COVER PAGE

Title

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

Authorship

Elizabeth C Oelsner
Norrina Bai Allen
Tauqeer Ali
Pramod Anugu
Howard Andrews
Alyssa Asaro
Pallavi P Balte
R Graham Barr
Alain G Bertoni
Jessica Bon
Rebekah Boyle
Arunee A Chang
Grace Chen
Shelley A Cole
Josef Coresh
Elaine Cornell

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.
Abstract word count (limit AJE: 200)

218

Word count (limit AJE: 4000)

3317
Footnotes Page

Abbreviations

ARIC Atherosclerosis Risk in Communities
BCL Biorepository and Central Laboratory
C4R Collaborative Cohort of Cohorts for COVID-19 Research
CARDIA Coronary Artery Risk Development in Young Adults
CONNECTS Collaborating Network of Networks for Evaluating COVID-19 and Therapeutic Strategies
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
PASC Post-Acute Sequelae of SARS-CoV-2 Infection
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

Running Head: C4R Study Design
Abstract

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

Keywords: COVID-19, cohort study, epidemiology
The adverse effects of the coronavirus disease 2019 (COVID-19) pandemic on United States (US) health, economy, and society are widespread and will likely continue well beyond the initial waves of infections (1). Lack of preparedness and inadequate implementation and uptake of standard public health interventions in the US has already contributed to over 23 million cases, one million hospitalizations and over 500,000 deaths from COVID-19 (2, 3), making COVID-19 the third-leading cause of death in the United States in 2020 and the second-leading cause of death in those over 85 years of age (4, 5). Furthermore, prolonged symptoms and clinical abnormalities are observed in some COVID-19 survivors, raising concerns that post-acute sequelae of COVID-19 could pose an additional long-term health burden (6).

Epidemiologists have marshalled the strengths of numerous complementary study designs to identify the incidence and major clinical and socio-demographic risk factors for COVID-19 illness, as well as to describe post-COVID-19 outcomes. In particular, case-based registries and large-scale electronic health record (EHR)- and health systems-based cohorts provided critical early insights into disease susceptibility and short- and long-term sequelae. Among these were findings that socio-economic disadvantage (5, 7-9) and pre-existing clinical conditions, such as obesity, heart conditions, or lung disease (10-19), are associated with greater risk of severe illness.

Nonetheless, clinical and survey databases pose several problems for COVID-19 epidemiology. Clinical case series lack rigorous control groups, have non-standardized, limited data collection, and are subject to ascertainment biases – including, but not limited to, reduced health care
access and quality among vulnerable communities. EHRs typically lack detailed information on health-related behaviors, such as smoking, so that controlling for confounders is challenging. Moreover, in the course of usual clinical care, clinically actionable diagnostic testing is performed for sick persons, but not well persons; hence, subclinical disease is not well detected, and genomic and other mechanistic biomarkers are generally lacking. Although inception cohorts with longitudinal follow-up of clinically ascertained cases of COVID-19 cases can address some of these knowledge gaps, survival bias, recall biases, and non-randomly missing data regarding pre-COVID health and behaviors are inevitable. In this context, strong assumptions are required to define phenotypes identified in COVID-19 survivors (e.g., fibrotic lung disease) as “sequelae” when they may have been present prior to the pandemic, and actually be antecedent risk factors or effect modifiers.

The Collaborative Cohort of Cohorts for COVID-19 Research (C4R) was established as a national, prospective study of adults at risk for incident COVID-19 that is relatively free of referral, survival, and recall biases. C4R includes fourteen US prospective cohort studies that, collectively, constitute a large, well-characterized, population-based sample that ranges in age from young adults to centenarians, and reflects the racial, ethnic, socioeconomic, and geographic diversity of the US. Using standardized protocols, C4R is aggressively attempting full ascertainment of SARS-CoV-2 infection and COVID-19 illness across all cohorts. C4R offers the additional major advantages of standardized data collection protocols, including high-quality clinical events surveillance dating back as far as 1971 in some studies, and robust retention rates.
For decades, the C4R cohorts have collected extensive longitudinal data on clinical and subclinical disease, behaviors, cognition, biomarkers, and social determinants of health. C4R will link this “pre-COVID” phenotyping to information on SARS-CoV-2 infection and acute and post-acute COVID-related illness. The integration of antecedent and illness-related data will provide a unique opportunity to understand mechanisms and modifiers of risk and resilience for SARS-CoV-2 infection and adverse COVID-19 outcomes. C4R will also support 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 will allow the assessment and differentiation of the effects of infection, illness, and pandemic-related social, economic, and behavioral changes.

Overall, C4R aims to provide a valuable scientific resource to (1) evaluate risk and resilience factors for adverse COVID-19 outcomes, including severe COVID-19 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 disease, and (3) examine disparities in COVID-19 risk and outcomes according to race, ethnicity, geography, and other social determinants of health.

METHODS

Cohort of cohorts
Fourteen prospective cohorts are collaborating in C4R (Table 1). Eight of the cohorts were designed to study cardiovascular disease epidemiology: Atherosclerosis Risk in Communities (ARIC) Study (20), Coronary Artery Risk Development in Young Adults (CARDIA) Study (21), Framingham Heart Study (FHS) (22), Hispanic Community Health Study/Study of Latinos (HCHS/SOL) (23-25), Jackson Heart Study (JHS) (26-28), Mediators of Atherosclerosis in South Asians Living in America (MASALA) Study (29, 30), Multi-Ethnic Study of Atherosclerosis (MESA) (31), and the Strong Heart Study (SHS) (32, 33). These cohorts generally recruited population-based samples, although only three (ARIC, CARDIA, FHS, HCHS/SOL) used representational sampling techniques at some or all sites. Four of the cardiovascular studies (ARIC, CARDIA, FHS, MESA) recruited multi-racial participants, and four were designed to study primarily specific race or ethnic groups (Hispanic/Latino participants in HCHS/SOL, Black participants in JHS, South Asian participants in MASALA, American Indian participants in SHS). Four multi-ethnic cohorts were established to study respiratory epidemiology: the Genetic Epidemiology of COPD (COPDGene) Study (34) and the SubPopulations and InteRmediate Outcome Measures in COPD Study (SPIROMICS) (35) were established as longitudinal case-control studies of cigarette smokers with and without COPD; Prevent Pulmonary Fibrosis (PrePF) is a study of early and established interstitial lung disease; and, the Severe Asthma Research Program (SARP) is a study of the entire range of mild to severe asthma, enriched for severe disease (36). Two studies – the Northern Manhattan Study (NOMAS) and the REasons for Geographic and Racial Differences in Stroke (REGARDS) – were established to study primarily neurological outcomes, including stroke and cognition. NOMAS is a multi-ethnic community study (37) and REGARDS is
a biracial (non-Hispanic Black, White) national sample of the continental US that oversampled Black people and those residing in the southeast (38).

The cohorts that comprise C4R have collected detailed data on their participants’ health and behavior for as long as fifty years of follow-up (Figure 1). As summarized in Table 2, C4R cohorts have performed extensive longitudinal phenotyping of subclinical and clinical disease as well as assessments of laboratory biomarkers, ‘Omics, imaging, diet, behavior, and social determinants of health, and they have extensive biorepositories of stored specimens. 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 C4R cohorts use similar or identical adjudication protocols to ascertain all-cause mortality. Ten cohorts ascertain cardiovascular events including myocardial infarction, stroke, and heart failure. Eight cohorts ascertain respiratory events such as COPD and asthma exacerbations. Seven cohorts ascertain incident cognitive impairment and/or dementia.

