UCSF
UC San Francisco Previously Published Works

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

Permalink
https://escholarship.org/uc/item/2d03t48x

Journal
American Journal of Epidemiology, 191(7)

ISSN
0002-9262

Authors
Oelsner, Elizabeth C
Krishnaswamy, Akshaya
Balte, Pallavi P
et al.

Publication Date
2022-06-27

DOI
10.1093/aje/kwac032

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

American Journal of Epidemiology **Submitted Manuscript**

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

Authors: Elizabeth C Oelsner, Akshaya Krishnaswamy, Pallavi P Balte, Norrina Bai Allen, Tauqeer Ali, Pramod Anugu, Howard 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 Pettee Gabriel, Linda Gallo, Jose Gutierrez, MeiLan 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 Kaplan, Gregory L Kinney, Anna Kucharska-Newton, Joyce S. Lee, Cora E Lewis, Deborah A. Levine, Emily B. Levitan, Bruce Levy, Barry Make, Kimberly Malloy, Jennifer J Manly, Carolina Mendoza-Puccini, Katie A Meyer, Yuan-I Min, Matthew Moll, Wendy C. Moore, Dave 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 Watson, Sally E. Wenzel, Karen Winters, Prescott G. Woodruff, Vanessa Xanthakis, Ying Zhang, Yiyi Zhang, For the C4R Investigators

ORIGINAL UNEDITED MANUSCRIPT

Published by Oxford University Press on behalf of the Johns Hopkins Bloomberg School of Public Health 2022. This work is written by (a) US Government employee(s) and is in the public domain in the US.

Correspondence Address: Elizabeth C Oelsner, MD MPH, Herbert Irving Associate Professor of Medicine, Division of General Medicine, Columbia University Irving Medical Center, 622 West 168th Street, PH9-105K New York, NY 10032 Email: eco7@cumc.columbia.edu Tel: 917-880-7099

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

ORIGINAL UNEDITED MANUSCRIPT

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** Division of Intramural Research, National Heart Lung and Blood Institute of the National Institutes of Health (Veronique Roger) Division of Cardiovascular Sciences, National Heart Lung and Blood Institute of the National Institutes of Health (Komal Arora, Sean Coady) Division of Lung Diseases, National Heart Lung and Blood Institute of the National Institutes of Health (Lisa Postow) National Institute of Neurological Disorders and Stroke of the National Institutes of Health (Carolina Mendoza-Puccini) **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

ORIGINAL UNEDITED MANUSCRIPT

(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, Arunee 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**

ORIGINAL UNEDITED MANUSCRIPT

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 Heath, 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

ORIGINAL UNEDITED MANUSCRIPT

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)

Funding: C4R is supported by NHLBI-CONNECTS OT2HL156812. 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,

ORIGINAL UNEDITED MANUSCRIPT

2U01HL096917from the NHLBI, NINDS, NIA and NIDCD. Ancillary studies funded additional data elements. 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). 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 & HHSN268201800007I), Northwestern University (HHSN268201800003I), University of Minnesota (HHSN268201800006I), and Kaiser Foundation Research Institute (HHSN268201800004I). This manuscript has been reviewed by CARDIA for scientific content. COPDGene was supported by Award Number U01 HL089897 and Award Number U01 HL089856 from the National Heart, Lung, and Blood Institute. 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.The Framingham Heart Study (FHS) acknowledges the support of Contracts NO1-HC-25195, HHSN268201500001I and 75N92019D00031 from the National Heart, Lung and Blood Institute. 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

ORIGINAL UNEDITED MANUSCRIPT

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, HHSN268201800011I 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, 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 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. Whole genome

ORIGINAL UNEDITED MANUSCRIPT

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 HHSN2682018000021) (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. NHLBI Pooled Cohorts Study was supported by grants NIH/NHLBI R21HL153700, K23HL130627, R21HL129924), R21HL121457.NOMAS was supported by NINDS R01 NS29993 and R01 AG066162.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.REGARDS 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 research by the principal and co-principal investigators of SARP were 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

ORIGINAL UNEDITED MANUSCRIPT

HL109168, U10 HL109152, U10 HL109086). Industry Partnerships also provided additional support: AstraZeneca, Boehringer-Ingelheim, Genetech, GlaxoSmithKline, MedImmune, Novartis, Regeneron, Sanofi and TEVA. Spirometers used in SARP III were provided by nSpire Health (Longmont, Colo).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. 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, & 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.

Data Availability Statement: Access to pooled C4R data is regulated by the C4R publications and presentations (P&P) policy, which is available on the C4R website. Data is made available for analyses on 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 is completed, C4R common data

ORIGINAL UNEDITED MANUSCRIPT

elements will be transferred as a limited dataset for public access on BioData Catalyst in accord with cohort-specific consents and commitments.

Thanks: We thank the participants of each cohort for their dedication to the studies. The authors thank the other investigators, the staff, and the participants of the REGARDS study for their valuable contributions. . We also acknowledge the dedication of the FHS study participants without whom this research would not be possible.

Conference presentation: None

Preprint: This manuscript was pre-published on medRxiv: https://www.medrxiv.org/content/10.1101/2021.03.19.21253986v1

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

Conflict of Interest: 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

ORIGINAL UNEDITED MANUSCRIPT

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.

Running Head: C4R Study Design

Key words: COVID-19, cohort study, epidemiology

ORIGINAL UNEDITED MANUSCRIPT

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
COVID-19	Coronavirus Disease 2019
DBS	Dried Blood Spot
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
SARS-CoV-2	Severe acute respiratory syndrome coronavirus 2
SPIROMICS	Subpopulations and Intermediate Outcome Measures in COPD Study
SHS	Strong Heart Study
W1Q	Wave 1 Questionnaire
W2Q	Wave 2 Questionnaire

ORIGINAL UNEDITED MANUSCRIPT

Abstract

The Collaborative Cohort of Cohorts for COVID-19 Research (C4R) is a national prospective study of adults comprising 14 established United States (US) prospective cohort studies.

Starting as early as 1971, C4R cohorts 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-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 reflects the racial, ethnic, socioeconomic, and geographic diversity of the US. 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 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 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 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.

Keywords: COVID-19, cohort study, epidemiology

Adverse effects of the coronavirus disease 2019 (COVID-19) pandemic on United States (US) health, economy, and society 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

ORIGINAL UNEDITED MANUSCRIPT

second-leading cause of death in those over 85 years of age (2, 3). Prolonged symptoms and clinical abnormalities are observed in some COVID-19 survivors, raising concerns that post-acute sequelae of SARS-CoV-2 infection (PASC) could pose an additional long-term health burden (4).

