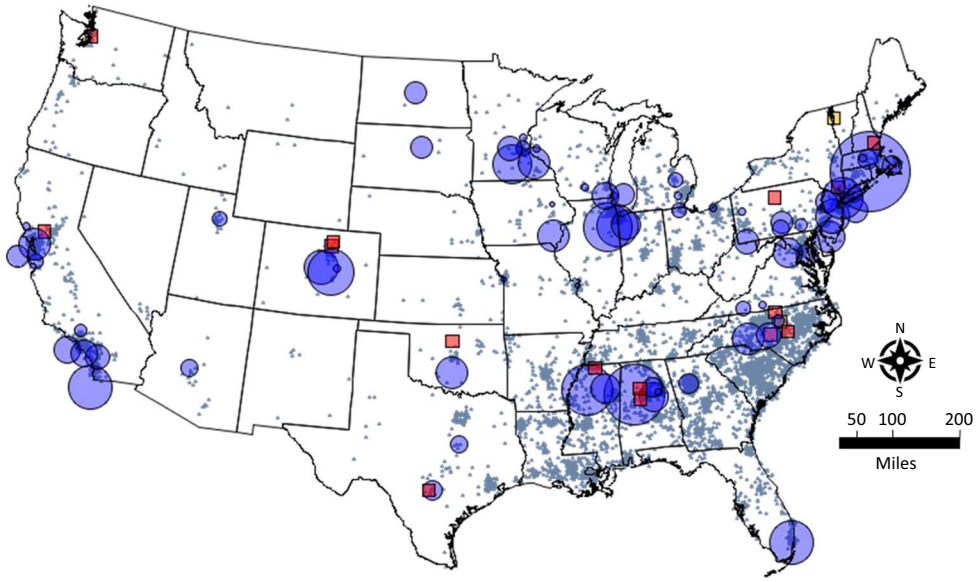

Figure 4. Participants, field/clinical centers, and coordinating centers in the Collaborative Cohort of Cohorts for COVID-19 Research (C4R) Study, 1991–2025. Blue circles indicate field/clinical centers, and their sizes are proportional to the number of participants at each field/clinical center. Participants in the Reasons for Geographic and Racial Differences in Stroke (REGARDS) Study, which does not have field/clinical centers, are shown by additional blue shading according to their geocoded home addresses. Red squares indicate coordinating centers involved in the study. Yellow squares indicate C4R central resources: the Data Coordination and Harmonization Center, the Biorepository and Central Laboratory, and the Administrative Coordinating Center. COVID-19, coronavirus disease 2019.

COVID-related events

C4R ascertains COVID-related hospitalizations and deaths that are identified via the C4R questionnaire or other surveillance methods available to the cohort investigators, including electronic health record linkages where available. Each cohort study uses its own established infrastructure for ascertainment of medical records and death certificates, including the National Death Index, the Centers for Medicare and Medicaid Services, *International Classification of Diseases, Tenth Revision, Clinical Modification*, codes (44), and linkage to records from local health departments. Cohort investigators review events locally at their field/coordinating centers or transfer records for central review by C4R. The C4R events review assesses severity and major complications of COVID-19 illness, including pneumonia, myocardial infarction, stroke, thromboembolism, and acute renal failure. The protocols use or are modeled after longstanding cohort protocols for classifying and validating cardiovascular (5–7, 9, 10, 15, 16, 18, 23, 45), respiratory (46), and thromboembolic (47) events. Protocols for ascertainment, review, and classification are available on the study website. As of 1 year after C4R funding, over 1,000 COVID-related hospitalizations or deaths had been ascertained across the consortium.

Dried blood spot collection

Kits for collection of dried blood spots are produced by the C4R Biorepository and Central Laboratory and shipped to cohort investigators (either to the individual field centers or to the cohort study coordinating center, based on the investigators' preference). Dried blood spot cards are labeled with a biospecimen identifier, which is linked to C4R identifiers that are maintained centrally and not shared with the Biorepository and Central Laboratory, through the use of a "linking key." Cohort field centers receive dried blood spot collection kits from the Biorepository and Central Laboratory and are responsible for recruitment, consent, and distribution to participants. Updated details regarding vaccination status are obtained at the time of consent for dried blood spot collection and immediately prior to mailing of the dried blood spot kit to the participant. Participant instructions, including a video, are provided by the cohort investigators and via the C4R website and/or cohort-specific websites. Participants mail the completed kits directly to the Biorepository and Central Laboratory or to the cohort field or coordinating center as an intermediary step. In cohort studies with upcoming in-person examinations, the dried blood spot may be collected in person by research staff. Dried blood spot collection was initiated in February 2021 (Web Figure 1) and is ongoing, with over 10,000 dried blood spots acquired as of September 30, 2021.

Serological analysis

After preprocessing of completed dried blood spot cards by the Biorepository and Central Laboratory, serological assays are performed by the New York State Wadsworth

Center's Bloodborne Viruses Laboratory under the Clinical Laboratory Improvement Amendments and New York State certification. The Bloodborne Viruses Laboratory performs a SARS-CoV-2 immunoglobulin G microsphere immunoassay using Luminex bead technology (Luminex Corporation, Austin, Texas) for qualitative detection of human immunoglobulin G antibodies to SARS-CoV-2 nucleocapsid (N) and spike subunit 1 (S1) antigens. Based on testing of 730 pre–COVID-19 dried blood spots and more than 1,100 dried blood spots from persons with laboratory-confirmed infection, specificity was 99.5% for both the N and S1 antigens, and sensitivity ranged from 90% to 96% for symptomatic individuals and from 77% to 91% for asymptomatic individuals. Sensitivity increased for both groups with amount of elapsed time since a positive polymerase chain reaction test, accounting for the range. This assay was used successfully to test over 57,000 dried blood spots for statewide serosurveys carried out from April to June 2020 as part of New York State's public health response. Serological results are reported by the Bloodborne Viruses Laboratory to the Biorepository and Central Laboratory and then to the cohort data coordinating centers, which are responsible for 1) recombining the results with the proper participants based on the "linking key" and 2) reporting results to participants according to usual cohort practices. Serological results are not known to have clinical relevance, and the Centers for Disease Control and Prevention does not currently recommend modifications to individual behavior or clinical care based on antibody status alone (48); hence, no protocols for "alert" findings were established, and participants may opt out of results return. Protocols for the serosurvey are available on the study website.

Since all current vaccines in use in the United States generate an immune response to the SARS-CoV-2 spike protein, antibody responses to vaccination versus viral infection may be distinguished by the anti–nucleocapsid assay results (49).

Quality control

Cohort investigators use established protocols for checking data completeness and accuracy at the field center and coordinating center levels. Dual data entry for C4R is encouraged but not required, since it is not feasible in all settings due to local impediments and COVID-related exigencies. Ten percent of event reviews are randomly selected for re-review. Reviewers not meeting standards receive regular feedback with recommendations for retraining and/or protocol modifications, as appropriate. Serological assays are repeated on a random 5% subsample of blind duplicates.

COVID-19 outcomes

C4R data define a spectrum of COVID-19 outcomes, including those listed in Web Table 5. Ascertainment of COVID-related hospitalizations and deaths characterizes, classifies, and validates moderate-to-severe COVID-19 illnesses. Questionnaires obtain self-reported information on the nature, severity, and duration of symptoms in the acute and postacute setting, supporting classification of symptomatic and asymptomatic infections and cases of prolonged recovery or postacute sequelae. Data on behaviors, attitudes, psychosocial impacts, and vaccinations are also collected. Seropositive individuals without self-reported infection may be reclassified as infected, whereas seronegative individuals with prior positive testing may be classified as seroreverted.