Collaboration

Most of the cohorts have a long history of collaboration in the genomics-oriented Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) consortium (39), the NHLBI Pooled Cohorts Study focusing on respiratory epidemiology (40), the Cross-Cohort Collaboration (CCC) for cardiovascular epidemiology (41), the Blood Pressure and Cognition (BP COG) Study (42), and the genetic sequencing and multi-omics-focused Trans-Omics for Precision
Medicine (TOPMed) Project (43). C4R is building and expanding 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 became 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 National Heart, Lung, and Blood Institute (NHLBI) funded C4R via an Other Transactional Authority (OTA) mechanism in October 2020. Additional funding for inclusion of the neurology-focused cohorts was provided via the OTA by the National Institutes of Neurological Disorders and Stroke (NINDS) and the National Institute of Aging (NIA).

Leadership for C4R is provided by an organizing committee that includes leading – and often, founding – principal investigators (PIs) from all C4R cohorts, PIs from the C4R Data Coordination and Harmonization Center (DCHC), PIs from the C4R Biorepository and Central Laboratory (BCL), and program officers from the NHLBI, NINDS, and NIA. This organizing committee developed master C4R protocols for COVID-19 data collection.

Consistent with an ancillary studies model, each cohort in C4R is directly responsible for accomplishing its own data collection in accordance with the master protocol and under the
supervision of its respective Observational Study Monitoring Board (OSMB), Steering Committee, and any other applicable regulatory authorities.

To promote and sustain this broad collaborative effort, C4R PIs invited additional investigators and cohort personnel to participate in C4R committees and working groups. Study materials, including protocols and meeting materials, are posted regularly on a password-protected investigator section of the C4R website (c4r.nih.org).

Participants

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

Of 73,119 active participants across the fourteen cohorts, 53,972 participants were readily available for recruitment into C4R. Anticipated socio-demographic characteristics of potential C4R participants, estimated from current active cohort participants, are shown in Table 3. Fifty-eight percent of potential participants are 65 years or older, and thus at high risk for severe COVID-19. The anticipated sample is racially and ethnically diverse, based on self-report (44), with approximately 6% American Indian participants, 2% Asian participants, 26% Black participants, and 20% Hispanic/Latino participants.
All forty-eight continental states are represented among C4R participants, including rural, suburban, and urban communities (Figure 2). In all, C4R is being conducted across forty field/clinical centers, many of which are associated with more than one C4R cohort; one cohort with extensive geographic reach, the REGARDS, operates via telephone and in-home exams only (38).

Data collection

COVID-19 questionnaires. C4R is ascertaining self-reported COVID-related experiences by questionnaire. Each cohort will deploy C4R questionnaires twice within 18 months following the initial outbreak in March 2020 via telephone, mail-in, online, email, or smartphone apps. Wave 1 questionnaires were developed as early as March 2020 in certain cohorts (45) and urgently administered in spring and summer 2020. Although these efforts pre-dated 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. Wave 2 questionnaires were fully standardized to include domains on COVID-19 infection, testing, hospitalization, symptoms, recovery, re-infection, contacts, vaccination, behavioral changes, sleep, memory loss, depression, anxiety, fatigue, and resilience. The C4R questionnaire was developed collaboratively to include validated and PhenX toolkit instruments (https://www.phenxtoolkit.org) (46-55) in order to optimize comparability with pre-pandemic assessments and across C4R and other epidemiology cohorts. The C4R questionnaire, including
translations into Spanish and Mandarin, are available on PhenX; Research Electronic Data Capture (REDCap (56, 57)) programming may be available on request.

**COVID-related events ascertainment.** C4R is ascertaining COVID-related hospitalizations and deaths that are identified via the C4R questionnaire or other surveillance methods available to the cohorts, including EHR linkages, where available. Each cohort is using its own established infrastructure for ascertainment of medical records and death certificates, including use of the National Death Index (NDI), the Centers for Medicare & Medicaid Services (CMS), International Classification of Diseases (ICD) codes (58), and linkage to records from local departments of health. Cohorts may review events locally at their Field/Coordinating Centers or transfer records for central review by C4R. The C4R events review is designed to assess 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 to classify and validate cardiovascular, respiratory (19), and thromboembolic (59) events. Protocols for ascertainment, review, and classification are available on the study website ([c4r-nih.org](http://c4r-nih.org)).

**Dried blood spot collection.** C4R is ascertaining serostatus by dried blood spot (DBS) in 2021. Cohort field centers receive DBS collection kits from the BCL and are responsible for recruitment, consent, and distribution to participants. Updated details regarding vaccination status are obtained at the time of DBS consent and immediately prior to mailing the DBS kit to the participant. Participants mail the completed kits directly to the BCL or to the cohort field or
coordinating center as an intermediary step. Participant instructions, including a video, are provided by the cohort and via the C4R website (c4r.nih.org) and/or cohort-specific websites. In cohorts with upcoming in-person exams, the DBS may be collected in-person by research staff.

C4R Common Data Elements

C4R data collection will define a spectrum of COVID-19 outcomes. Ascertainment of COVID-related hospitalizations and deaths will characterize, classify, and validate moderate-to-severe COVID-19 illnesses. In addition to identifying these events, questionnaires are being used to obtain self-reported information on the nature, severity, and duration of symptoms during acute infection and in the post-acute setting. This will support classification of symptomatic and asymptomatic infections, as well as cases of prolonged recovery or post-acute sequelae of SARS-CoV-2 infection (PASC). Data on behaviors, attitudes, psychosocial impacts, and vaccinations will also be collected. Seropositive individuals without self-reported infection will be reclassified as infected, whereas seronegative individuals with prior positive testing by self-report or health records will be classified as sero-reverted.

Data management

C4R data collection is coordinated centrally at the DCHC at Columbia University Irving Medical Center. Electronic data collection forms are being programmed into REDCap for use or adaptation by the cohort coordinating centers. Metadata on completion of questionnaires,
events ascertainment, and DBS kit collection status are reported and reviewed bi-weekly to ensure operational milestones are met. Participants are assigned a C4R study identifier by cohort-specific coordinating centers that is used for participant-level data transfers and analyses.

Biorepository and central laboratory and serology measurement

The C4R BCL at the University of Vermont is responsible for establishing a C4R biorepository of DBS, plus other biospecimens that may be collected in the future, and for performing and/or coordinating performance of any centralized clinical and biomarker assays and serology assays. Individual DBS Collection kits are produced by the BCL and shipped to the cohorts (either to the individual field centers or the cohort coordinating center, based on cohort preference). Kits and DBS cards are labeled with a biospecimen identifier, which is linked to C4R identifiers that are maintained centrally and not shared with the BCL, through the use of a “linking key.” Filled DBS cards are returned to the BCL, and batches prepared for serology assays performed by the New York State Wadsworth Center’s Bloodborne Viruses Laboratory (BVL) under CLIA and New York State certification. The BVL performs a SARS-CoV-2 IgG Microsphere Immunoassay using Luminex bead technology for qualitative detection of human IgG antibodies to SARS-CoV-2 nucleocapsid (N) and spike subunit 1 (S1) antigens. Based on testing 730 pre-COVID DBS and >1100 DBS from individuals with laboratory-confirmed infection, specificity is 99.5% for both N and S1 and sensitivity ranged from 90 to 96% for symptomatic individuals and 77 to 91% for asymptomatic individuals. Sensitivity increased for both groups with time from positive PCR
test, accounting for the range. This assay was used successfully to test over 57,000 DBS for statewide serosurveys from April-June as part of New York State’s public health response.