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. 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's longstanding 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, the C4R cohorts 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 post-acute 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

ORIGINAL UNEDITED MANUSCRIPT

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 cohorts are collaborating in C4R (**Table 1**). Eight of the cohorts were designed to study cardiovascular disease epidemiology: Atherosclerosis Risk in Communities (ARIC) Study (5), Coronary Artery Risk Development in Young Adults (CARDIA) Study (6), Framingham Heart Study (FHS) (7), Hispanic Community Health Study/Study of Latinos (HCHS/SOL) (8-10), Jackson Heart Study (JHS) (11-13), Mediators of Atherosclerosis in South Asians Living in America (MASALA) Study (14, 15), Multi-Ethnic Study of Atherosclerosis (MESA)

ORIGINAL UNEDITED MANUSCRIPT

(16), and the Strong Heart Study (SHS) (17, 18). 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 (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 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 (21). 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 (22) 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 (23).

These cohorts have collected detailed data on participants' health and behavior for as long as fifty years (**Figures 1, 2**). 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

ORIGINAL UNEDITED MANUSCRIPT

biorepositories of stored specimens (**Web Table 1**). 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 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 and governance

Most C4R cohorts have a history of collaboration in the genomics-oriented Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) consortium (24), the NHLBI Pooled Cohorts Study focusing on respiratory epidemiology (25), the Cross-Cohort Collaboration (CCC) for cardiovascular epidemiology (26), the Blood Pressure and Cognition (BP COG) Study (27), and the genetic sequencing and multi-omics-focused Trans-Omics for Precision Medicine (TOPMed) Project (28). 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

ORIGINAL UNEDITED MANUSCRIPT

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).

The collaborative governance structure for C4R is summarized in **Figure 3**. The administrative coordinating center (ACC) 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 (OSMB) convened by CONNECTS.

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) at Columbia University Irving Medical Center, PIs from the C4R Biorepository and Central Laboratory (BCL) at the University of Vermont, and program officers from the NHLBI, NINDS, and NIA. Master C4R protocols for COVID-19 data collection were drafted, piloted, and refined using a consensus-driven approach by dedicated sub-committees that included content-area experts from each cohort. Draft protocols were reviewed, refined, and approved by the organizing committee. Protocols were designed to provide flexibility for multi-modal data collection, including remote options that could be accomplished by both on-site and off-site study personnel, and are available at the study website (c4r-nih.org).

ORIGINAL UNEDITED MANUSCRIPT

Consistent with an ancillary studies model, each cohort is directly responsible for accomplishing its own data collection in accordance with the master protocols and under the supervision of its respective OSMB, Steering Committee, and any other applicable regulatory authorities. To maintain full participation, cohorts were permitted to adapt protocols to cohort-specific needs.

The BCL is responsible for establishing a C4R biorepository of dried blood spots (DBS), plus other biospecimens that may be collected in the future, and for performing and/or coordinating performance of centralized clinical and biomarker assays and serology assays.

Data collection is coordinated centrally at the DCHC. Electronic data collection forms are programmed into REDCap for use or adaptation by the cohort data coordinating centers (DCCs). Metadata on completion of questionnaires, events ascertainment, and DBS are reported to the DCHC bi-weekly. The DCHC maintains the C4R website, which includes a password-protected investigator section with regular status updates and study materials.

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

Participants

ORIGINAL UNEDITED MANUSCRIPT

Cohort participants previously consented for in-person, telephone, and/or email 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 1st, 2020 and had not withdrawn consent for cohort participation were considered eligible for enrollment into C4R. Of 72,358 participants who were believed to meet these inclusion criteria, the C4R PIs estimated that 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 socio-demographic 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 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 (29), 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 4**). 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 (23).

ORIGINAL UNEDITED MANUSCRIPT

COVID-19 questionnaires

Each cohort was funded to deploy COVID-19 questionnaires twice within 18 months following the initial outbreak in March 2020 via telephone, mail-in, online, email, or smartphone apps. Both the first questionnaire administration, called Wave 1, and the second questionnaire administration, called Wave 2, are attempted across the entire target population.

Various COVID-19 questionnaires were developed as early as March 2020 in certain cohorts (30) and administered in spring and summer 2020. Although these efforts pre-dated 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, re-infection, contacts, vaccination, behavioral changes, sleep, memory loss, depression, anxiety, fatigue, and resilience. This C4R questionnaire includes validated and PhenX toolkit instruments (31-40) in order to optimize comparability with pre-pandemic assessments and across C4R and other cohorts. The C4R questionnaire, including translations into Spanish and Mandarin, is available on PhenX and the C4R website; Research Electronic Data Capture (REDCap (41, 42)) programming is available on request. The C4R questionnaire was used by three cohorts in Wave

ORIGINAL UNEDITED MANUSCRIPT

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 are similar to the target population (**Table 2**).

COVID-related events

C4R ascertains 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 uses its own established infrastructure for ascertainment of medical records and death certificates, including the National Death Index, the Centers for Medicare & Medicaid Services, International Classification of Diseases codes (43), and linkage to records from local departments of health. Cohorts 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 to classify and validate cardiovascular, respiratory (44), and thromboembolic (45) events. Protocols for ascertainment, review, and classification are available on the study website. As of one year after C4R funding, over 1,000 COVID-related hospitalizations or deaths had been ascertained across the consortium.

ORIGINAL UNEDITED MANUSCRIPT

DBS collection

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). 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." 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. Participant instructions, including a video, are provided by the cohort and via the C4R website and/or cohort-specific websites. Participants mail the completed kits directly to the BCL or to the cohort field or coordinating center as an intermediary step. In cohorts with upcoming in-person exams, the DBS may be collected in-person by research staff. DBS deployment was initiated in February 2021 (**Web Figure 1**) and is ongoing, with over 10,000 DBS completed as of September 30, 2021.

Serology

After pre-processing of completed DBS cards by the BCL, serology assays are performed by the New York State Wadsworth Center's Bloodborne Viruses Laboratory (BVL) under CLIA and New

ORIGINAL UNEDITED MANUSCRIPT

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 2020 as part of New York State's public health response. Serology results are reported by the BVL to the BCL, and then to the cohort DCCs, 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 known to have clinical relevance, and the CDC does not currently recommend modifications to individual behavior or clinical care based on antibody status alone (46); 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 U.S. generate an immune response to the Spike protein, antibody responses to vaccination versus viral infection may be distinguished by the anti-nucleocapsid assay results (47).