Harmonization

Harmonization of COVID-19 and prepandemic data is performed centrally by the Data Coordination and Harmonization Center on the C4R Analysis Commons to define COVID-19 outcomes and to align prepandemic data for large-scale, longitudinal analyses. This effort leverages prior harmonization efforts across C4R cohorts in the TOPMed Project, the NHLBI Pooled Cohorts Study, the Blood Pressure and Cognition Study, and the CHARGE Working Groups (25, 26, 28, 50–56). Core measures that have already been harmonized across the majority of C4R cohorts are available on the study website; additional variable harmonization is guided by scientific priority and the data needs of approved manuscript proposals. As in prior published efforts (26, 57), major steps in data harmonization include identification of variables of interest, review of available data in consultation with cohort-specific investigators and analysts at cohort data coordinating centers, and qualitative assessments of data collection instruments and data dictionaries. Variables are aligned to determine differences in measurement and classification. Next, candidate variables for harmonization are transferred to the C4R Analysis Commons for quantitative assessment, relabeling, and recoding according to a common C4R standard. Quantitative comparisons are made within and between individuals and cohorts to identify outliers and missing data. Data quality issues are investigated and corrected in collaboration with cohort data coordinating centers. Harmonized and derived variables, plus codes used to generate them, are shared with the source cohorts.

Because of its significance to COVID-19 epidemiology, particular emphasis is being placed on harmonizing prepandemic physiological (26), neurocognitive (28, 58–63), and imaging-based (64–71) phenotyping data collected within the decade prior to the outbreak (Table 3). Harmonization of lung computed tomography scans is being accomplished using deep learning (72–75) and other methods, which will be discussed in separately published reports.

Data management

The C4R Commons Agreement, modeled on the CHARGE Analysis Commons Consortium Agreement (76), is expediting cross-cohort data harmonization and sharing, as allowed (77). Following review and approval, cohort-specific agreements permit COVID-19 and prepandemic data to be uploaded to the C4R Analysis Commons, which is

Table 3. Estimated Numbers of Participants With Recent Prepandemic Deep Phenotyping for Harmonization in the C4R Study, by Cohort, United States, 2010–2020[a]

Measure	Cohort														
	ARIC Study	CARDIA Study	COPDGene Study	FHS	HCHS/SOL	JHS	MASALA Study	MESA	NOMAS	PrePF Study	REGARDS Study	SARP	SPIROMICS	SHS	C4R Study
Physical function															
Timed walk	3,140		3,879	2,257				2,473					1,064		12,813
Hand grip strength	3,140		1,342	6,232					628				1,160		12,502
Spirometry	3,612	3,119	4,000	5,914	8,400	2,317		3,502		1,123		380	1,149	94	33,610
DLCO	457		3,075	5,914						1,123					10,569
Resting oxygen level	11		4,000		8,400			3,973					1,069		17,453
CT scanning															
Lung CT			3,775	2,799				3,459		1,059		229	1,138		12,459
Cardiac CT	2,267	3,068		2,799		2,317	500	2,801							13,752
Dual-energy CT								731					330		1,061
Any CT	2,267	3,068	3,775	2,799		2,317	500	3,549		1,059		229	1,138		20,701
Cardiac measures															
Cardiac MRI						1,421		2,556					67		4,044
Echocardiogram	2,973	3,115			6,427	2,317		2,920	502				82	1,901	20,237
ECG	2,409			7,258	8,400	2,317	500	3,802	502		7,778			1,910	34,876
Brain measures															
Brain MRI	771	653		3,543	1,245			1,113	803						8,128
Neurocognitive testing															
Long testing	3,589	3,354		2,477	8,400			1,487	600		528			817	21,252
Short testing			1,360	3,934	8,400			3,788	1,260		8,000			817	27,559
Any testing	3,589	3,354	1,360	3,800	8,400			3,857	1,260		8,000			817	34,437
Sleep and activity measures															
Polysomnography	11	835			8,400	913		1,779							11,938
Actimetry	513	1,397		4,100	8,400	852		1,828							17,090
ECG monitoring	2,257							1,510	300						4,067

Table continues

Table 3. Continued

Measure	Cohort														
	ARIC Study	CARDIA Study	COPDGene Study	FHS	HCHS/ SOL	JHS	MASALA Study	MESA	NOMAS	PrePF Study	REGARDS Study	SARP	SPIROMICS	SHS	C4R Study
Biomarkers															
Blood sample	5,046	4,221	4,000	7,258	8,400	2,317	500	4,683	1,267	2,500	8,000	380	1,800	2,701	53,073
Urine sample	5,046	4,221		7,258	8,400	2,317	500	3,900	1,267		8,000		1,800	2,701	45,410
GWAS	4,541	3,799	4,000	6,817	8,400	2,317		4,455		2,250		380	1,620		38,579
RNA sequencing (blood sample)			800	2,730				1,082		1,000		342	1,800		7,754
Metabolomics				3,025	8,000	2,317		812		100				2,701	16,955
Methylation		3,799	4,000	1,900	3,000	1,752		1,212		1,000				2,325	18,988
Proteomics				2,813		1,852		812		250					5,727
Sputum/ bronchoscopy												380	1,000		1,380
Gut microbiome		607			8,000										8,607

Abbreviations: ARIC, Atherosclerosis Risk in Communities; CARDIA, Coronary Artery Risk Development in Young Adults; COPD, chronic obstructive pulmonary disease; COPDGene, Genetic Epidemiology of COPD; COVID-19, coronavirus disease 2019; C4R, Collaborative Cohort of Cohorts for COVID-19 Research; CT, computed tomography; DLCO, diffusing capacity of the lung for carbon monoxide; ECG, electrocardiogram; FHS, Framingham Heart Study; GWAS, genome-wide association 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; MRI, magnetic resonance imaging; 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] If the most recent examination was conducted prior to 2010, data are not included.

located on the NHLBI's cloud computing platform, BioData Catalyst (https://biodatacatalyst.nhlbi.nih.gov/) (78). This platform has enterprise-grade compliance and security certification and is HIPAA-compliant (Health Insurance Portability and Accountability Act of 1996) and compliant with Database of Genotypes and Phenotypes and Clinical Laboratory Improvement Amendments security best practices. Nonetheless, no protected health information is retained in the C4R Analysis Commons. Participants are assigned a C4R study identifier by the cohort-specific data coordinating center that is distinct from the original cohort participant identifier. Time-to-assessment and time-to-event are calculated as latencies with a random offset.

Access to pooled C4R data is regulated by the C4R publications and presentations policy, which is available on the C4R website (https://c4r.nih.org). Data are made available for analyses in the C4R Analysis Commons for investigators with manuscript proposals approved by the C4R publications and cohort coordinating committees, as well as by each cohort included in a given proposal. Investigators are not permitted to access or analyze data for which there are relevant consent restrictions. Data downloads by investigators from the C4R Analysis Commons are prohibited. Recommendations to address common data issues anticipated for C4R analyses, such as missing data, are developed by the C4R Statistical Subcommittee, and related programming is shareable in the C4R Analysis Commons. Once harmonization and related quality control are completed, C4R common data elements will be transferred as a limited data set for public access on BioData Catalyst in accord with cohort-specific consents and commitments.

DISCUSSION

C4R leverages existing US cohort studies to develop a large, multiethnic, pooled cohort of participants with incident COVID-19 and COVID-unaffected participants. C4R includes a diverse population of US adults, including older and socially disadvantaged populations that have especially high risk of adverse COVID-19 outcomes. C4R is distinguished from other large studies of COVID-19 by its wealth of prepandemic phenotyping, providing unique opportunities to evaluate a range of risk and resilience factors for SARS-CoV-2 infection and adverse COVID-19 outcomes, including severe COVID-19 illness, postacute sequelae, and other long-term effects of the pandemic response. Unlike case registries and electronic-health-record–based studies, C4R's repeated examinations and cognitive assessments completed before and after the COVID-19 pandemic also provide important opportunities to estimate the social and behavioral impact of the COVID-related pandemic response on changes in long-term mental and physical health across multiple domains.

C4R constitutes a historic initiative to standardize and expedite data collection by US cohort studies despite major operational and societal challenges. Commitment from cohort principal investigators with experience in cross-cohort collaboration was critical to expediting the study design, protocol development, approval, and deployment. The large scale of data collection was made possible not only by a cultural shift toward multistudy collaborative efforts occasioned by experience in CHARGE and TOPMed but also by the established cohort infrastructure, including sophisticated data coordinating centers that were prepared to implement new protocols rapidly and experienced and professional clinical staff with long-standing relationships with cohort participants. Differences in the data collected by individual cohort studies are being reconciled through harmonization, where possible, or else they will be accounted for in analysis plans, in which cohorts without a necessary data element may be excluded. A similar approach is being used with respect to pre-COVID data heterogeneity; fortunately, many cohort studies used similar or identical protocols for data collection in the pre-COVID era, due often to overlapping investigative groups, facilitating harmonization efforts. The challenges of sharing data across 14 cohort studies are well documented (79) and have been surmounted in recent decades by several strategies, including meta-analysis of results generated by cohort-specific data coordinating centers and analyses of pooled cohort data assembled at a single center or in an NHLBI repository. To enhance accessibility, increase security, and expedite high-priority analyses, C4R adapted and extended prior data management models by building a cloud-based enclave for assembly, harmonization, and analysis of pooled data.