Serology results are reported by the BVL to the C4R BCL, and then to the cohort coordinating centers, which are responsible for a) recombining the results with the proper participants based on the “linking key”, and b) reporting results to participants according to usual cohort practices. Serological results are not believed to have clinical relevance, and the CDC does not currently recommend modifications to individual behavior or clinical care based on antibody status alone (60); hence, no protocols for “alert” findings have been established, and participants may opt out of results return. Protocols for the serosurvey are available on the study website (c4r.nih.org).

Since all current vaccines in use in the U.S. generate an immune response to the Spike protein, we anticipate being able to distinguish vaccination from viral infection by the use of the anti-nucleocapsid assay results (61).

Harmonization

Harmonization of COVID-19 and pre-pandemic data will be performed centrally to define COVID-19 common data elements and to align pre-pandemic data for large-scale, longitudinal analyses. This effort will leverage prior harmonization efforts across C4R cohorts in the TOPMed Project, the NHLBI Pooled Cohorts Study, the BP COG Study, and the CHARGE Working Groups (10, 40, 42, 62-67). Due to their significance to COVID-19 epidemiology, particular emphasis will
be placed on harmonizing the large amount of deep pre-pandemic physiologic (40), neurocognitive (68-74), and imaging-based (75-82) phenotyping collected within the decade prior to the outbreak using deep-learning (18, 83-85) and other methods (Table 4).

Quality control

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

Data sharing

The C4R Commons Agreement, modeled on the CHARGE Analysis Commons Consortium Agreement (86), will expedite cross-cohort data harmonization and sharing, as allowed (87). Following review and approval, cohort-specific agreements would permit COVID-19 and pre-pandemic data to be uploaded to the NIH-supported cloud computing platform, hosted by BioData Catalyst. Access to the pooled C4R dataset would be granted to investigators involved in core harmonization efforts and those with manuscript proposals approved by C4R.
publications and cohort coordinating committees. Once harmonization and related quality control is completed, C4R common data elements will be transferred as a limited dataset for public access on BioData Catalyst in accord with cohort-specific consents and commitments.

Governance

The administrative coordinating center for C4R is the NHBLI CONNECTS program (nhlbi-connects.org). Metadata on operational progress is submitted biweekly to CONNECTS for tracking and review purposes. Central functions of C4R are overseen by a C4R OSMB convened by CONNECTS.

DISCUSSION

C4R will leverage existing American cohort studies to develop a large, multi-ethnic, pooled cohort of participants with incident COVID-19 and COVID-free participants that is relatively free of referral, survival, and recall biases compared to clinically based inception cohorts of COVID-19 patients. C4R includes a highly 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 unparalleled wealth of pre-pandemic 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, PASC, and other long-term effects of the pandemic response. Unlike case registries and EHR-
based studies, C4R’s repeated exams and cognitive assessments before and after COVID-19 also provide important opportunities to estimate the social and behavioral impact of the COVID-19-related pandemic response on changes in long-term mental and physical health across multiple domains.

C4R will provide important opportunities for future studies using a range of epidemiologic study designs. For example, nested within C4R, longitudinal cohort studies of COVID-affected and unaffected participants could repeat a variety of subclinical measures (e.g., echocardiography, lung imaging, neuro-cognitive assessment) to define reliably the consequences of COVID-19 infection. Ongoing high-quality events follow-up will allow assessment of long-term clinical health outcomes following COVID-19 and the pandemic period. The extensive biobanks maintained by the cohorts could support measurement of prior viral infections, immune-phenotypes, metabo-types, ‘Omics, and other pre-COVID characteristics that may be risk determinants or modifiers for COVID-19 susceptibility and vaccine effectiveness. The fact that the cohorts continue to follow their participants provides a dynamic resource to study emerging questions in COVID-19 epidemiology, including but not limited to viral variants and vaccination. And, 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 the strategic transformation of population studies (88).
ACKNOWLEDGEMENTS

Author affiliations

Albert Einstein College of Medicine

- Department of Epidemiology and Population Health, Albert Einstein College of Medicine, Bronx, New York, United States (Carmen R Isasi, Robert C Kaplan)

Boston Children’s Hospital/ Harvard Medical School

- Department of Medicine, Division of Allergy and Immunology, Boston Children’s Hospital, Boston Massachusetts (Wanda Phipatanakul)

Boston University

- Department of Pathology & Laboratory Medicine, Boston University School of Medicine, Boston, Massachusetts, United States (Joel M Henderson)
- Departments of Medicine and Epidemiology, Boston University, Boston, Massachusetts, United States (Ramachandran S Vasan)
- Department of Medicine, School of Medicine, and Department of Biostatistics, School of Public Health, Boston University, Boston, Massachusetts, United States (Vanessa Xanthakis)

Brigham and Women’s Hospital/ Harvard Medical School

- Pulmonary and Critical Care Medicine, Brigham and Women's Hospital, Boston, Massachusetts (Bruce Levy, Matthew R Moll)
- Channing Division of Network Medicine, Brigham and Women's Hospital, Boston, MA (Matthew Moll).
Colorado Anschutz Medical Campus

- Division of Pulmonary Sciences and Critical Care, Department of Medicine, University of Colorado Anschutz Medical Campus, Aurora, Colorado, United States (Alyssa Asaro, Joyce Lee, David Schwartz)
- Department of Epidemiology, School of Public Health, University of Colorado Anschutz Medical Campus, Aurora, Colorado, United States (Gregory Kinney)

Columbia University

- Division of General Medicine, Department of Medicine, Columbia University Medical Center, New York, New York, United States (Pallavi P Balte, R Graham Barr, Akshaya Krishnaswamy, Elizabeth C Oelsner, Priya Palta, Yiyi Zhang)
- Columbia Data Coordinating Center, New York State Psychiatric Institute and Columbia College of Physicians and Surgeons, Columbia University Medical Center, New York, New York, United States (Howard F Andrews)
- Department of Neurology, Vagelos College of Physicians and Surgeons, Columbia University, New York, New York, United States (Mitchell Elkind, Jose Gutierrez, Jennifer J Manly)
- Department of Epidemiology, Mailman School of Public Health, New York, NY United States (Ryan T Demmer [adjunct], Mitchell Elkind)

Georgetown University

- MedStar Health Research Institute and Department of Medicine, Georgetown University, Washington, District of Columbia, United States (Jason G Umans)
Johns Hopkins University

- Department of Epidemiology, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, Maryland, United States (Josef Coresh)
- Division of Cardiology, Department of Medicine, Johns Hopkins University, Baltimore, Maryland, United States (Wendy S Post)

Lundquist Institute

- Lundquist Institute, Los Angeles, California, United States (Jerome I Rotter)

National Institutes of Health (intramural)

- Division of Intramural Research, National Heart Lung and Blood Institute of the National Institutes of Health (Veronique Roger)

National Jewish Health

- Division of Pulmonary, Critical Care & Sleep Medicine, Department of Medicine, National Jewish Health, Denver, Colorado, United States (Barry Make)
- Division of Rheumatology, Department of Medicine, National Jewish Health, Denver, Colorado, United States (Elizabeth A Regan)
- National Jewish Health, Denver, Colorado, United States (Grace Chen)

New York State Department of Health

- Bloodborne Viruses Laboratory, Wadsworth Center, New York State Department of Health, Albany, New York, United States (Monica Parker)