Quality control

ORIGINAL UNEDITED MANUSCRIPT

Cohorts 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% sub-sample of blind duplicates.

COVID-19 outcomes

C4R data defines 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 post-acute setting, supporting classification of symptomatic and asymptomatic infections and cases of prolonged recovery or PASC. 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 sero-reverted.

ORIGINAL UNEDITED MANUSCRIPT

Harmonization

Harmonization of COVID-19 and pre-pandemic data are performed centrally by the DCHC on the C4R Analysis Commons to define COVID-19 outcomes and to align pre-pandemic 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 BP COG Study, and the CHARGE Working Groups (24, 25, 27, 48-54). 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 data needs of approved manuscript proposals. As in prior published efforts (25, 55), 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 DCCs, 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 DCCs. Harmonized and derived variables, plus codes used to generate them, are shared with the source cohorts.

Due to their significance to COVID-19 epidemiology, particular emphasis is being placed on harmonizing pre-pandemic physiologic (25), neurocognitive (27, 56-61), and imaging-based (62-

ORIGINAL UNEDITED MANUSCRIPT

69) phenotyping collected within the decade prior to the outbreak (**Table 3**). Harmonization of lung computed tomography is being accomplished using deep learning (70-73) and other methods, which will be published in separate reports.

Data management

The C4R Commons Agreement, modeled on the CHARGE Analysis Commons Consortium Agreement (74), is expediting cross-cohort data harmonization and sharing, as allowed (75). Following review and approval, cohort-specific agreements permit COVID-19 and pre-pandemic data to be uploaded to the C4R Analysis Commons, which is located on the NHLBI cloud computing platform, BioData Catalyst powered by Seven Bridges. This platform has enterprise-grade compliance and security certification and is HIPAA-compliant and compliant with dbGaP and CLIA security best practices. Nonetheless, no Protected Health Information (PHI) is retained on the C4R Analysis Commons. Participants are assigned a C4R study identifier by the cohort-specific DCC 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 (P&P) policy, which is available on the C4R website. Data is made available for analyses on 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

ORIGINAL UNEDITED MANUSCRIPT

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 Sub-Committee and related programming is shareable on the C4R Analysis Commons. 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.

DISCUSSION

C4R leverages existing American cohort studies to develop a large, multi-ethnic, 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 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.

ORIGINAL UNEDITED MANUSCRIPT

C4R constitutes a historic initiative to standardize and expedite data collection by US cohort studies despite major operational and societal challenges. Commitment from cohort PIs with experience in cross-cohort collaboration was critical to expediting study design, protocol development, approval, and deployment. The large scale of data collection was made possible not only by a culture shift toward multi-study collaborative efforts occasioned by experience in CHARGE and TOPMed, but also by the established cohort infrastructure, including sophisticated DCCs that were prepared to implement new protocols rapidly, and experienced and professional clinical staff with longstanding relationships with cohort participants. Differences in data collected by cohort are being reconciled by 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 cohorts 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 data sharing across fourteen cohorts are well documented (76) and have been surmounted in recent decades by several strategies, including meta-analysis of results generated by cohort-specific DCCs and analyses of pooled cohort data assembled at a single center or on 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 sampled to be representative of the US population; to some extent, this may be addressed

ORIGINAL UNEDITED MANUSCRIPT

analytically by weighting approaches. Participation in C4R is voluntary, which may lead to selection biases. Nonetheless, we do not observe major differences in the socio-demographic characteristics of eligible participants compared to 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 is sparse in C4R compared to certain EHR resources, but, conversely, data on pre-COVID biomarkers, multi-omics, physiology, organ structure, and symptomatology may be richer in C4R and less subject to diagnostic or referral biases. Due to antibody waning, only a subset of self-reported COVID-19 in C4R will be validated by C4R serology, and negative serology will not be suitable to rule out infection. Nonetheless, serology provides opportunities to reclassify some participants with subclinical infection and to examine antibody responses to vaccination and the impact on subsequent "breakthrough" infection risk. Events adjudication is primarily validating more severe illness, although many cohorts are ascertaining positive test results among non-hospitalized 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 measures (e.g., echocardiography, lung imaging, neuro-cognitive assessment) to define reliably 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, metabo-types, 'Omics, and other

ORIGINAL UNEDITED MANUSCRIPT

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 strategic transformation of population studies (77).

ACKNOWLEDGEMENTS

Members of the study group

Tauqueer Ali, Nori Allen, Howard Andrews, Pramod Anugu, Komal Arora, Alex Arynchyn, Alyssa Asaro, Pallavi P. Balte, Mike Bancks, R. Graham Barr, Lori Bateman, Alain Bertoni, Alexander G. Bick, Eugene Bleecker, David Bluemke, Rebekah Boyle, Matt Budoff, Jonathan Cardwell, Jeffrey Carr, Jackie Causey, Ann Chang, Grace Chen, Sean Coady, Shelley Cole, Josef Coresh, Elaine Cornell, Adolfo Correa, David Couper, Emily Culbertson, Jeffrey L. Curtis, Mary Cushman, Charles DeCarli, Ryan T. Demmer, Richard B. Devereux, Nicholas DiPersio, Brett Dolezal, Margaret Doyle, Mitch Elkind, Sean Fain, Matt Feinstein, James Floyd, Christine Freeman, Amanda Fretts, Kelley Gabriel, Linda Gallo, Sina Gharib, Caroline Goode, Parag Goyal, Mike Griswold, MeiLan Han, David Hanna, Annette Hastie, Susan R. Heckbert, Joel Henderson, Karen D. Hinckley Stukovsky, Eric Hoffman, Udo Hoffman, Mady Hornig, Virginia J. Howard, Carmen R. Isasi, David R. Jacobs, Cashell Jaquis, Suzanne E. Judd, Debora Kamin Mukaz, Alka M. Kanaya, Namratha R. Kandula, Robert Kaplan, Sadiya Khan, Jonathan King, Gregory L. Kinney, Anna Kucharska-Newton, Smith Kumi, Andrew F. Laine, Joyce Lee, Rozenn Lemaitre, Bonnie Levin, Deborah Levine, Emily B. Levitan, Bruce Levy, Joao Lima, David Lynch, Barry Make, Kimberly Malloy, Jennifer J. Manly, Dave Mauger, Katy Melius, Carolina Mendoza-Puccini, Sharon Merkin, Katie Meyer, Deborah Meyers, Yuan-I (Nancy) Min, Melissa Minotti, Nathalie Moise, Wendy Moore, Claudia Moy, Karen Mutalik, Ilya Nasrallah, Cheryl Nelson, Lauren Nelson, Patricia Noel, Anna Nordvig, George O'Connor, Michelle Odden, Elizabeth Oelsner, Marcia O'Leary, Wanda O'Neal, Victor Ortega, Priya Palta, Nathalie Pamir, George Papanicolaou, Brenda Phillips, Wanda Phipatanakul, Tim Plante, Yashashwi Pokharel, Wendy S. Post, Lisa Postow, Bruce Psaty, Shaun M. Purcell, Laura Raffield, Vasan Ramachandran, Susan Redline, Liz Regan, Alex Reiner, Carlos