C4R has certain limitations. Although it includes a diverse, nationwide sample, it was not constructed to be representative of the US population; to some extent, this may be addressed analytically through the use of weighting approaches. Participation in C4R is voluntary, which may lead to selection biases. Nonetheless, we do not observe major differences in the sociodemographic characteristics of eligible participants compared with participants completing the wave 1 questionnaire, and the preponderance of women in the sample may be explained to some extent by the advanced age of contributing cohorts. Data on acute COVID-19 are sparse in C4R compared with certain electronic health record resources, but, conversely, data on pre-COVID biomarkers, multiomics, physiology, organ structure, and symptomatology may be richer in C4R and less subject to diagnostic or referral biases. Because of antibody-waning, only a subset of self-reported COVID-19 cases in C4R will be validated by C4R serology, and negative serology will not be suitable to rule out infection. Nonetheless, serological analysis provides opportunities to reclassify some participants with subclinical infection and to examine antibody responses to vaccination and the impact on subsequent "breakthrough" infection risk. Adjudication of events consists primarily of validating more severe illness, although many cohort investigators are ascertaining positive test results among nonhospitalized participants to confirm history of infection.

C4R provides opportunities for future studies using a range of epidemiologic study designs (Figure 5). For example, nested within C4R, longitudinal cohort studies of COVID-affected and unaffected participants could repeat particular measurements (e.g., echocardiography, lung imaging, neurocognitive assessment) to reliably define the

consequences of COVID-19 infection. Ongoing high-quality events follow-up allows assessment of long-term clinical outcomes following COVID-19 and the pandemic period. Extensive biobanks maintained by the cohorts could support measurement of prior viral infections, immune phenotypes, metabotypes, -omics, and other pre-COVID characteristics that may be risk determinants or modifiers for COVID-19 susceptibility and vaccine effectiveness. The fact that the cohort investigators continue to follow their participants provides a dynamic resource for studying emerging questions in COVID-19 epidemiology, including but not limited to viral variants and vaccination. Additionally, C4R provides a model for cross-cohort collaboration and active data-sharing that will promote consortium-based epidemiologic work on biological, social, and epidemiologic questions beyond the COVID-19 pandemic, in alignment with recommendations for strategic transformation of population studies (80).

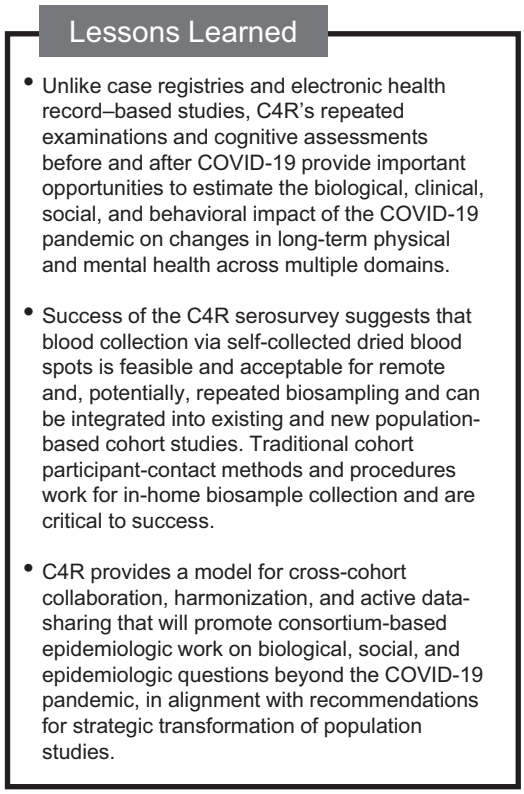

Figure 5. Lessons learned from the Collaborative Cohort of Cohorts for COVID-19 Research (C4R) Study. COVID-19, coronavirus disease 2019.

ACKNOWLEDGMENTS

Author affiliations: Division of General Medicine, Department of Medicine, Columbia University Medical Center, New York, New York, United States (Elizabeth C. Oelsner, Akshaya Krishnaswamy, Pallavi P. Balte, R. Graham Barr, Priya Palta, Yiyi Zhang); Northwestern Center for Epidemiology and Population Health, Department of Preventive Medicine, Feinberg School of Medicine, Northwestern University, Chicago, Illinois, United States (Norrina Bai Allen, Namratha R. Kandula); Center for American Indian Health Research, University of Oklahoma Health Sciences Center, Oklahoma City, Oklahoma, United States (Tauqeer Ali, Kimberly Malloy, Ying Zhang); Department of Biostatistics and Epidemiology, Hudson College of Public Health, University of Oklahoma, Oklahoma City, Oklahoma, United States (Tauqeer Ali, Kimberly Malloy, Ying Zhang); Department of Medicine, University of Mississippi Medical Center, Jackson, Mississippi, United States (Pramod Anugu, Adolfo Correa, Mario Sims, Yuan-I (Nancy) Min); Department of Population Health Science, John D. Bower School of Population Health, University of Mississippi, Jackson, Mississippi, United States (Pramod Anugu, Adolfo Correa, Mario Sims); Columbia Data Coordinating Center, Mailman School of Public Health, Columbia University, New York, New York, United States (Howard F. Andrews); New York State Psychiatric Institute, New York, New York, United States (Howard F. Andrews); Department of Biostatistics, Vagelos College of Physicians and Surgeons, Columbia University, New York, New York, United States (Howard F. Andrews); Columbia University Medical Center, New York, New York, United States (Howard F. Andrews); Division of Cardiovascular Sciences, National Heart, Lung, and Blood Institute, Bethesda, Maryland, United States (Komal Arora, Sean Coady); Division of Pulmonary Sciences and Critical Care Medicine, Department of Medicine, School of Medicine, University of Colorado, Aurora, Colorado, United States (Alyssa Asaro, Joyce S. Lee, David A. Schwartz); Department of Epidemiology and Prevention, Department of Medicine, School of Medicine, Wake Forest University, Winston-Salem, North Carolina, United States (Alain G. Bertoni); Division of Pulmonary, Allergy and Critical Care Medicine, Department of Medicine, School of Medicine, University of Pittsburgh, Pittsburgh, Pennsylvania, United States (Jessica Bon); Laboratory for Clinical Biochemistry Research, Department of Pathology and Laboratory Medicine, Larner College of Medicine, University of Vermont, Burlington, Vermont, United States (Rebekah Boyle, Elaine Cornell, Russell P. Tracy); Department of Obstetrics, Gynecology and Reproductive Sciences, School of Medicine, University of California, San Francisco, San Francisco, California, United States (Arunee A. Chang, Michael Schembri); Department of Medicine, School of Medicine, University of California, San Francisco, San Francisco, California, United States (Alka M. Kanaya, Prescott G. Woodruff); Division of General Internal Medicine, Department of Medicine, Feinberg School of Medicine, Northwestern University, Chicago, Illinois, United States (Namratha R. Kandula); National Jewish Health, Denver, Colorado, United States (Grace Chen); Population Health Program, Texas Biomedical Research Institute, San Antonio, Texas, United States (Shelley A. Cole); Department of Epidemiology, Bloomberg School of Public Health, Johns Hopkins University, Baltimore,