Northwestern
• Center for Epidemiology and Population Health, Department of Preventive Medicine, Northwestern University, Chicago, Illinois, United States (Norinna Bai Allen, Namratha R Kandula)

• Division of General Internal Medicine, Department of Medicine, Northwestern University, Chicago, Illinois, United States (Namratha R Kandula)

Pennsylvania State University

• Division of Biostatistics and Bioinformatics, Department of Public Health Sciences, Pennsylvania State University, State College, Pennsylvania, United States (Dave Mauger)

San Diego State University

• South Bay Latino Research Center, Department of Psychology, San Diego State University, San Diego, California, United States (Linda Gallo, Gregory A Talavera)

Texas Biomedical Research Institute

• Population Health Program, Texas Biomedical Research Institute, San Antonio, Texas, United States (Shelley A Cole)

University of Alabama at Birmingham

• Department of Biostatistics, University of Alabama at Birmingham, Birmingham, Alabama, United States (Suzanne E Judd)

• Department of Epidemiology, University of Alabama at Birmingham, Birmingham, Alabama, United States (Kelley Pettee Gabriel, Virginia J Howard, Cora E Lewis, Emily B Levitan)
• Division of Preventive Medicine, School of Medicine, University of Alabama at Birmingham, Birmingham, Alabama, United States (James M Shikany)

UCLA

• The Institute for Translational Genomics and Population Sciences, Department of Pediatrics, The Lundquist Institute for Biomedical Innovation at Harbor-UCLA Medical Center, Torrance, CA USA (Jerome I Rotter)

• Division of Cardiology, Department of Medicine, UCLA Medical Center, Los Angeles, California, United States (Karol E Watson)

University of California San Francisco

• Department of Medicine, University of California San Francisco, San Francisco, California, United States (Alka M. Kanaya, Arune A. Chang)

• Department of Obstetrics, Gynecology & Reproductive Sciences, University of California San Francisco, San Francisco, California, United States (Michael Schembri)

• Division of Pulmonary, Critical Care, Allergy, and Sleep Medicine, Department of Medicine, University of California San Francisco, San Francisco, California, United States (Prescott G. Woodruff)

University of Kentucky

• Department of Epidemiology, College of Public Health, University of Kentucky, Lexington, KY (Anna M Kucharska-Newton).

University of Miami
- Department of Neurology and Evelyn F. McKnight Brain Institute, University of Miami, Miami, Florida, United States (Tatjana Rundek, Ralph L. Sacco)

University of Michigan

- Division of Pulmonary and Critical Care, Department of Medicine, University of Michigan, Ann Arbor, Michigan, United States (MeiLan K. Han)
- Division of General Medicine, Department of Medicine, University of Michigan, Ann Arbor, Michigan, United States (Deborah A Levine)

University of Minnesota

- 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)

University of Mississippi Medical Center

- Departments of Medicine and Population Health Science, School of Public Health, University of Mississippi Medical Center, Jackson, Mississippi, United States (Pramod Anugu, Adolfo Correa, Mario Sims, Yuan-I Min)
- School of Nursing, University of Mississippi Medical Center, Jackson, Mississippi, United States (Karen Winters)

University of North Carolina

- Collaborative Studies Coordinating Center, Department of Biostatistics, Gillings School of Global Public Health, University of North Carolina, Chapel Hill, North Carolina, United States (David Couper, Kimberly Ring)
• Department of Epidemiology, Gillings School of Global Public Health, University of North Carolina, Chapel Hill, North Carolina, United States (Anna Kucharska-Newton)

• Department of Nutrition, Gillings School of Global Public Health, University of North Carolina, Chapel Hill, North Carolina, United States (Katie Meyer)

University of Oklahoma Health Sciences Center

• Center for American Indian Health Research, Department of Biostatistics and Epidemiology, Hudson College of Public Health, University of Oklahoma Health Sciences Center, Oklahoma City, Oklahoma, United States (Tauqeer Ali, Kimberly Malloy, Ying Zhang)

University of Pittsburgh

• Department of Environmental and Occupational Health, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, Pennsylvania, United States (Sally E Wenzel)

• Division of Pulmonary, Allergy, and Critical Care Medicine, Department of Medicine, University of Pittsburgh, Pittsburgh, Pennsylvania, United States (Jessica Bon)

UT Health San Antonio

• Glenn Biggs Institute for Alzheimer's & Neurodegenerative Diseases, Graduate School of Biomedical Sciences, UT Health San Antonio, San Antonio, Texas, United States (Sudha Seshadri)

University of Vermont
• Laboratory for Clinical Biochemistry Research, Department of Pathology & Laboratory Medicine, Larner College of Medicine, University of Vermont, Burlington, Vermont, United States (Rebekah Boyle, Elaine Cornell, Russell Tracy)

• Department of Medicine, Vermont Center for Cardiovascular and Brain Health, Larner College of Medicine at the University of Vermont, Burlington, Vermont, United States (Mary Cushman, Debora Kamin Mukaz)

University of Washington

• Department of Epidemiology, School of Public Health, University of Washington, Seattle, Washington, United States (Amanda M Fretts, Robert Kaplan, Bruce Psaty)

• Department of Health Services, School of Public Health, University of Washington, Seattle, Washington, United States (Bruce M Psaty)

• Division of General Medicine, Department of Medicine, University of Washington, Seattle, Washington, United States (Bruce Psaty)

• Department of Biostatistics, School of Public Health, University of Washington, Seattle, Washington, United States (Karen Hinckley Stukovsky)

Wake Forest School of Medicine

• Department of Epidemiology and Prevention, Department of Medicine, Wake Forest School of Medicine, Winston-Salem, North Carolina, United States (Alain G Bertoni)

• Division of Pulmonary, Critical Care, Allergy, and Immunologic Diseases, Department of Medicine, Wake Forest School of Medicine (Wendy C Moore, Victor E Ortega)
# Author contributions

All authors meet ICJME criteria for authorship.