ORIGINAL UNEDITED MANUSCRIPT

Rodriguez, Veronique Roger, Jerry Rotter, Ralph Sacco, Monika M. Safford, Michael Schembri, David Schwartz, Sudha Seshadri, Amil Shah, Sanjiv J. Shah, James Shikany, Mario Sims, Kumi Smith, Sylvia Smoller, Elsayed Soliman, Scott Solomon, Daniela Sotres-Alvarez, Meg Stewart, Kevin Strobino, Patrice Sutherland, Bonnie Swenor, Gregory Talavera, Greg Terry, Russ Tracy, Janis Tse, Jason Umans, Ramachandran S. Vasan, Louis Vuga, Molly Wagster, Henry Wang, George Washko, Sally Wentzel, Cynthia West, Carla Wilson, Prescott Woodruff, Jackie Wright, Vanessa Xanthakis, Ya Yuan, Neil Zakai, Ying Zhang, Yiyi Zhang.

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

References

1. Kissler SM, Tedijanto C, Goldstein E, Grad YH, Lipsitch M. Projecting the transmission dynamics of SARS-CoV-2 through the postpandemic period. Science. 2020.
2. Woolf SH, Chapman DA, Lee JH. COVID-19 as the Leading Cause of Death in the United States. JAMA. 2021;325(2):123-4.
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).
4. Huang C, Huang L, Wang Y, Li X, Ren L, Gu X, et al. 6-month consequences of COVID-19 in patients discharged from hospital: a cohort study. The Lancet. 2021;397(10270):220-32.
5. The Atherosclerosis Risk in Communities (ARIC) Study: design and objectives. The ARIC investigators. Am J Epidemiol. 1989;129(4):687-702.
6. Friedman GD, Cutter GR, Donahue RP, Hughes GH, Hulley SB, Jacobs DR, Jr., et al. CARDIA: study design, recruitment, and some characteristics of the examined subjects. J Clin Epidemiol. 1988;41(11):1105-16.
7. Tsao CW, Vasan RS. Cohort Profile: The Framingham Heart Study (FHS): overview of milestones in cardiovascular epidemiology. Int J Epidemiol. 2015;44(6):1800-13.
8. Daviglus ML, Talavera GA, Aviles-Santa ML, Allison M, Cai J, Criqui MH, 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):1775-84.
9. Lavange LM, Kalsbeek WD, Sorlie PD, Aviles-Santa LM, Kaplan RC, Barnhart J, et al. Sample design and cohort selection in the Hispanic Community Health Study/Study of Latinos. Ann Epidemiol. 2010;20(8):642-9.
10. Sorlie PD, Aviles-Santa LM, Wassertheil-Smoller S, Kaplan RC, Daviglus ML, Giachello AL, et al. Design and implementation of the Hispanic Community Health Study/Study of Latinos. Ann Epidemiol. 2010;20(8):629-41.
11. Carpenter MA, Crow R, Steffes M, Rock W, Heilbraun J, Evans G, et al. Laboratory, reading center, and coordinating center data management methods in the Jackson Heart Study. Am J Med Sci. 2004;328(3):131-44.

ORIGINAL UNEDITED MANUSCRIPT

12. Keku E, Rosamond W, Taylor HA, Jr., Garrison R, Wyatt SB, Richard M, et al. Cardiovascular disease event classification in the Jackson Heart Study: methods and procedures. Ethn Dis. 2005;15(4 Suppl 6):S6-62-70.
13. Taylor HA, Jr., Wilson JG, Jones DW, Sarpong DF, Srinivasan A, Garrison RJ, 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):S6-4-17.
14. Kanaya AM, Chang A, Schembri M, Puri-Taneja A, Srivastava S, Dave SS, 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, Budoff MJ, Hulley S, Vittinghoff E, et al. Mediators of Atherosclerosis in South Asians Living in America (MASALA) study: objectives, methods, and cohort description. Clin Cardiol. 2013;36(12):713-20.
16. Bild DE, Bluemke DA, Burke GL, Detrano R, Diez Roux AV, Folsom AR, et al. Multi-Ethnic Study of Atherosclerosis: objectives and design. Am J Epidemiol. 2002;156(9):871-81.
17. Lee ET, Welty TK, Fabsitz R, Cowan LD, Le NA, Oopik AJ, et al. The Strong Heart Study. A study of cardiovascular disease in American Indians: design and methods. Am J Epidemiol. 1990;132(6):1141-55.
18. North KE, Howard BV, Welty TK, Best LG, Lee ET, Yeh JL, et al. Genetic and environmental contributions to cardiovascular disease risk in American Indians: the strong heart family study. Am J Epidemiol. 2003;157(4):303-14.
19. Regan EA, Hokanson JE, Murphy JR, Make B, Lynch DA, Beaty TH, et al. Genetic epidemiology of COPD (COPDGene) study design. COPD. 2010;7(1):32-43.
20. Couper D, LaVange LM, Han M, Barr RG, Bleecker E, Hoffman EA, et al. Design of the Subpopulations and Intermediate Outcomes in COPD Study (SPIROMICS). Thorax. 2014;69(5):491-4.
21. Teague WG, Phillips BR, Fahy JV, Wenzel SE, Fitzpatrick AM, Moore WC, et al. Baseline Features of the Severe Asthma Research Program (SARP III) Cohort: Differences with Age. J Allergy Clin Immunol Pract. 2018;6(2):545-54 e4.
22. Sacco RL, Boden-Albala B, Gan R, Chen X, Kargman DE, Shea S, 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):259-68.
23. Howard VJ, Cushman M, Pulley L, Gomez CR, Go RC, Prineas RJ, et al. The reasons for geographic and racial differences in stroke study: objectives and design. Neuroepidemiology. 2005;25(3):135-43.
24. Psaty BM, O'Donnell CJ, Gudnason V, Lunetta KL, Folsom AR, Rotter JI, 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.
25. Oelsner EC, Balte PP, Cassano PA, Couper D, Enright PL, Folsom AR, et al. Harmonization of Respiratory Data From 9 US Population-Based Cohorts: The NHLBI Pooled Cohorts Study. Am J Epidemiol. 2018;187(11):2265-78.
26. Cross Cohort Collaboration [updated May 10, 2021. Available from: https://chs-nhlbi.org/node/6539.