Maryland, United States (Josef Coresh); Collaborative Studies Coordinating Center, Department of Biostatistics, Gillings School of Global Public Health, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, United States (David Couper, Kimberly Ring); Vermont Center for Cardiovascular and Brain Health, Department of Medicine, Larner College of Medicine, University of Vermont, Burlington, Vermont, United States (Mary Cushman, Debora Kamin Mukaz); Department of Epidemiology, Mailman School of Public Health, Columbia University, New York, New York, United States (Ryan T. Demmer, Mitchell S. V. Elkind); Division of Epidemiology and Community Health, School of Public Health, University of Minnesota, Minneapolis, Minnesota, United States (Ryan T. Demmer, Aaron R. Folsom, David R. Jacobs, Jr.); Department of Epidemiology, School of Public Health, University of Washington, Seattle, Washington, United States (Amanda M. Fretts, Robert C. Kaplan, Bruce M. Psaty); Department of Epidemiology, School of Public Health, University of Alabama at Birmingham, Birmingham, Alabama, United States (Kelley P. Gabriel, Virginia J. Howard, Cora E. Lewis, Emily B. Levitan); South Bay Latino Research Center, Department of Psychology, College of Sciences, San Diego State University, San Diego, California, United States (Linda C. Gallo, Gregory A. Talavera); Department of Neurology, Vagelos College of Physicians and Surgeons, Columbia University, New York, New York, United States (Mitchell S. V. Elkind, Jose Gutierrez, Jennifer J. Manly); Division of Pulmonary and Critical Care, Department of Medicine, University of Michigan, Ann Arbor, Michigan, United States (MeiLan K. Han); Department of Pathology and Laboratory Medicine, School of Medicine, Boston University, Boston, Massachusetts, United States (Joel M. Henderson); Department of Epidemiology and Population Health, Albert Einstein College of Medicine, New York, New York, United States (Carmen R. Isasi, Robert C. Kaplan); Department of Biostatistics, School of Public Health, University of Alabama at Birmingham, Birmingham, Alabama, United States (Suzanne E. Judd); Department of Epidemiology, School of Public Health, University of Colorado, Aurora, Colorado, United States (Gregory L. Kinney); Department of Epidemiology, College of Public Health, University of Kentucky, Lexington, Kentucky, United States (Anna M. Kucharska-Newton); Department of Epidemiology, Gillings School of Global Public Health, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, United States (Anna Kucharska-Newton); Division of General Medicine, Department of Medicine, University of Michigan Medical School, Ann Arbor, Michigan, United States (Deborah A. Levine); Department of Pulmonary and Critical Care Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts, United States (Bruce D. Levy, Matthew R. Moll); Division of Pulmonary, Critical Care and Sleep Medicine, Department of Medicine, National Jewish Health, Denver, Colorado, United States (Barry J. Make); National Institute of Neurological Disorders and Stroke, Bethesda, Maryland, United States (Carolina Mendoza-Puccini); Department of Nutrition, Gillings School of Global Public Health, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, United States (Katie A. Meyer); Channing Division of Network Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts, United States (Matthew R. Moll); Division of Pulmonary, Critical Care, Allergy and Immunologic Diseases, Department of Medicine, School of Medicine, Wake Forest University, Winston-Salem, North Carolina, United States (Wendy C. Moore, Victor E. Ortega); Division of Biostatistics and Bioinformatics, Department of Public Health Sciences, College of Medicine, Pennsylvania State University, State College, Pennsylvania, United States (David Mauger); Bloodborne Viruses Laboratory, Wadsworth Center, New York State Department of Health, Albany, New York, United States (Monica M. Parker); Department of Medicine, Division of Allergy and Immunology, Boston Children's Hospital and Harvard Medical School, Boston, Massachusetts, United States (Wanda Phipatanakul); Division of Cardiology, Department of Medicine, School of Medicine, Johns Hopkins University, Baltimore, Maryland, United States (Wendy S. Post); Division of Lung Diseases, National Heart, Lung, and Blood Institute, Bethesda, Maryland, United States (Lisa Postow); Department of Health Systems and Population Health, School of Public Health, University of Washington, Seattle, Washington, United States (Bruce M. Psaty); Division of General Medicine, Department of Medicine, School of Medicine, University of Washington, Seattle, Washington, United States (Bruce M. Psaty); Division of Rheumatology, Department of Medicine, National Jewish Health, Denver, Colorado, United States (Elizabeth A. Regan); Division of Intramural Research, National Heart, Lung, and Blood Institute, Bethesda, Maryland, United States (Veronique L. Roger); Institute for Translational Genomics and Population Sciences, Department of Pediatrics, Lundquist Institute for Biomedical Innovation at Harbor-UCLA Medical Center, Torrance, California, United States (Jerome I. Rotter); Department of Neurology, Miller School of Medicine, University of Miami, Miami, Florida, United States (Tatjana Rundek, Ralph L. Sacco); Evelyn F. McKnight Brain Institute, University of Miami, Miami, Florida, United States (Tatjana Rundek, Ralph L. Sacco); Glenn Biggs Institute for Alzheimer's and Neurodegenerative Diseases, Graduate School of Biomedical Sciences, University of Texas, San Antonio, Texas, United States (Sudha Seshadri); University of Texas Health Science Center at San Antonio, San Antonio, Texas, United States (Sudha Seshadri); Division of Preventive Medicine, School of Medicine, University of Alabama at Birmingham, Birmingham, Alabama, United States (James M. Shikany); Department of Biostatistics, School of Public Health, University of Washington, Seattle, Washington, United States (Karen Hinckley Stukovsky); MedStar Health Research Institute, Hyattsville, Maryland, United States (Jason G. Umans); Georgetown-Howard Universities Center for Clinical and Translational Science, Washington, DC, United States (Jason G. Umans); Department of Medicine, School of Medicine, Boston University, Boston,

Massachusetts, United States (Ramachandran S. Vasan, Vanessa Xanthakis); Department of Epidemiology, School of Public Health, Boston University, Boston, Massachusetts, United States (Ramachandran S. Vasan); Division of Cardiology, Department of Medicine, Ronald Reagan UCLA Medical Center, Los Angeles, California, United States (Karol E. Watson); Department of Environmental and Occupational Health, Graduate School of Public Heath, University of Pittsburgh, Pittsburgh, Pennsylvania, United States (Sally E. Wenzel); School of Nursing, University of Mississippi, Jackson, Mississippi, United States (Karen Winters); and Department of Biostatistics, School of Public Health, Boston University, Boston, Massachusetts, United States (Vanessa Xanthakis).