## Funding

<table>
<thead>
<tr>
<th>Study</th>
<th>Funding</th>
</tr>
</thead>
<tbody>
<tr>
<td>C4R</td>
<td>C4R is supported by NHLBI-CONNECTS OT2HL156812</td>
</tr>
<tr>
<td>ARIC</td>
<td>The Atherosclerosis Risk in Communities study has been funded in whole or in part with Federal funds from the National Heart, Lung, and Blood Institute, National Institutes of Health, Department of Health and Human Services, under Contract nos. (HHSN268201700001I, HHSN268201700002I, HHSN268201700003I, HHSN268201700005I, HHSN268201700004I). Neurocognitive data are collected by U01 2U01HL096812, 2U01HL096814, 2U01HL096899, 2U01HL096902, 2U01HL096917 from the NHLBI, NINDS, NIA and NIDCD. Ancillary studies funded additional data elements.</td>
</tr>
<tr>
<td>BP COG</td>
<td>The BP COG Study is supported by the National Institute of Neurological Disorders and Stroke, National Institutes of Health, and US Department of Health and Human Services (grant R01 NS102715).</td>
</tr>
<tr>
<td>CARDIA</td>
<td>The Coronary Artery Risk Development in Young Adults Study (CARDIA) is conducted and supported by the National Heart, Lung, and Blood Institute (NHLBI) in collaboration with the University of Alabama at Birmingham (HHSN268201800005I &amp; HHSN268201800007I), Northwestern University (HHSN268201800003I), University of Minnesota (HHSN268201800006I), and Kaiser Foundation Research Institute (HHSN268201800004I). This manuscript has been reviewed by CARDIA for scientific content.</td>
</tr>
</tbody>
</table>
| COPDGene | Grant Support and Disclaimer  
The project described was supported by Award Number U01 HL089897 and Award Number U01 HL089856 from the National Heart, Lung, and Blood Institute. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Heart, Lung, and Blood Institute or the National Institutes of Health.  
COPD Foundation Funding  
COPDGene is also supported by the COPD Foundation through contributions made to an Industry Advisory Board comprised of AstraZeneca, Boehringer-Ingelheim, Genentech, GlaxoSmithKline, Novartis, Pfizer, Siemens, and Sunovion. |
| FHS | The Framingham Heart Study (FHS) acknowledges the support of Contracts NO1-HC-25195, HHSN268201500001I and 75N92019D00031 from the National Heart, Lung and Blood Institute. We also acknowledge the dedication of the FHS study participants without whom this research would not be possible. |
The Hispanic Community Health Study/Study of Latinos is a collaborative study supported by contracts from the National Heart, Lung, and Blood Institute (NHLBI) to the University of North Carolina (HHSN268201300001I / N01-HC-65233), University of Miami (HHSN268201300004I / N01-HC-65234), Albert Einstein College of Medicine (HHSN268201300002I / N01-HC-65235), University of Illinois at Chicago – HHSN268201300003I / N01-HC-65236 Northwestern Univ), and San Diego State University (HHSN268201300005I / N01-HC-65237). The following Institutes/Centers/Offices have contributed to the HCHS/SOL through a transfer of funds to the NHLBI: National Institute on Minority Health and Health Disparities, National Institute on Deafness and Other Communication Disorders, National Institute of Dental and Craniofacial Research, National Institute of Diabetes and Digestive and Kidney Diseases, National Institute of Neurological Disorders and Stroke, NIH Institution-Office of Dietary Supplements.

The Jackson Heart Study (JHS) is supported and conducted in collaboration with Jackson State University (HHSN268201800013I), Tougaloo College (HHSN268201800014I), the Mississippi State Department of Health (HHSN268201800015I) and the University of Mississippi Medical Center (HHSN268201800010I, HHSN268201800011 I and HHSN268201800012I) contracts from the National Heart, Lung, and Blood Institute (NHLBI) and the National Institute on Minority Health and Health Disparities (NIMHD).

MASALA was supported by Grant Number R01HL093009 from the National Heart, Lung, And Blood Institute and the National Center for Research Resources and the National Center for Advancing Translational Sciences, National Institutes of Health, through UCSF-CTSI Grant Number UL1RR024131.

MESA and the MESA SHARe projects are conducted and supported by the National Heart, Lung, and Blood Institute (NHLBI) in collaboration with MESA investigators. Support for MESA is provided by contracts and grants 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, 75N9202000007, 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 National Heart, Lung, and Blood Institute; R01-AG058969 from the National Institute on Aging; and by grants UL1-TR-000040, UL1-TR-001079, and UL1-TR-001420 from the National Center for Advancing Translational Sciences (NCATS). Funding for SHAre genotyping was provided by NHLBI Contract N02-HL-64278.

This publication was developed under the Science to Achieve Results (STAR) research assistance agreements, No. RD831697 (MESA Air) and RD-83830001 (MESA Air Next Stage), awarded by the U.S Environmental Protection Agency. It has not been formally reviewed by the EPA. The views expressed in this document are solely those
of the authors and the EPA does not endorse any products or commercial services mentioned in this publication.

Whole genome sequencing (WGS) for the Trans-Omics in Precision Medicine (TOPMed) program was supported by the NHLBI. WGS for “NHLBI TOPMed: Multi-Ethnic Study of Atherosclerosis (MESA)” (phs001416.v1.p1) was performed at the Broad Institute of MIT and Harvard (3U54HG003067-13S1). Centralized read mapping and genotype calling, along with variant quality metrics and filtering were provided by the TOPMed Informatics Research Center (3R01HL-117626-02S1, contract HHSN268201800002I) (Broad RNA Seq, Proteomics HHSN268201600034I, UW RNA Seq HHSN268201600032I, USC DNA Methylation HHSN268201600034I, Broad Metabolomics HHSN268201600038I). Phenotype harmonization, data management, sample-identity QC, and general study coordination, were provided by the TOPMed Data Coordinating Center (3R01HL-120393; U01HL-120393; contract HHSN268180001I).

The provision of genotyping data was supported in part by the National Center for Advancing Translational Sciences, CTSI grant UL1TR001881, and the National Institute of Diabetes and Digestive and Kidney Disease Diabetes Research Center (DRC) grant DK063491 to the Southern California Diabetes Endocrinology Research Center.

The authors thank the other investigators, the staff, and the participants of the MESA study for their valuable contributions. A full list of participating MESA investigators and institutions can be found at [http://www.mesa-nhlbi.org](http://www.mesa-nhlbi.org).

<table>
<thead>
<tr>
<th>NHLBI Pooled Cohorts Study</th>
<th>NIH/NHLBI R21HL153700, K23HL130627, R21HL129924), R21HL121457</th>
</tr>
</thead>
<tbody>
<tr>
<td>NOMAS</td>
<td>Supported by NINDS R01 NS29993 and R01 AG066162.</td>
</tr>
<tr>
<td>PrePF</td>
<td>Prevent Pulmonary Fibrosis cohort was established in 2000 and has been supported by: X01-ES101947, R01-HL095393, RC2-HL1011715, R21/33-HL120770, R01-HL097163, X01-HL134585, UH2/3-HL123442, P01-HL092870, UG3/UH3-HL151865, and DoD W81XWH-17-1-0597.</td>
</tr>
<tr>
<td>REGARDS</td>
<td>This research project is supported by cooperative agreement U01 NS041588 co-funded by the National Institute of Neurological Disorders and Stroke (NINDS) and the National Institute on Aging (NIA), National Institutes of Health, Department of Health and Human Service. The content is solely the responsibility of the authors and does not necessarily represent the official views of 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. The authors thank the other investigators, the staff, and the participants of the REGARDS study for their valuable contributions.</td>
</tr>
<tr>
<td>Study</td>
<td>Funding Information</td>
</tr>
<tr>
<td>-------</td>
<td>--------------------</td>
</tr>
<tr>
<td><strong>SARP</strong></td>
<td>The research by the principal and co-principal investigators was funded by National Institutes of Health/National Heart Lung Blood Institute (NIH/NHLBI) Severe Asthma Research Program: U10 HL109164, U10 HL109257, U10 HL109146), U10 HL109172, U10 HL109250, U10 HL109168, U10 HL109152, U10 HL109086). Industry Partnerships also provided additional support: AstraZeneca, Boehringer-Ingelhein, Genetech, GlaxoSmithKline, MedImmune, Novartis, Regeneron, Sanofi and TEVA. Spirometers used in SARP III were provided by nSpire Health (Longmont, Colo).</td>
</tr>
<tr>
<td><strong>SPIROMICS</strong></td>
<td>SPIROMICS has been funded by contracts from the NIH/NHLBI (HHSN268200900013C, HHSN268200900014C, HHSN268200900015C, HHSN268200900016C, HHSN268200900017C, HHSN268200900018C, HHSN268200900019C, HHSN268200900020C), grants from the NIH/NHLBI (U01 HL137880 and U24 HL141762), and supplemented by contributions made through the Foundation for the NIH and the COPD Foundation from AstraZeneca/MedImmune; Bayer; Bellerophon Therapeutics; Boehringer-Ingelheim Pharmaceuticals, Inc.; Chiesi Farmaceutici S.p.A.; Forest Research Institute, Inc.; GlaxoSmithKline; Grifols Therapeutics, Inc.; Ikaria, Inc.; Novartis Pharmaceuticals Corporation; Nycomed GmbH; ProterixBio; Regeneron Pharmaceuticals, Inc.; Sanofi; Sunovion; Takeda Pharmaceutical Company; and Theravance Biopharma and Mylan.</td>
</tr>
<tr>
<td><strong>SHS</strong></td>
<td>The Strong Heart Study has been funded in whole or in part with federal funds from the National Heart, Lung, and Blood Institute, National Institute of Health, Department of Health and Human Services, under contract numbers 75N92019D00027, 75N92019D00028, 75N92019D00029, &amp; 75N92019D00030. The study was previously supported by research grants: R01HL109315, R01HL109301, R01HL109284, R01HL109282, and R01HL109319 and by cooperative agreements: U01HL41642, U01HL41652, U01HL41654, U01HL65520, and U01HL65521.</td>
</tr>
</tbody>
</table>
Thank you’s