ORIGINAL UNEDITED MANUSCRIPT

27. Levine DA, Gross AL, Briceno EM, Tilton N, Kabeto MU, Hingtgen SM, et al. Association Between Blood Pressure and Later-Life Cognition Among Black and White Individuals. JAMA Neurol. 2020;77(7):810-9.
28. Taliun D, Harris DN, Kessler MD, Carlson J, Szpiech ZA, Torres R, et al. Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. Nature. 2021;590(7845):290-9.
29. Flanagin A, Frey T, Christiansen SL, Bauchner H. The Reporting of Race and Ethnicity in Medical and Science Journals: Comments Invited. JAMA. 2021.
30. Multi-Ethnic Study of Atherosclerosis. Oelsner MESA COVID-19 Questionnaire 2020. https://www.phenxtoolkit.org/toolkit_content/PDF/MESA_Questionnaire_Annotated.pdf. Accessed October 13, 2021.
31. Behavioral Risk Factor Surveillance System. Questionnaire 2019. https://www.cdc.gov/brfss/questionnaires/. Updated August 26, 2021. Accessed October 13, 2021.
32. Health and Retirement Survey. HRS Biennial Interview Questionnaires 2020. https://hrs.isr.umich.edu/documentation/questionnaires. Accessed October 13, 2021.
33. MACS-WIHS Combined Cohort Study. BASELINE COVID-19 ABBREVIATED QUESTIONNAIRE (BLCOVID) 2020. https://www.nlm.nih.gov/dr2/MACS-WIHS_questionnaire_BLCOVID-040620.pdf. Updates April 6, 2020. Accessed October 13, 2021.
34. Andresen EM, Malmgren JA, Carter WB, Patrick DL. 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.
35. Cohen S, Kamarck T, Mermelstein R. A global measure of perceived stress. J Health Soc Behav. 1983;24(4):385-96.
36. Levine DW, Kripke DF, Kaplan RM, Lewis MA, Naughton MJ, Bowen DJ, et al. Reliability and validity of the Women's Health Initiative Insomnia Rating Scale. Psychol Assess. 2003;15(2):137-48.
37. Pilkonis PA, Choi SW, Reise SP, Stover AM, Riley WT, Cella D, et al. Item banks for measuring emotional distress from the Patient-Reported Outcomes Measurement Information System (PROMIS(R)): depression, anxiety, and anger. Assessment. 2011;18(3):263-83.
38. RAND Corporation. Social Support Survey Instrument. https://www.rand.org/health-care/surveys_tools/mos/social-support/survey-instrument.html. Accessed October 13, 2021.
39. Russell DW. UCLA Loneliness Scale (Version 3): reliability, validity, and factor structure. J Pers Assess. 1996;66(1):20-40.
40. Smith BW, Dalen J, Wiggins K, Tooley E, Christopher P, Bernard J. The brief resilience scale: assessing the ability to bounce back. Int J Behav Med. 2008;15(3):194-200.
41. Harris PA, Taylor R, Minor BL, Elliott V, Fernandez M, O'Neal L, et al. The REDCap consortium: Building an international community of software platform partners. J Biomed Inform. 2019;95:103208.
42. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform. 2009;42(2):377-81.
43. Centers for Disease Control and Prevention. ICD-10-CM Official Coding and Reporting Guidelines April 1, 2020 through September 30, 2020. Accessed October 13, 2021.

ORIGINAL UNEDITED MANUSCRIPT

44. Oelsner EC, Loehr LR, Henderson AG, Donohue KM, Enright PL, Kalhan R, et al. Classifying Chronic Lower Respiratory Disease Events in Epidemiologic Cohort Studies. Ann Am Thorac Soc. 2016;13(7):1057-66.
45. Zakai NA, McClure LA, Judd SE, Safford MM, Folsom AR, Lutsey PL, et al. Racial and regional differences in venous thromboembolism in the United States in 3 cohorts. Circulation. 2014;129(14):1502-9.
46. 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. Accessed October 13, 2021.
47. Laing ED, Sterling SL, Richard SA, Epsi NJ, Coggins S, Samuels EC, et al. Antigen-based multiplex strategies to discriminate SARS-CoV-2 natural and vaccine induced immunity from seasonal human coronavirus humoral responses. medRxiv. 2021.
48. Balte PP, Chaves PHM, Couper DJ, Enright P, Jacobs DR, Jr., Kalhan R, et al. Association of Nonobstructive Chronic Bronchitis With Respiratory Health Outcomes in Adults. JAMA Intern Med. 2020.
49. Bhatt SP, Balte PP, Schwartz JE, Cassano PA, Couper D, Jacobs DR, Jr., et al. Discriminative Accuracy of FEV1:FVC Thresholds for COPD-Related Hospitalization and Mortality. JAMA. 2019;321(24):2438-47.
50. Bhatt SP, Schwartz JE, Oelsner EC. FEV1:FVC Thresholds for Defining Chronic Obstructive Pulmonary Disease-Reply. JAMA. 2019;322(16):1611-2.
51. Cornelius T BP, Bhatt SP, Cassano PA, Currow D, Jacobs DR, Johnson M, Kalhan R, Kronmal R, Loehr LR, O'Connor GT, Schwartz JE, Smith BM, White WB, Yende S, Oelsner EC. . A dyadic growth modeling approach to weight gain and lung function loss: The NHLBI Pooled Cohorts Study. . Am J Epidemiol. 2020 (Accepted).
52. Oelsner EC, Balte PP, Bhatt SP, Cassano PA, Couper D, Folsom AR, 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.
53. Zhang Y, Vittinghoff E, Pletcher MJ, Allen NB, Zeki Al Hazzouri A, Yaffe K, et al. Associations of Blood Pressure and Cholesterol Levels During Young Adulthood With Later Cardiovascular Events. J Am Coll Cardiol. 2019;74(3):330-41.
54. Oelsner EC, Balte PP, Grams ME, Cassano PA, Jacobs DR, Barr RG, 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):321-32.
55. Stilp AM, Emery LS, Broome JG, Buth EJ, Khan AT, Laurie CA, et al. A System for Phenotype Harmonization in the NHLBI Trans-Omics for Precision Medicine (TOPMed) Program. Am J Epidemiol. 2021.
56. Griffith L, van den Heuvel E, Fortier I, Hofer S, Raina P, Sohel N, Payette H, Wolfson C, Belleville S. Harmonization of Cognitive Measures in Individual Participant Data and Aggregate Data Meta-Analysis. Methods Research Report. (Prepared by the McMaster University Evidence-based Practice Center under Contract No. 290-2007-10060-I.) AHRQ Publication No.13-EHC040-EF. Rockville, MD: Agency for Healthcare Research and Quality; March 2013. www.effectivehealthcare.ahrq.gov/reports/final.cfm. .