The Collaborative Cohort of Cohorts for COVID-19 Research (C4R) Study is supported by National Heart, Lung, and Blood Institute (NHLBI)–Collaborating Network of Networks for Evaluating COVID-19 and Therapeutic Strategies (CONNECTS) grant OT2HL156812, with cofunding from the National Institute of Neurological Disorders and Stroke (NINDS) and the National Institute on Aging (NIA). The Atherosclerosis Risk in Communities Study has been funded in whole or in part by the NHLBI, National Institutes of Health (NIH), US Department of Health and Human Services, under contracts 75N92022D00001, 75N92022D00002, 75N92022D00003, 75N92022D00004, and 75N92022D00005. Neurocognitive data are collected under grants U01 2U01HL096812, 2U01HL096814, 2U01HL096899, 2U01HL096902, and 2U01HL096917 from the NHLBI, the NINDS, the NIA, and the National Institute on Deafness and Other Communication Disorders. Ancillary studies funded additional data elements. The Blood Pressure and Cognition Study is supported by the NINDS (grant R01 NS102715). The Coronary Artery Risk Development in Young Adults (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 HHSN268201800004I). The Genetic Epidemiology of COPD (COPDGene) Study was supported by awards U01 HL089897 and U01 HL089856 from the NHLBI. COPDGene is also supported by the COPD Foundation through contributions made to an industry advisory board comprised of AstraZeneca AB (Cambridge, United Kingdom), Boehringer-Ingelheim (Ingelheim am Rhein, Germany), Genentech, Inc. (South San Francisco, California), GlaxoSmithKline plc (London, United Kingdom), Novartis International AG (Basel, Switzerland), Pfizer, Inc. (New York, New York), Siemens AG (Berlin, Germany), and Sunovion Pharmaceuticals Inc. (Marlborough, Massachusetts). The Framingham Heart Study has received support from the NHLBI (grant N01-HC-25195, contract HHSN268201500001I, and grant 75N92019D00031). The Hispanic Community Health Study/Study of Latinos (HCHS/SOL) is a collaborative study supported by contracts between the NHLBI and the University of North Carolina (contract HHSN268201300001I/N01-HC-65233), the University of Miami (contract HHSN268201300004I/N01-HC-65234), Albert Einstein College of Medicine (contract HHSN268201300002I/N01-HC-65235), the University of Illinois at Chicago (contract HHSN268201300003I/N01-HC-65236 (Northwestern University)), and San Diego State University (contract HHSN268201300005I/N01-HC-65237). The following institutes/centers/offices have contributed to the HCHS/SOL through a transfer of funds to the NHLBI: the National Institute on Minority Health and Health Disparities, the National Institute on Deafness and Other Communication Disorders, the National Institute of Dental and Craniofacial Research, the National Institute of Diabetes and Digestive and Kidney Diseases, the NINDS, and the NIH Office of Dietary Supplements. The Jackson Heart Study is supported by and conducted in collaboration with Jackson State University (contract HHSN268201800013I), Tougaloo College (contract HHSN268201800014I), the Mississippi State Department of Health (contract HHSN268201800015I), the University of Mississippi Medical Center (contracts HHSN268201800010I, HHSN268201800011I, and HHSN268201800012I), the NHLBI, and the National Institute on Minority Health and Health Disparities. The Mediators of Atherosclerosis in South Asians Living in America (MASALA) Study was supported by grant R01HL093009 from the NHLBI, the National Center for Research Resources, and the National Center for Advancing Translational Sciences, NIH, through University of California, San Francisco–Clinical and Translational Science Institute grant UL1RR024131. The Multi-Ethnic Study of Atherosclerosis (MESA) and the MESA SNP Health Association Resource (SHARe) are conducted and supported by the NHLBI in collaboration with the MESA investigators. Support for MESA is provided by grants and contracts 75N92020D00001, HHSN268201500003I, N01-HC-95159, 75N92020D00005, N01-HC-95160, 75N92020D00002, N01-HC-95161, 75N92020D00003, N01-HC-95162, 75N92020D00006, N01-HC-95163, 75N92020D00004, N01-HC-95164, 75N92020D00007, N01-HC-95165, N01-HC-95166, N01-HC-95167, N01-HC-95168, N01-HC-95169, R01-HL077612, R01-HL093081, R01-HL130506, R01-HL127028, R01-HL127659, R01-HL098433, R01-HL101250, and R01-HL135009 from the NHLBI; grant R01-AG058969 from the NIA; and grants UL1-TR-000040, UL1-TR-001079, and UL1-TR-001420 from the National Center for Advancing Translational Sciences. Funding for SHARe genotyping was provided by NHLBI contract N02-HL-64278. This publication was developed under Science to Achieve Results (STAR) research assistance agreements RD831697 (MESA Air) and RD-83830001 (MESA Air Next Stage), awarded by the Environmental Protection Agency. Whole genome sequencing for the Trans-Omics in Precision Medicine (TOPMed) Program was supported by the NHLBI. Whole genome sequencing for the MESA component of the TOPMed Study (Database of Genotypes

and Phenotypes accession no. phs001416.v1.p1) was performed at the Broad Institute of MIT and Harvard (grant 3U54HG003067-13S1). Centralized read mapping and genotype calling, along with variant quality metrics and filtering, were provided by the TOPMed Informatics Research Center (grant 3R01HL-117626-02S1 and contract HHSN268201800002I) (Broad RNA Seq, Proteomics HHSN268201600034I, UW RNA Seq HHSN268201600032I, USC DNA Methylation HHSN268201600034I, Broad Metabolomics HHSN268201600038I). Phenotype harmonization, data management, sample-identity quality control, and general study coordination were provided by the TOPMed Data Coordinating Center (grants 3R01HL-120393 and U01HL-120393 and contract HHSN268180001I). The provision of genotyping data was supported in part by the National Center for Advancing Translational Sciences, Clinical and Translational Science Institute grant UL1TR001881, and National Institute of Diabetes and Digestive and Kidney Diseases Diabetes Research Center grant DK063491 to the Southern California Diabetes Endocrinology Research Center. The NHLBI Pooled Cohorts Study was supported by grants R21HL153700, K23HL130627, R21HL129924, and R21HL121457 from the NIH/NHLBI. The Northern Manhattan Study was supported by grants R01 NS29993 and R01 NS48134 from the NINDS and grant R01 AG066162 from the NIA. The Prevent Pulmonary Fibrosis cohort study was established in 2000 and has been supported by NIH awards Z01-ES101947, R01-HL095393, RC2-HL1011715, R21/33-HL120770, R01-HL097163, Z01-HL134585, UH2/3-HL123442, P01-HL092870, UG3/UH3-HL151865, and DoD W81XWH-17-1-0597. The Reasons for Geographic and Racial Differences in Stroke (REGARDS) Study is supported by cooperative agreement U01 NS041588, cofunded by the NINDS and the NIA. Research by the principal and co–principal investigators of the Severe Asthma Research Program was funded by the NIH/NHLBI (grants U10 HL109164, U10 HL109257, U10 HL109146, U10 HL109172, U10 HL109250, U10 HL109168, U10 HL109152, and U10 HL109086). Additional support was provided through industry partnerships with the following companies: AstraZeneca, Boehringer-Ingelheim, Genentech, GlaxoSmithKline, MedImmune, Inc. (Gaithersburg, Maryland), Novartis, Regeneron Pharmaceuticals, Inc. (Tarrytown, New York), Sanofi S.A. (Paris, France), and Teva Pharmaceuticals USA (North Wales, Pennsylvania). Spirometers used in Severe Asthma Research Program III were provided by nSpire Health, Inc. (Longmont, Colorado). The Subpopulations and Intermediate Outcome Measures in COPD Study (SPIROMICS) has been funded by contracts with the NIH/NHLBI (contracts HHSN268200900013C, HHSN268200900014C, HHSN268200900015C, HHSN268200900016C, HHSN268200900017C, HHSN268200900018C, HHSN268200900019C, and HHSN268200900020C) and grants from the NIH/NHLBI (grants U01 HL137880 and U24 HL141762) and supplemented through contributions made to the Foundation for the NIH and the COPD Foundation by AstraZeneca, MedImmune, Bayer Corporation (Whippany, New Jersey), Bellerophon Therapeutics (Warren, New Jersey), Boehringer-Ingelheim, Chiesi Farmaceutici S.p.A. (Parma, Italia), the Forest Research Institute, Inc. (Jersey City, New Jersey), GlaxoSmithKline, Grifols Therapeutics, Inc. (Research Triangle Park, North Carolina), Ikaria, Inc. (Hampton, New Jersey), Novartis, Nycomed Pharma GmbH (Zurich, Switzerland), ProterixBio, Inc. (Billerica, Massachusetts), Regeneron, Sanofi, Sunovion, Takeda Pharmaceutical Company (Tokyo, Japan), Theravance Biopharma, Inc. (South San Francisco, California), and Mylan N.V. (White Sulphur Springs, West Virginia). The Strong Heart Study has been funded in whole or in part by the NHLBI (contracts 75N92019D00027, 75N92019D00028, 75N92019D00029, and 75N92019D00030). The Strong Heart Study was previously supported by research grants R01HL109315, R01HL109301, R01HL109284, R01HL109282, and R01HL109319 and cooperative agreements U01HL41642, U01HL41652, U01HL41654, U01HL65520, and U01HL65521.

Access to pooled C4R data is regulated by the C4R Study's publications and presentations policy, which is available on the C4R website (https://c4r-nih.org). Data are made available for analyses in the C4R Analysis Commons for investigators with manuscript proposals approved by the C4R publications and cohort coordinating committees, as well as by each cohort included in a given proposal. Once harmonization and related quality control have been completed, C4R common data elements will be transferred as a limited data set for public access on BioData Catalyst, in accord with cohort-specific consents and commitments.

We thank the participants in each cohort study for their dedication. We thank the staff and participants of the Jackson Heart Study. We also thank the MESA investigators and the staff and participants of the REGARDS Study for their valuable contributions. In addition, we acknowledge the dedication of the Framingham Heart Study participants, without whom this research would not have been possible.