We thank the participants of each cohort for their dedication to the studies.

Members of the study group

Ali, Tauqueer
Allen, Nori
Andrews, Howard
Anugu, Pramod
Arora, Komal
Arynchyn, Alex
Asaro, Alyssa
Balte, Pallavi P
Bancks, Mike
Barr, R Graham
Bateman, Lori
Bertoni, Alain
Bick, Alexander G
Bleecker, Eugene
Bluemke, David
Boyle, Rebekah
Budoff, Matt
Cardwell, Jonathan
Floyd, James
Freeman, Christine
Fretts, Amanda
Gabriel, Kelley
Gallo, Linda
Gharib, Sina
Goode, Caroline
Goyal, Parag
Griswold, Mike
Han, MeiLan
Hanna, David
Hastie, Annette
Heckbert, Susan R.
Henderson, Joel
Hinckley Stukovsky, Karen
Hoffman, Eric
Hoffman, Udo
Hornig, Mady
Howard, Virginia J
Isasi, Carmen R
Jacobs, David R Jr
Jaquis, Cashell
Judd, Suzanne E
Kamin Mukaz, Debora
Kanaya, Alka M
Kandula, Namratha R
Kaplan, Robert
Khan, Sadiya
King, Jonathan
Kinney, Gregory L
Kucharska-Newton, Anna
Kumi, Smith
Laine, Andrew F.
Lee, Joyce
Lemaitre, Rozenn
Levin, Bonnie
Levine, Deborah
Levitan, Emily B.
Levy, Bruce
Lima, Joao
Lynch, David
Make, Barry
Malloy, Kimberly
Manly, Jennifer J
Mauger, Dave
Melius, Katy
Mendoza-Puccini, Carolina
Merkin, Sharon
Meyer, Katie
Meyers, Deborah
Min, Yuan-I (Nancy)
Minotti, Melissa
Moise, Nathalie
Moore, Wendy
Moy, Claudia
Mutalik, Karen
Nasrallah, Ilya
Nelson, Cheryl
Nelson, Lauren
Noel, Patricia
Nordvig, Anna
O'Connor, George
Odden, Michelle
Oelsner, Elizabeth
O'Leary, Marcia
O'Neal, Wanda
Vuga, Louis
Wagster, Molly
Wang, Henry
Washko, George
Wentzel, Sally
West, Cynthia
Wilson, Carla
Woodruff, Prescott
Wright, Jackie
Xanthakis, Vanessa
Yuan, Ya
Zakai, Neil
Zhang, Ying
Zhang, Yiyi

Disclaimer

The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Conflict of Interest Statement
Ali, Asaro, Bertoni, Boyle, Cole, Coresh, Cornell, Correa, Cushman, Demmer, Folsom, Fretts,

Howard, Isasi, Jacobs, Kamin Mukaz, Kandula, Malloy, Meyer, Oelsner, Parker, Pettee, Ring,

Roger, Schembri, Seshadri, Shikany, Tracy, Vasan, Winters, Zhang: None

Mitchell Elkind receives royalties from UpToDate for a chapter on neurological complications of COVID-19; receives study drug in kind from the BMS-Pfizer Alliance for Eliquis and ancillary funding from Roche, both for an NIH-funded trial of stroke prevention.

MeiLan K Han reports consulting for BI, GSK, AZ, Merck, Mylan, Verona, Teva, Cipla, Chiesi and Sanofi. She reports research support from Novartis and Sunovion.

Emily B. Levitan has research funding from Amgen and has received consulting fees for a research project sponsored by Novartis.

Wanda Phipatanakul has funding and trial medication support from Astra Zeneca, GSK, Merck, Genentech, Novartis, Regeneron, Sanofi for asthma studies and consulting from GSK, Genentech, Novartis, Sanofi, Regeneron for asthma therapeutics.

Bruce M Psaty serves on the Steering Committee of the Yale Open Data Access Project funded by Johnson & Johnson.

David Schwartz is founder and chief scientific officer of Eleven P15, a company dedicated to the diagnosis, prevention, and treatment of early presentations of pulmonary fibrosis. Joyce Lee serves as a consultant for Eleven P15.

Sally Wenzel receives funding for consulting and clinical trials from AstraZeneca, GSK, Sanofi-Genzyme, Novartis, Knopp; she also receives research support from Pieris and Regeneron.
Prescott Woodruff has a research grant from Genentech and is a consultant for Sanofi,
Regeneron, Clarus ventures, 23andMe, Theravance, Astra Zeneca, Glenmark Pharmaceuticals,
Amgen, and NGM Pharma, outside the submitted work.
Table 1. Characteristics of participants in C4R cohorts on March 1, 2020.

<table>
<thead>
<tr>
<th>Cohort</th>
<th>N, C4R target population</th>
<th>Current age range, years</th>
<th>NHW</th>
<th>B</th>
<th>H/L</th>
<th>As</th>
<th>Am Ind</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>ARIC</td>
<td>5,046</td>
<td>75-97</td>
<td>77</td>
<td>23</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>CARDIA</td>
<td>4,221</td>
<td>53-66</td>
<td>50</td>
<td>50</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>COPDGene</td>
<td>4,000</td>
<td>50-90</td>
<td>65</td>
<td>35</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>FHS</td>
<td>7,574</td>
<td>26-108</td>
<td>86</td>
<td>3</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>7</td>
</tr>
<tr>
<td>HCHS/SOL</td>
<td>8,400</td>
<td>30-87</td>
<td>0</td>
<td>0</td>
<td>100</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>JHS</td>
<td>2,900</td>
<td>38-102</td>
<td>0</td>
<td>100</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>MASALA</td>
<td>500</td>
<td>50-94</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>100</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>MESA</td>
<td>4,683</td>
<td>65-103</td>
<td>38</td>
<td>27</td>
<td>24</td>
<td>12</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>NOMAS</td>
<td>1,267</td>
<td>62-106</td>
<td>12</td>
<td>14</td>
<td>72</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>PrePF</td>
<td>2,500</td>
<td>40-80</td>
<td>92</td>
<td>3</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>REGARDS</td>
<td>8,000</td>
<td>57-105</td>
<td>62</td>
<td>38</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>SARP</td>
<td>380</td>
<td>18-80</td>
<td>75</td>
<td>25</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>SPIROMICS</td>
<td>1,800</td>
<td>47-87</td>
<td>82</td>
<td>4</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>SHS</td>
<td>2,701</td>
<td>31-105</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>100</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

NHW = Non-Hispanic White. B = Black. H/L = Hispanic/Latinx. As = Asian American. Am Ind = American Indian. Of note, ARIC did not inquire regarding Hispanic/Latino ethnicity, hence White participants cannot be definitely defined as non-Hispanic.