ORIGINAL UNEDITED MANUSCRIPT

57. Griffith LE, van den Heuvel E, Fortier I, Sohel N, Hofer SM, Payette H, et al. Statistical approaches to harmonize data on cognitive measures in systematic reviews are rarely reported. J Clin Epidemiol. 2015;68(2):154-62.
58. Gross AL, Jones RN, Fong TG, Tommet D, Inouye SK. Calibration and validation of an innovative approach for estimating general cognitive performance. Neuroepidemiology. 2014;42(3):144-53.
59. Gross AL, Sherva R, Mukherjee S, Newhouse S, Kauwe JS, Munsie LM, et al. Calibrating longitudinal cognition in Alzheimer's disease across diverse test batteries and datasets. Neuroepidemiology. 2014;43(3-4):194-205.
60. Gross AL, Mungas DM, Crane PK, Gibbons LE, MacKay-Brandt A, Manly JJ, et al. Effects of education and race on cognitive decline: An integrative study of generalizability versus study-specific results. Psychol Aging. 2015;30(4):863-80.
61. Langa KM, Ryan LH, McCammon RJ, Jones RN, Manly JJ, Levine DA, et al. The Health and Retirement Study Harmonized Cognitive Assessment Protocol Project: Study Design and Methods. Neuroepidemiology. 2020;54(1):64-74.
62. Oelsner EC, Ortega VE, Smith BM, Nguyen JN, Manichaikul AW, Hoffman EA, 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):721-31.
63. Smith BM, Kirby M, Hoffman EA, Kronmal RA, Aaron SD, Allen NB, et al. Association of Dysanapsis With Chronic Obstructive Pulmonary Disease Among Older Adults. JAMA. 2020;323(22):2268-80.
64. Smith BM, Hoffman EA, Rabinowitz D, Bleecker E, Christenson S, Couper 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):987-96.
65. Hame Y, Angelini E, Hoffman E, Barr G, Laine A. Adaptive quantification and longitudinal analysis of pulmonary emphysema with a hidden Markov measure field model. IEEE Transactions on Medical Imaging. 2014;33(7):1527-40.
66. Wang M, Aaron CP, Madrigano J, Hoffman EA, Angelini E, Yang 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):546-56.
67. Yang J, Angelini E, Smith BM, Balte P, Hoffman E, Barr RG, et al. Unsupervised Discovery of Spatially- Informed Lung Texture Patterns for Pulmonary Emphysema: The MESA COPD Study. International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)2017. p. 116-24.
68. Yang J, Angelini ED, Balte PP, Hoffman EA, Wu CO, Venkatesh BA, et al. Emphysema quantification on cardiac CT scans using hidden Markov measure field model: The MESA Lung Study. International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI); Athenes, Greece: Springer; 2016. p. 624-31.
69. Yang J, Vetterli T, Balte PP, Barr RG, Laine AF, Angelini ED. Unsupervised Domain Adaption With Adversarial Learning (UDAA) for Emphysema Subtyping on Cardiac CT Scans: The Mesa Study. IEEE International Symposium on Biomedical Imaging (ISBI)2019. p. 289-93.
70. Ravi D, Wong C, Deligianni F, Berthelot M, Andreu-Perez J, Lo B, et al. Deep Learning for Health Informatics. IEEE J Biomed Health Inform. 2017;21(1):4-21.

ORIGINAL UNEDITED MANUSCRIPT

71. Pham T TT, Phung D and Venkatesh S. . Deepcare: A deep dynamic memory model for predictive medicine. Paper presented at: Pacific-Asia Conference on Knowledge Discovery and Data Mining; 2016.
72. Miotto R, Li L, Kidd BA, Dudley JT. Deep Patient: An Unsupervised Representation to Predict the Future of Patients from the Electronic Health Records. Sci Rep. 2016;6:26094.
73. He K, ZHang X, Ren S, Sun J. (2015-12-10). "Deep Residual Learning for Image Recognition". arXiv:1512.03385.
74. Brody JA, Morrison AC, Bis JC, O'Connell JR, Brown MR, Huffman JE, et al. Analysis commons, a team approach to discovery in a big-data environment for genetic epidemiology. Nat Genet. 2017;49(11):1560-3.
75. Health NIo. NIH Tribal Consultation Report: NIH Draft Policy for Data Management and Sharing. 2020.
76. Psaty BM, Rich SS, Boerwinkle E. Innovation in Genomic Data Sharing at the NIH. N Engl J Med. 2019;380(23):2192-5.
77. Roger VL, Boerwinkle E, Crapo JD, Douglas PS, Epstein JA, Granger CB, 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):363-8.

ORIGINAL UNEDITED MANUSCRIPT

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

Cohort	N	Current age range, years	Sex, % Female	Race/ethnicity, % NHW	B	H/L	As	Am Ind	Other	Original research focus
ARIC	6,690	75-97	63	77	23	0[a]	0	0	0	Cardiovascular
CARDIA	4,590	53-66	56	50	50	0	0	0	0	Cardiovascular
COPDGene	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	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	Neurologic
PrePF	5,000	40-80	55	92	3	3	0	0	0	Pulmonary
REGARDS	12,766	57-105	58	62	38	0	0	0	0	Neurologic
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

Am Ind = American Indian; As = Asian American; B = Black; H/L = Hispanic/Latinx; NHW = Non-Hispanic White.