Members of the C4R study group: Tauqueer Ali, Norrina Bai Allen, Howard F. Andrews, Pramod Anugu, Komal Arora, Alex Arynchyn, Alyssa Asaro, Pallavi P. Balte, Michael Bancks, R. Graham Barr, Lori Bateman, Alain G. Bertoni, Alexander G. Bick, Eugene R. Bleecker, David A. Bluemke, Rebekah Boyle, Matthew Budoff, Jonathan Cardwell, Jeffrey Carr, Jackie Causey, Arunee A. Chang, Grace Chen, Sean Coady, Shelley A. Cole, Josef Coresh, Elaine Cornell, Adolfo Correa, David Couper, Emily Culbertson, Jeffrey L. Curtis, Mary Cushman, Charles S. DeCarli, Ryan T. Demmer, Richard B. Devereux, Nicholas DiPersio, Brett A. Dolezal, Margaret Doyle, Mitchell S. V. Elkind, Sean B. Fain, Matthew Feinstein, James Floyd, Christine Freeman, Amanda M. Fretts, Kelley P. Gabriel, Linda C. Gallo, Sina A. Gharib, Caroline Goode, Parag Goyal, Michael Griswold, MeiLan K. Han, David Hanna, Annette T. Hastie, Susan R. Heckbert, Joel M. Henderson, Karen D. Hinckley Stukovsky, Eric A. Hoffman, Udo Hoffman, Mady Hornig, Virginia J. Howard, Carmen R. Isasi, David R. Jacobs, Jr., Cashell Jaquis, Suzanne E. Judd,

Debora Kamin Mukaz, Alka M. Kanaya, Namratha R. Kandula, Robert C. Kaplan, Sadiya S. Khan, Jonathan King, Gregory L. Kinney, Akshaya Krishnaswamy, Anna Kucharska-Newton, M. Kumi Smith, Andrew F. Laine, Joyce S. Lee, Rozenn Lemaitre, Bonnie E. Levin, Deborah A. Levine, Emily B. Levitan, Bruce D. Levy, Joao A. C. Lima, David A. Lynch, Barry J. Make, Kimberly Malloy, Jennifer J. Manly, David Mauger, Katy Melius, Carolina Mendoza-Puccini, Sharon S. Merkin, Katie A. Meyer, Deborah Meyers, Yuan-I (Nancy) Min, Melissa Minotti, Nathalie Moise, Wendy C. Moore, Claudia Moy, Karen Mutalik, Ilya M. Nasrallah, Cheryl Nelson, Lauren Nelson, Patricia Noel, Anna S. Nordvig, George T. O'Connor, Michelle Odden, Elizabeth C. Oelsner, Marcia O'Leary, Wanda O'Neal, Victor E. Ortega, Priya Palta, Nathalie Pamir, George Papanicolaou, Brenda R. Phillips, Wanda Phipatanakul, Timothy B. Plante, Yashashwi Pokharel, Wendy S. Post, Lisa Postow, Bruce M. Psaty, Shaun M. Purcell, Laura Raffield, Ramachandran S. Vasan, Susan Redline, Elizabeth A. Regan, Alex Reiner, Carlos Rodriguez, Veronique L. Roger, Jerome I. Rotter, Ralph L. Sacco, Monika M. Safford, Michael Schembri, David A. Schwartz, Sudha Seshadri, Amil Shah, Sanjiv J. Shah, James Shikany, Mario Sims, Sylvia Smoller, Elsayed Soliman, Scott Solomon, Daniela Sotres-Alvarez, Meg Stewart, Kevin Strobino, Patrice Sutherland, Bonnie Swenor, Gregory A. Talavera, James G. Terry, Russell P. Tracy, Janis Tse, Jason G. Umans, Louis J. Vuga, Molly V. Wagster, Henry Wang, George R. Washko, Sally E. Wenzel, Cynthia West, Carla Wilson, Prescott G. Woodruff, Jackie Wright, Vanessa Xanthakis, Ya Yuan, Neil A. Zakai, Ying Zhang, and Yiyi Zhang.

A full list of participating REGARDS investigators and institutions can be found at https://www.uab.edu/soph/regardsstudy/.

A preliminary version of this article was published in *medRxiv* (81).

The views expressed in this article are those of the authors and do not necessarily represent the views of the NHLBI, the NIH, or the US Department of Health and Human Services; the NINDS; or the NIA. Representatives of the NINDS were involved in the review of the manuscript but were not directly involved in the collection, management, analysis, or interpretation of the data. This article has not been formally reviewed by the Environmental Protection Agency. The views expressed in this document are solely those of the authors, and the Environmental Protection Agency does not endorse any products or commercial services mentioned in this publication.

M. S. V. Elkind receives royalties from UpToDate, Inc. (Waltham, Massachusetts) for a book chapter on neurological complications of COVID-19 and receives study drug in kind from the Bristol-Myers Squibb–Pfizer Alliance for Eliquis (New York, New York) and ancillary funding from Roche, Inc. (New York, New York), both for an NIH-funded trial of stroke prevention. M. K. Han reports consulting for Boehringer Ingelheim, GlaxoSmithKline (GSK), AstraZeneca, Merck & Co., Inc. (Kenilworth, New Jersey), Mylan N.V., Verona Pharma plc (Cardiff, United Kingdom), Teva Pharmaceuticals USA (North Wales, Pennsylvania), Cipla Ltd. (Mumbai, India), Chiesi Farmaceutici, and Sanofi. M. K. Han also reports receiving research support from Novartis and Sunovion. E. B. Levitan has received research funding from AMGen, Inc. (Thousand Oaks, California) and has received consulting fees for a research project sponsored by Novartis. W. Phipatanakul has received funding and trial medication support from AstraZeneca, GSK, Merck, Genentech, Novartis, Regeneron, and Sanofi for asthma studies and has received consulting fees from GSK, Genentech, Novartis, Sanofi, and Regeneron for asthma therapeutics. B. M. Psaty serves on the Steering Committee of the Yale Open Data Access Project, funded by Johnson & Johnson (New Brunswick, New Jersey). D. A. Schwartz is the founder and chief scientific officer of Eleven P15 Inc. (Denver, Colorado), a company dedicated to the diagnosis, prevention, and treatment of early presentations of pulmonary fibrosis. J. S. Lee serves as a consultant for Eleven P15. S. E. Wenzel receives funding for consulting and clinical trials from AstraZeneca, GSK, Sanofi-Genzyme (Cambridge, Massachusetts), Novartis, and Knopp Biosciences LLC (Pittsburgh, Pennsylvania); she also receives research support from Pieris Pharmaceuticals GmbH (Hallbergmoos, Germany) and Regeneron. P. G. Woodruff has received a research grant from Genentech and is a consultant for Sanofi, Regeneron, Clarus Ventures LLC (Cambridge, Massachusetts), 23andMe, Inc. (Sunnyvale, California), Theravance Biopharma, Inc. (South San Francisco, California), AstraZeneca, Glenmark Pharmaceuticals Ltd. (Mumbai, India), AMGen, and NGM Biopharmaceuticals Inc. (South San Francisco, California) outside the scope of this work. None of the other authors have any potential conflicts of interest to declare.

REFERENCES

1. Kissler SM, Tedijanto C, Goldstein E, et al. Projecting the transmission dynamics of SARS-CoV-2 through the postpandemic period. *Science.* 2020;368(6493):860–868.
2. Woolf SH, Chapman DA, Lee JH. COVID-19 as the leading cause of death in the United States. *JAMA.* 2021;325(2):123–124.
3. Andrasfay T, Goldman N. Reductions in 2020 US life expectancy due to COVID-19 and the disproportionate impact on the black and Latino populations. *Proc Natl Acad Sci U S A.* 2021;118(5):e2014746118.
4. Huang C, Huang L, Wang Y, et al. 6-month consequences of COVID-19 in patients discharged from hospital: a cohort study. *Lancet.* 2021;397(10270):220–232.
5. The ARIC Investigators. The Atherosclerosis Risk in Communities (ARIC) Study: design and objectives. *Am J Epidemiol.* 1989;129(4):687–702.
6. Friedman GD, Cutter GR, Donahue RP, et al. CARDIA: study design, recruitment, and some characteristics of the examined subjects. *J Clin Epidemiol.* 1988;41(11):1105–1116.
7. Tsao CW, Vasan RS. Cohort profile: the Framingham Heart Study (FHS): overview of milestones in cardiovascular epidemiology. *Int J Epidemiol.* 2015;44(6):1800–1813.