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
Table 2. Density of phenotype data in C4R, 1971-2025.

<table>
<thead>
<tr>
<th>Incident events/diagnoses</th>
<th>ARIC</th>
<th>CARDIA</th>
<th>COPD Gene</th>
<th>FHS</th>
<th>HCHS/SOL</th>
<th>JHS</th>
<th>MAS-ALA</th>
<th>MESA</th>
<th>NOMAS</th>
<th>PrePF</th>
<th>REG-ARDS</th>
<th>SARP</th>
<th>SPIROMICS</th>
<th>SHS</th>
</tr>
</thead>
<tbody>
<tr>
<td>MI, CVA, TIA, HF</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>COPD, asthma</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>Cog. Impairment/Dementia</td>
<td>A</td>
<td>A**</td>
<td>A</td>
<td>A</td>
<td>A**</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>Diabetes</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>Chronic kidney disease</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>Mortality</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>Echocardiography</td>
<td>2</td>
<td>3</td>
<td>6</td>
<td>1</td>
<td>1</td>
<td>1**</td>
<td>1**</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1**</td>
<td>1</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>Cardiac MRI</td>
<td>1</td>
<td>4</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Carotid IMT/plaque</td>
<td>4</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>1**</td>
<td>1</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>Brain MRI</td>
<td>4**</td>
<td>3**</td>
<td>3</td>
<td>1**</td>
<td>1**</td>
<td>1**</td>
<td>1**</td>
<td>1**</td>
<td>1</td>
<td>1</td>
<td>1**</td>
<td>1</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>CT coronary artery calcium</td>
<td>1</td>
<td>4</td>
<td>1</td>
<td>2</td>
<td>1**</td>
<td>1**</td>
<td>1**</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Pulmonary function</td>
<td>3</td>
<td>7**</td>
<td>3</td>
<td>6</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td>6</td>
<td>5</td>
</tr>
<tr>
<td>CT lung imaging</td>
<td>1**</td>
<td>3</td>
<td>1</td>
<td></td>
<td>1</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Polysomnography</td>
<td>1</td>
<td>2**</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2*</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Cognitive assessments</td>
<td>6*</td>
<td>3**</td>
<td>1</td>
<td>9</td>
<td>3**</td>
<td>1**</td>
<td>4</td>
<td>20</td>
<td>6*</td>
<td>40</td>
<td>6*</td>
<td>1</td>
<td>6**</td>
<td>2</td>
</tr>
<tr>
<td>Psychological assessments</td>
<td>6*</td>
<td>9**</td>
<td>9</td>
<td>9</td>
<td>3**</td>
<td>4**</td>
<td>2**</td>
<td>7**</td>
<td>20</td>
<td>20</td>
<td>2**</td>
<td>2</td>
<td>6**</td>
<td>2</td>
</tr>
<tr>
<td>Geocoding</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>Metabolic (lipids, glucose)</td>
<td>8**</td>
<td>10**</td>
<td>9</td>
<td>3**</td>
<td>4**</td>
<td>3**</td>
<td>7**</td>
<td>3**</td>
<td>3**</td>
<td>3**</td>
<td>2</td>
<td>2</td>
<td>6**</td>
<td>6**</td>
</tr>
<tr>
<td>Inflammatory</td>
<td>8**</td>
<td>4</td>
<td>2</td>
<td>2**</td>
<td>4**</td>
<td>1**</td>
<td>6**</td>
<td>2</td>
<td>2**</td>
<td>2**</td>
<td>2</td>
<td>2</td>
<td>6**</td>
<td>6**</td>
</tr>
<tr>
<td>Blood counts</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>6</td>
<td>5</td>
<td>6*</td>
<td>6**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Renal (blood)</td>
<td>8**</td>
<td>10**</td>
<td>6</td>
<td>3**</td>
<td>2**</td>
<td>3**</td>
<td>7**</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>6**</td>
<td>6**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Renal (urine)</td>
<td>5**</td>
<td>2</td>
<td>7</td>
<td>3**</td>
<td>4**</td>
<td>1**</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>6**</td>
<td>6**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stored blood</td>
<td>8**</td>
<td>10**</td>
<td>3</td>
<td>9</td>
<td>3**</td>
<td>4**</td>
<td>3**</td>
<td>7**</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>6</td>
<td>5</td>
</tr>
<tr>
<td>GWAS, N</td>
<td>12527</td>
<td>3622</td>
<td>10000</td>
<td>8482</td>
<td>12670</td>
<td>3443</td>
<td>4287</td>
<td>1267</td>
<td>2500</td>
<td>3851</td>
<td>1418</td>
<td>5</td>
<td>2900</td>
<td>3476</td>
</tr>
<tr>
<td>WGS (T = TOPMed)</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
</tr>
<tr>
<td>RNA Seq-blood</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>RNA Seq-lung tissue</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DNA methylation</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Proteomics</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1**</td>
<td>2</td>
<td>1</td>
<td>1**</td>
<td>2</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Metabolomics</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
</tr>
</tbody>
</table>

*Maximum number of measures per participant. **Includes planned measurement at upcoming exam. A = Ascertained. T = on TOPMed. Data reflected in the table may be collected in only a subset of the cohort (e.g., as part of an Ancillary Study).
Table 3. Planned enrollment of C4R target population, estimated based on characteristics of current active cohort participants, 2020.

<table>
<thead>
<tr>
<th>C4R Cohort</th>
<th>N</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>53,972</td>
<td>100.00</td>
</tr>
<tr>
<td>Sex</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>31,034</td>
<td>57.50</td>
</tr>
<tr>
<td>Male</td>
<td>22,938</td>
<td>42.50</td>
</tr>
<tr>
<td>Race</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asian</td>
<td>1,255</td>
<td>2.33</td>
</tr>
<tr>
<td>American Indian and Alaska Native</td>
<td>3,015</td>
<td>5.59</td>
</tr>
<tr>
<td>Black</td>
<td>13,937</td>
<td>25.82</td>
</tr>
<tr>
<td>Native Hawaiian or Pacific Islander</td>
<td>32</td>
<td>0.06</td>
</tr>
<tr>
<td>White</td>
<td>29,104</td>
<td>53.92</td>
</tr>
<tr>
<td>Other</td>
<td>3,156</td>
<td>5.85</td>
</tr>
<tr>
<td>Multiple</td>
<td>2,037</td>
<td>3.77</td>
</tr>
<tr>
<td>Unknown</td>
<td>1,436</td>
<td>2.66</td>
</tr>
<tr>
<td>Ethnicity</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hispanic</td>
<td>10,862</td>
<td>20.12</td>
</tr>
<tr>
<td>Non-Hispanic</td>
<td>34,899</td>
<td>64.66</td>
</tr>
<tr>
<td>Unknown</td>
<td>8,211</td>
<td>15.21</td>
</tr>
<tr>
<td>Age group</td>
<td></td>
<td></td>
</tr>
<tr>
<td>18-29</td>
<td>663</td>
<td>1.23</td>
</tr>
<tr>
<td>30-64</td>
<td>21,920</td>
<td>40.61</td>
</tr>
<tr>
<td>65+</td>
<td>31,389</td>
<td>58.16</td>
</tr>
</tbody>
</table>
Table 4. Estimated number of participants with recent pre-pandemic deep phenotyping for harmonization in C4R, by cohort, 2010-2020. If the most recent exam was prior to 2010, data are not included. CT = computed tomography. ECG = electrocardiogram. GWAS = genome-wide association study. MRI = magnetic resonance imaging. Neurocog = neurocognitive.