ORIGINAL UNEDITED MANUSCRIPT

ARIC = Atherosclerosis Risk in Communities Study; C4R = Collaborative Cohort of Cohorts for COVID-19 Research; CARDIA = Coronary Artery Risk Development in Young Adults; COPDGene= Genetic Epidemiology of COPD; FHS = Framingham Heart Study; HCHS/SOL = Hispanic Community Health Study/Study of Latinos; JHS = Jackson Heart Study; MASALA = Mediators of Atherosclerosis in South Asians Living in America; MESA = Multi-Ethnic Study of Atherosclerosis; NOMAS = Northern Manhattan Study; PrePF = Prevent Pulmonary Fibrosis; REGARDS = REasons for Geographic and Racial Differences in Stroke; SARP = Severe Asthma Research Program; SPIROMICS = Subpopulations and Intermediate Outcome Measures in COPD Study; SHS = Strong Heart Study

[a] ARIC did not inquire regarding Hispanic/Latino ethnicity, hence White participants cannot be definitely defined as non-Hispanic.

ORIGINAL UNEDITED MANUSCRIPT

Table 2. Characteristics of the C4R target population compared to participants completing the C4R Wave 1 Questionnaire, March 1, 2020 – September 30, 2021.

	Target population (n=53,143)		Participants completing C4R wave 1 questionnaires through September 30, 2021 (n=45,262)	
Cohort	N	%	N	%
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 and Alaska Native	2,971	5.59	1,923	4.25
Black	13,722	25.82	9,860	21.78
Native Hawaiian or 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
Multiple	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				
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

ORIGINAL UNEDITED MANUSCRIPT

Cohort				
ARIC	5,046	9.50	5,466	12.08
CARDIA	4,221	7.94	2,530	5.59
COPDGene	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	500	0.94	460	1.02
MESA	4,683	8.81	3,450	7.62
NOMAS	1,256	2.36	887	1.96
PrePF	2,500	4.70	614	1.36
REGARDS	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

C4R = Collaborative Cohort of Cohorts for COVID-19 Research. 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

ORIGINAL UNEDITED MANUSCRIPT

Table 3. Estimated number of participants with recent pre-pandemic deep phenotyping for harmonization in C4R, by cohort, United States, 2010-2020.[a]

Measures	ARIC	CARDIA	COPDGene	FHS	HCHS/SOL	JHS	MASALA	MESA	NOMAS	PrePF	REGARDS	SARP	SPIROMICS	SHS	C4R
Physical function															
Timed walk	3,140		3,879	2,257				2,473					1,064		12,813
Hand-grip	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 O2	11		4,000		8,400			3,973					1,069		17,453
CT															
CT Lung			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															
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															
Brain MRI	771	653		3,543	1,245			1,113	803						8,128
Neurocog-long	3,589	3,354		2,477	8,400			1,487	600		528			817	21,252
Neurocog-short			1,360	3,934	8,400			3,788	1,260		8,000			817	27,559
Neurocog-any	3,589	3,354	1,360	3,800	8,400			3,857	1,260		8,000			817	34,437
Sleep+Activity															
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
Biomarkers															
Blood	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	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
RNAseq (blood)			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/bronch												380	1,000		1,380
Gut microbiome		607			8,000										8,607

ORIGINAL UNEDITED MANUSCRIPT

Bronch = bronchoscopy; CT = computed tomography; DLCO = diffusing capacity of the lung for carbon monoxide; ECG = electrocardiogram; GWAS = genome-wide association study; MRI = magnetic resonance imaging; neurocog = neurocognitive; O2 = oxygen;
ARIC = Atherosclerosis Risk in Communities Study; C4R = Collaborative Cohort of Cohorts for COVID-19 Research; CARDIA = Coronary Artery Risk Development in Young Adults; COPDGene= Genetic Epidemiology of COPD; FHS = Framingham Heart Study; HCHS/SOL = Hispanic Community Health Study/Study of Latinos; JHS = Jackson Heart Study; MASALA = Mediators of Atherosclerosis in South Asians Living in America; MESA = Multi-Ethnic Study of Atherosclerosis; NOMAS = Northern Manhattan Study; PrePF = Prevent Pulmonary Fibrosis; REGARDS = REasons for Geographic and Racial Differences in Stroke; SARP = Severe Asthma Research Program; SPIROMICS = Subpopulations and Intermediate Outcome Measures in COPD Study; SHS = Strong Heart Study

[a] If the most recent exam was prior to 2010, data are not included.

ORIGINAL UNEDITED MANUSCRIPT

Figure Legends

Figure 1. Longitudinal pre-COVID follow-up and planned follow-up for cardiovascular cohorts collaborating in C4R, by cohort, United States, 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. ARIC = Atherosclerosis Risk in Communities Study; CARDIA = Coronary Artery Risk Development in Young Adults; FHS-Gen3 = Framingham Heart Study Third Generation; FHS-O = Framingham Offspring 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; SHS = Strong Heart Study; Exam = examination. In ARIC, 424 gave restricted consent; JHS[i] 1,626 participants recruited from ARIC; CARDIA includes withdrawal of consent by one participant; MESA is MESA + 257 new recruits into the MESA Air Pollution Study.