8. Daviglus ML, Talavera GA, Avilés-Santa ML, et al. Prevalence of major cardiovascular risk factors and cardiovascular diseases among Hispanic/Latino individuals of diverse backgrounds in the United States. *JAMA.* 2012; 308(17):1775–1784.
9. Lavange LM, Kalsbeek WD, Sorlie PD, et al. Sample design and cohort selection in the Hispanic Community Health Study/Study of Latinos. *Ann Epidemiol.* 2010;20(8): 642–649.
10. Sorlie PD, Avilés-Santa LM, Wassertheil-Smoller S, et al. Design and implementation of the Hispanic Community Health Study/Study of Latinos. *Ann Epidemiol.* 2010;20(8): 629–641.
11. Carpenter MA, Crow R, Steffes M, et al. Laboratory, reading center, and coordinating center data management methods in the Jackson Heart Study. *Am J Med Sci.* 2004;328(3): 131–144.
12. Keku E, Rosamond W, Taylor HA Jr, et al. Cardiovascular disease event classification in the Jackson Heart Study: methods and procedures. *Ethn Dis.* 2005;15(4 suppl 6): S6-62–S6-70.
13. Taylor HA Jr, Wilson JG, Jones DW, et al. Toward resolution of cardiovascular health disparities in African Americans: design and methods of the Jackson Heart Study. *Ethn Dis.* 2005;15(4 suppl 6):S6-4–S6-17.
14. 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.* 2019;3(2-3):97–104.
15. Kanaya AM, Kandula N, Herrington D, et al. Mediators of Atherosclerosis in South Asians Living in America (MASALA) Study: objectives, methods, and cohort description. *Clin Cardiol.* 2013;36(12):713–720.
16. Bild DE, Bluemke DA, Burke GL, et al. Multi-Ethnic Study of Atherosclerosis: objectives and design. *Am J Epidemiol.* 2002;156(9):871–881.
17. Lee ET, Welty TK, Fabsitz R, et al. The Strong Heart Study: a study of cardiovascular disease in American Indians: design and methods. *Am J Epidemiol.* 1990;132(6):1141–1155.
18. North KE, Howard BV, Welty TK, et al. Genetic and environmental contributions to cardiovascular disease risk in American Indians: the Strong Heart Family Study. *Am J Epidemiol.* 2003;157(4):303–314.
19. Regan EA, Hokanson JE, Murphy JR, et al. Genetic Epidemiology of COPD (COPDGene) Study design. *COPD.* 2010;7(1):32–43.
20. Couper D, LaVange LM, Han M, et al. Design of the Subpopulations and Intermediate Outcomes in COPD Study (SPIROMICS). *Thorax.* 2014;69(5):491–494.
21. Mathai SK, Humphries S, Kropski JA, et al. *MUC5B* variant is associated with visually and quantitatively detected preclinical pulmonary fibrosis. *Thorax*. 2019;74(12): 1131–1139.
22. Teague WG, Phillips BR, Fahy JV, et al. Baseline features of the Severe Asthma Research Program (SARP III) cohort: differences with age. *J Allergy Clin Immunol Pract.* 2018; 6(2):545–554.e4.
23. Sacco RL, Boden-Albala B, Gan R, et al. Stroke incidence among white, black, and Hispanic residents of an urban community: the Northern Manhattan Stroke Study. *Am J Epidemiol.* 1998;147(3):259–268.
24. Howard VJ, Cushman M, Pulley L, et al. The Reasons for Geographic and Racial Differences in Stroke Study: objectives and design. *Neuroepidemiology.* 2005;25(3): 135–143.
25. Psaty BM, O'Donnell CJ, Gudnason V, et al. Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium: design of prospective meta-analyses of genome-wide association studies from 5 cohorts. *Circ Cardiovasc Genet.* 2009;2(1):73–80.
26. 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.* 2018; 187(11):2265–2278.
27. CHS Coordinating Center, University of Washington. Cross-Cohort Collaboration Consortium. https://chs-nhlbi.org/node/6539. Published 2001. Updated May 10, 2021. Accessed October 13, 2021.
28. Levine DA, Gross AL, Briceño EM, et al. Association between blood pressure and later-life cognition among black and white individuals. *JAMA Neurol.* 2020;77(7):810–819.
29. Taliun D, Harris DN, Kessler MD, et al. Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. *Nature.* 2021;590(7845):290–299.
30. Harris PA, Taylor R, Minor BL, et al. The REDCap Consortium: building an international community of software platform partners. *J Biomed Inform.* 2019;95:103208.
31. Harris PA, Taylor R, Thielke R, et al. Research Electronic Data Capture (REDCap)—a metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform.* 2009;42(2): 377–381.
32. Flanagin A, Frey T, Christiansen SL, et al. The reporting of race and ethnicity in medical and science journals: comments invited. *JAMA.* 2021;325(11):1049–1052.
33. Oelsner EC. *MESA COVID-19 Questionnaire*. Bethesda, MD: National Library of Medicine; 2020. https://www.nlm.nih.gov/dr2/Oelsner_MESA_COVID-19_Questionnaire_FINAL_04102020.pdf. Accessed October 13, 2021.
34. Behavioral Risk Factor Surveillance System, Centers for Disease Control and Prevention. BRFSS questionnaires. https://www.cdc.gov/brfss/questionnaires/. Updated August 26, 2021. Accessed October 13, 2021.
35. Institute for Social Research, University of Michigan. Health and Retirement Study. Biennial interview questionnaires. https://hrs.isr.umich.edu/documentation/questionnaires. Accessed October 13, 2021.
36. Women's Interagency HIV Study (WIHS)/Multicenter AIDS Cohort Study (MACS). *Baseline COVID-19 Abbreviated Questionnaire (BLCOVID)*. Bethesda, MD: National Library of Medicine; 2020. https://www.nlm.nih.gov/dr2/MACS-WIHS_questionnaire_BLCOVID-040620.pdf. Published April 6, 2020. Accessed October 13, 2021.
37. Andresen EM, Malmgren JA, Carter WB, et al. Screening for depression in well older adults: evaluation of a short form of the CES-D (Center for Epidemiologic Studies Depression Scale). *Am J Prev Med.* 1994;10(2):77–84.
38. Cohen S, Kamarck T, Mermelstein R. A global measure of perceived stress. *J Health Soc Behav.* 1983;24(4):385–396.
39. Levine DW, Kripke DF, Kaplan RM, et al. Reliability and validity of the Women's Health Initiative Insomnia Rating Scale. *Psychol Assess.* 2003;15(2):137–148.
40. Pilkonis PA, Choi SW, Reise SP, et al. Item banks for measuring emotional distress from the Patient-Reported Outcomes Measurement Information System (PROMIS®): depression, anxiety, and anger. *Assessment.* 2011;18(3): 263–283.
41. RAND Corporation. Social Support Survey instrument. https://www.rand.org/health-care/surveys_tools/mos/social-support/survey-instrument.html. Accessed October 13, 2021.