<table>
<thead>
<tr>
<th>Physical function</th>
<th>ARIC</th>
<th>CARD-IA</th>
<th>COPD-Gene</th>
<th>HCHS/SOL</th>
<th>JHS</th>
<th>MAS-ALA</th>
<th>MESA</th>
<th>MESA</th>
<th>NOM-AS</th>
<th>PrePF</th>
<th>REGARDS</th>
<th>SARP</th>
<th>SPIROMICS</th>
<th>SHS</th>
<th>C4R</th>
</tr>
</thead>
<tbody>
<tr>
<td>Timed walk</td>
<td>3140</td>
<td>3879</td>
<td>2431</td>
<td>2473</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hand-grip</td>
<td>3140</td>
<td>1342</td>
<td>2257</td>
<td>628</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spirometry</td>
<td>3612</td>
<td>3119</td>
<td>4000</td>
<td>6232</td>
<td>9436</td>
<td>5006</td>
<td>3334</td>
<td>2517</td>
<td></td>
<td>1123</td>
<td>1123</td>
<td>380</td>
<td>1149</td>
<td>94</td>
<td>40002</td>
</tr>
<tr>
<td>DLCO</td>
<td>457</td>
<td>1342</td>
<td>2257</td>
<td>6232</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Resting O2</td>
<td>11</td>
<td>4000</td>
<td>5914</td>
<td>9195</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3751</td>
<td>3165</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CT Lung</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cardiac CT</td>
<td>2267</td>
<td>3068</td>
<td>2799</td>
<td>2944</td>
<td>1144</td>
<td>2643</td>
<td>2802</td>
<td></td>
<td></td>
<td>1059</td>
<td>229</td>
<td>1138</td>
<td>11889</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Any CT</td>
<td>2267</td>
<td>3068</td>
<td>3264</td>
<td>2944</td>
<td>1144</td>
<td>3371</td>
<td>3938</td>
<td></td>
<td></td>
<td>1059</td>
<td>229</td>
<td>1138</td>
<td>25221</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dual energy CT</td>
<td>2799</td>
<td></td>
<td>724</td>
<td>720</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>724</td>
<td>720</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cardiac</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cardiac MRI</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Echocardiogram</td>
<td>2973</td>
<td>3115</td>
<td>0</td>
<td>6427</td>
<td>5081</td>
<td>2849</td>
<td>2920</td>
<td>502</td>
<td></td>
<td>7778</td>
<td>67</td>
<td>6511</td>
<td>15846</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ECG</td>
<td>2409</td>
<td></td>
<td>0</td>
<td>9697</td>
<td>5289</td>
<td>892</td>
<td>3584</td>
<td>3803</td>
<td>502</td>
<td></td>
<td>1910</td>
<td>1910</td>
<td>35864</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Brain</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Brain MRI</td>
<td>771</td>
<td>653</td>
<td>1421</td>
<td>2466</td>
<td>2557</td>
<td>1059</td>
<td>229</td>
<td>1138</td>
<td></td>
<td>67</td>
<td>6511</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Neurocog-long</td>
<td>3589</td>
<td>3354</td>
<td>1245</td>
<td>1095</td>
<td>1117</td>
<td>803</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>817</td>
<td>31783</td>
<td>5684</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Neurocog-short</td>
<td>1360</td>
<td>2477</td>
<td>8400</td>
<td>1480</td>
<td>2000</td>
<td>600</td>
<td>528</td>
<td></td>
<td></td>
<td>817</td>
<td>30467</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Neurocog-any</td>
<td>3589</td>
<td>3354</td>
<td>3934</td>
<td>3652</td>
<td>4570</td>
<td>1260</td>
<td>8000</td>
<td></td>
<td>817</td>
<td>38936</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sleep+Activity</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Polysomnography</td>
<td>11</td>
<td>835</td>
<td>9195</td>
<td>913</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1718</td>
<td>1779</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Actimetry</td>
<td>513</td>
<td>1397</td>
<td>0</td>
<td>9012</td>
<td>852</td>
<td>1764</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ECG monitoring</td>
<td>2257</td>
<td></td>
<td>4100</td>
<td>1480</td>
<td>1557</td>
<td>300</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Biomarkers</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Blood</td>
<td>5046</td>
<td>4221</td>
<td>4000</td>
<td>7574</td>
<td>8400</td>
<td>2900</td>
<td>1144</td>
<td>3731</td>
<td>4683</td>
<td>1267</td>
<td>2500</td>
<td>8000</td>
<td>380</td>
<td>1800</td>
<td>2701</td>
</tr>
<tr>
<td>Urine</td>
<td>5046</td>
<td>4221</td>
<td>4000</td>
<td>7574</td>
<td>8400</td>
<td>2900</td>
<td>1144</td>
<td>3688</td>
<td>4683</td>
<td>1267</td>
<td>8000</td>
<td>1800</td>
<td>2701</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GWAS</td>
<td>4541</td>
<td>3799</td>
<td>4000</td>
<td>6817</td>
<td>12670</td>
<td>2610</td>
<td>3979</td>
<td>4215</td>
<td>2250</td>
<td></td>
<td>380</td>
<td>1620</td>
<td>46230</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RNaseq (blood)</td>
<td>800</td>
<td>2730</td>
<td>1040</td>
<td>2973</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>342</td>
<td>4027</td>
<td>11912</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Metabolomics</td>
<td>3025</td>
<td>8000</td>
<td>2750</td>
<td>787</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4000</td>
<td></td>
<td>20562</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Methylation</td>
<td>3799</td>
<td>4000</td>
<td>1900</td>
<td>3000</td>
<td>1752</td>
<td>1179</td>
<td>1830</td>
<td></td>
<td>2325</td>
<td></td>
<td>19346</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Proteomics</td>
<td>2813</td>
<td>1852</td>
<td>786</td>
<td>2000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>7451</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sputum/bronch</td>
<td>380</td>
<td>1000</td>
<td>1380</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gut microbiome</td>
<td>607</td>
<td>8000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Figure 1. Longitudinal pre-COVID follow up and planned follow-up of C4R participants, by cohort, 1971-2025. Some visits were overlapping, which is not shown; instead, midpoints of the visits are indicated. COVID-era exams are shaded in blue. Solid lines indicate cohort follow up, which typically includes regular contact by telephone and mail and ongoing events ascertainment.  

^424 gave restricted consent; ^b Includes 1,626 participants recruited from ARIC; ^c Withdrawal of consent by one participant; ^d MESA + 257 new recruits into the MESA Air Pollution Study.
Figure 2. C4R participants, field/clinical centers, and coordinating centers. Blue circles indicate field/clinical centers, and the size is proportional to the number of participants at that field/clinical center. Participants in the REGARDS, 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.
References

2. CDC COVID Data Tracker 2021 [Available from: https://covid.cdc.gov/covid-data-tracker/#cases_casesper100klast7days.