42. Russell DW. UCLA Loneliness Scale (version 3): reliability, validity, and factor structure. *J Pers Assess.* 1996;66(1): 20–40.
43. Smith BW, Dalen J, Wiggins K, et al. The Brief Resilience Scale: assessing the ability to bounce back. *Int J Behav Med.* 2008;15(3):194–200.
44. Centers for Disease Control and Prevention. *ICD-10-CM Official Coding and Reporting Guidelines April 1, 2020 Through September 30, 2020*. Hyattsville, MD: National Center for Health Statistics; 2020. https://www.cdc.gov/nchs/data/icd/COVID-19-guidelines-final.pdf. Accessed October 13, 2021.
45. Wyatt SB, Diekelmann N, Henderson F, et al. A community-driven model of research participation: the Jackson Heart Study Participant Recruitment and Retention Study. *Ethn Dis*. 2003;13(4):438–455.
46. Oelsner EC, Loehr LR, Henderson AG, et al. Classifying chronic lower respiratory disease events in epidemiologic cohort studies. *Ann Am Thorac Soc.* 2016;13(7):1057–1066.
47. Zakai NA, McClure LA, Judd SE, et al. Racial and regional differences in venous thromboembolism in the United States in 3 cohorts. *Circulation.* 2014;129(14):1502–1509.
48. Centers for Disease Control and Prevention. Interim guidelines for COVID-19 antibody testing. https://www.cdc.gov/coronavirus/2019-ncov/lab/resources/antibody-tests-guidelines.html#anchor_1590264293982. Updated January 24, 2022. Accessed October 13, 2021.
49. Laing ED, Sterling SL, Richard SA, et al. Antigen-based multiplex strategies to discriminate SARS-CoV-2 natural and vaccine induced immunity from seasonal human coronavirus humoral responses [preprint]. *medRxiv.* 2021. (https://doi.org/10.1101/2021.02.10.21251518). Accessed October 13, 2021.
50. Balte PP, Chaves PHM, Couper DJ, et al. Association of nonobstructive chronic bronchitis with respiratory health outcomes in adults. *JAMA Intern Med.* 2020;180(5): 676–686.
51. Bhatt SP, Balte PP, Schwartz JE, et al. Discriminative accuracy of FEV_1:FVC thresholds for COPD-related hospitalization and mortality. *JAMA.* 2019;321(24): 2438–2447.
52. Bhatt SP, Schwartz JE, Oelsner EC. FEV_1:FVC thresholds for defining chronic obstructive pulmonary disease—reply. *JAMA.* 2019;322(16):1611–1612.
53. Cornelius TBP, Schwartz JE, Balte P, et al. A dyadic growth modeling approach for examining associations between weight gain and lung function decline. *Am J Epidemiol.* 2020; 189(10):1173–1184.
54. Oelsner EC, Balte PP, Bhatt SP, et al. Lung function decline in former smokers and low-intensity current smokers: a secondary data analysis of the NHLBI Pooled Cohorts Study. *Lancet Respir Med.* 2020;8(1):34–44.
55. Zhang Y, Vittinghoff E, Pletcher MJ, et al. Associations of blood pressure and cholesterol levels during young adulthood with later cardiovascular events. *J Am Coll Cardiol.* 2019; 74(3):330–341.
56. Oelsner EC, Balte PP, Grams ME, et al. Albuminuria, lung function decline, and risk of incident chronic obstructive pulmonary disease. The NHLBI Pooled Cohorts Study. *Am J Respir Crit Care Med.* 2019;199(3):321–332.
57. Stilp AM, Emery LS, Broome JG, et al. A system for phenotype harmonization in the National Heart, Lung, and Blood Institute Trans-Omics for Precision Medicine (TOPMed) Program. *Am J Epidemiol.* 2021;190(10): 1977–1992.
58. Griffith L, van den Heuvel E, Fortier I, et al. *Harmonization of Cognitive Measures in Individual Participant Data and Aggregate Data Meta-Analysis*. (AHRQ publication no. 13-EHC040-EF). Rockville, MD: Agency for Healthcare Research and Quality; 2013.
59. Griffith LE, van den Heuvel E, Fortier I, et al. Statistical approaches to harmonize data on cognitive measures in systematic reviews are rarely reported. *J Clin Epidemiol.* 2015;68(2):154–162.
60. Gross AL, Jones RN, Fong TG, et al. Calibration and validation of an innovative approach for estimating general cognitive performance. *Neuroepidemiology.* 2014;42(3): 144–153.
61. Gross AL, Sherva R, Mukherjee S, et al. Calibrating longitudinal cognition in Alzheimer's disease across diverse test batteries and datasets. *Neuroepidemiology.* 2014;43(3-4): 194–205.
62. Gross AL, Mungas DM, Crane PK, et al. Effects of education and race on cognitive decline: an integrative study of generalizability versus study-specific results. *Psychol Aging.* 2015;30(4):863–880.
63. Langa KM, Ryan LH, McCammon RJ, et al. The Health and Retirement Study Harmonized Cognitive Assessment Protocol Project: study design and methods. *Neuroepidemiology.* 2020;54(1):64–74.
64. Oelsner EC, Ortega VE, Smith BM, et al. A genetic risk score associated with chronic obstructive pulmonary disease susceptibility and lung structure on computed tomography. *Am J Respir Crit Care Med.* 2019;200(6):721–731.
65. Smith BM, Kirby M, Hoffman EA, et al. Association of dysanapsis with chronic obstructive pulmonary disease among older adults. *JAMA.* 2020;323(22):2268–2280.
66. Smith BM, Hoffman EA, Rabinowitz D, et al. Comparison of spatially matched airways reveals thinner airway walls in COPD. The Multi-Ethnic Study of Atherosclerosis (MESA) COPD Study and the Subpopulations and Intermediate Outcomes in COPD Study (SPIROMICS). *Thorax.* 2014; 69(11):987–996.
67. Hame Y, Angelini E, Hoffman E, et al. Adaptive quantification and longitudinal analysis of pulmonary emphysema with a hidden Markov measure field model. *IEEE Trans Med Imaging.* 2014;33(7):1527–1540.
68. Wang M, Aaron CP, Madrigano J, et al. Association between long-term exposure to ambient air pollution and change in quantitatively assessed emphysema and lung function. *JAMA.* 2019;322(6):546–556.
69. Yang J, Angelini ED, Balte PP, et al. Unsupervised discovery of spatially-informed lung texture patterns for pulmonary emphysema: the MESA COPD Study. In: Descoteaux M, Maier-Hein L, Franz A, et al., eds. *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2017*. (Lecture Notes in Computer Science, vol. 10433). New York, NY: Springer Publishing Company; 2017:116–124.
70. Yang J, Angelini ED, Balte PP, et al. Emphysema quantification on cardiac CT scans using hidden Markov measure field model: the MESA Lung Study. In: Ourselin S, Joskowicz L, Sabuncu MR, et al., eds. *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2016*. (Lecture Notes in Computer Science, vol. 9900). New York, NY: Springer Publishing Company; 2016: 624–631.
71. Yang J, Vetterli T, Balte PP, et al. Unsupervised domain adaption with adversarial learning (UDAA) for emphysema subtyping on cardiac CT scans: the MESA Study. In: *2019 IEEE 16th International Symposium on Biomedical Imaging*

(ISBI 2019). Piscataway, NJ: Institute of Electrical and Electronics Engineers; 2019:289–293.
72. Ravi D, Wong C, Deligianni F, et al. Deep learning for health informatics. *IEEE J Biomed Health Inform.* 2017;21(1):4–21.
73. Pham T, Tran T, Phung D, et al. DeepCare: a deep dynamic memory model for predictive medicine. In: Bailey J, Khan L, Washio T, et al., eds. *Advances in Knowledge Discovery and Data Mining: 20th Pacific-Asia Conference, PAKDD 2016. Auckland, New Zealand, April 19–22, 2016. Proceedings, Part II*. (Lecture Notes in Artificial Intelligence, vol. 9652). Berlin, Germany: Springer-Verlag; 2016:30–41.
74. Miotto R, Li L, Kidd BA, et al. Deep patient: an unsupervised representation to predict the future of patients from the electronic health records. *Sci Rep*. 2016;6(1):26094.
75. He K, Zhang X, Ren S, et al. Deep residual learning for image recognition [preprint]. *arXiv*. 2015. (https://doi.org/10.48550/arXiv.1512.03385). Accessed October 13, 2021.
76. Brody JA, Morrison AC, Bis JC, et al. Analysis commons, a team approach to discovery in a big-data environment for genetic epidemiology. *Nat Genet*. 2017;49(11):1560–1563.
77. Tribal Health Research Office, National Institutes of Health. *NIH Tribal Consultation Report: NIH Draft Policy for Data Management and Sharing*. Bethesda, MD: National Institutes of Health; 2020.
78. National Heart, Lung, and Blood Institute. The NHLBI BioData Catalyst. https://doi.org/10.5281/zenodo.3822858. Published August 31, 2020. Accessed April 12, 2022.
79. Psaty BM, Rich SS, Boerwinkle E. Innovation in genomic data sharing at the NIH. *N Engl J Med.* 2019;380(23):2192–2195.
80. Roger VL, Boerwinkle E, Crapo JD, et al. Strategic transformation of population studies: recommendations of the working group on epidemiology and population sciences from the National Heart, Lung, and Blood Advisory Council and Board of External Experts. *Am J Epidemiol.* 2015;181(6):363–368.
81. Oelsner EC, Allen NB, Ali T, et al. Collaborative Cohort of Cohorts for COVID-19 Research (C4R) Study: study design [preprint]. *medRxiv*. (https://www.medrxiv.org/content/10.1101/2021.03.19.21253986v1). Accessed October 13, 2021.