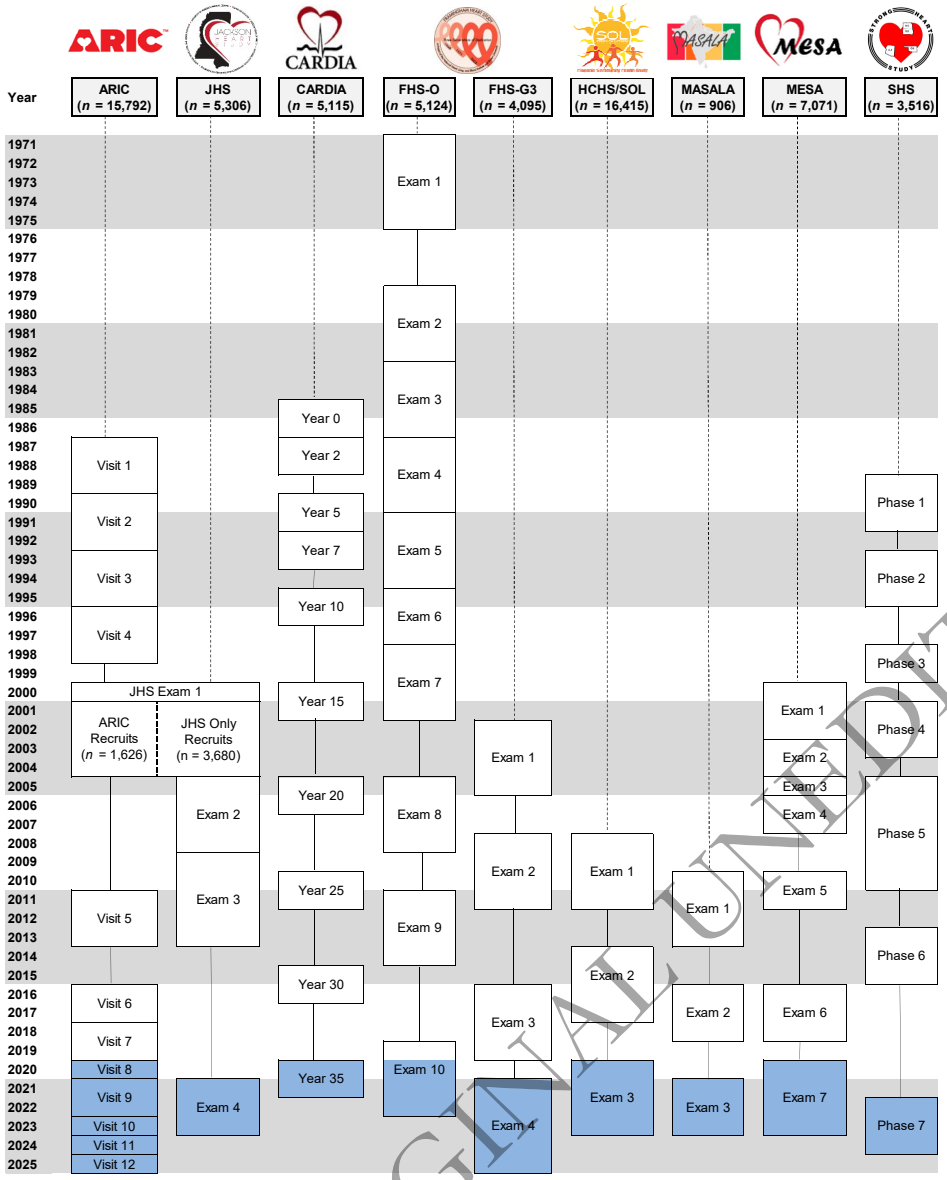
Figure 2. Longitudinal pre-COVID follow-up and planned follow-up for pulmonary and neurological cohorts in C4R, by cohort, United States, 1991-2025. Some visits were overlapping, which is not shown; instead, midpoints of the visits are indicated. COVID-era exams are shaded in blue. Solid lines indicate cohort follow up, which typically includes regular contact by telephone and mail and ongoing events ascertainment. COPDGene= Genetic Epidemiology of COPD; Exam = examination; 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.

Figure 3. C4R organizational chart. CONNECTS = Collaborating Network of Networks for Evaluating COVID-19 and Therapeutic Strategies. DSMB = data safety monitoring board. OSMB = observational studies monitoring board. 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

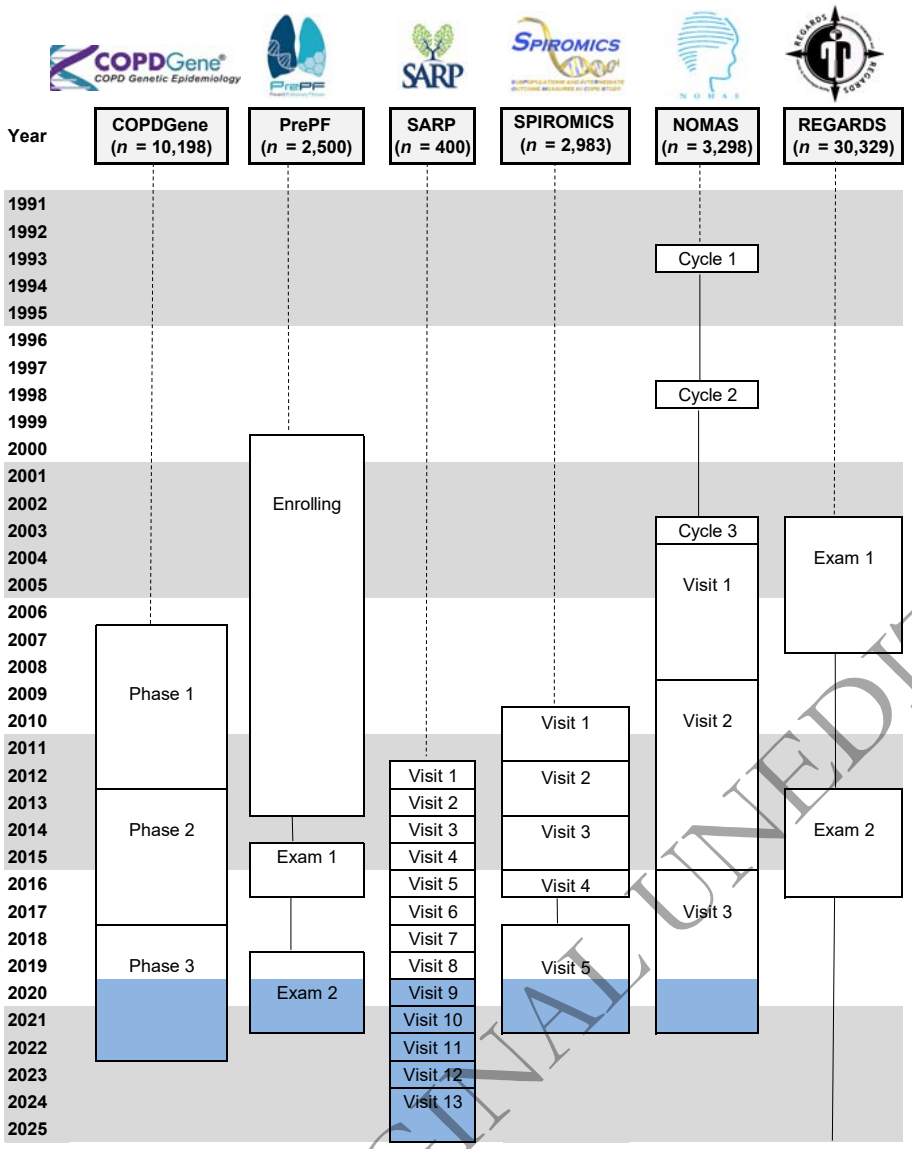
Figure 4. C4R participants, field/clinical centers, and coordinating centers. Blue circles indicate field/clinical centers, and the size is proportional to the number of participants at that field/clinical center. Participants in the REGARDS, which does not have field/clinical centers, are shown by additional 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.

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

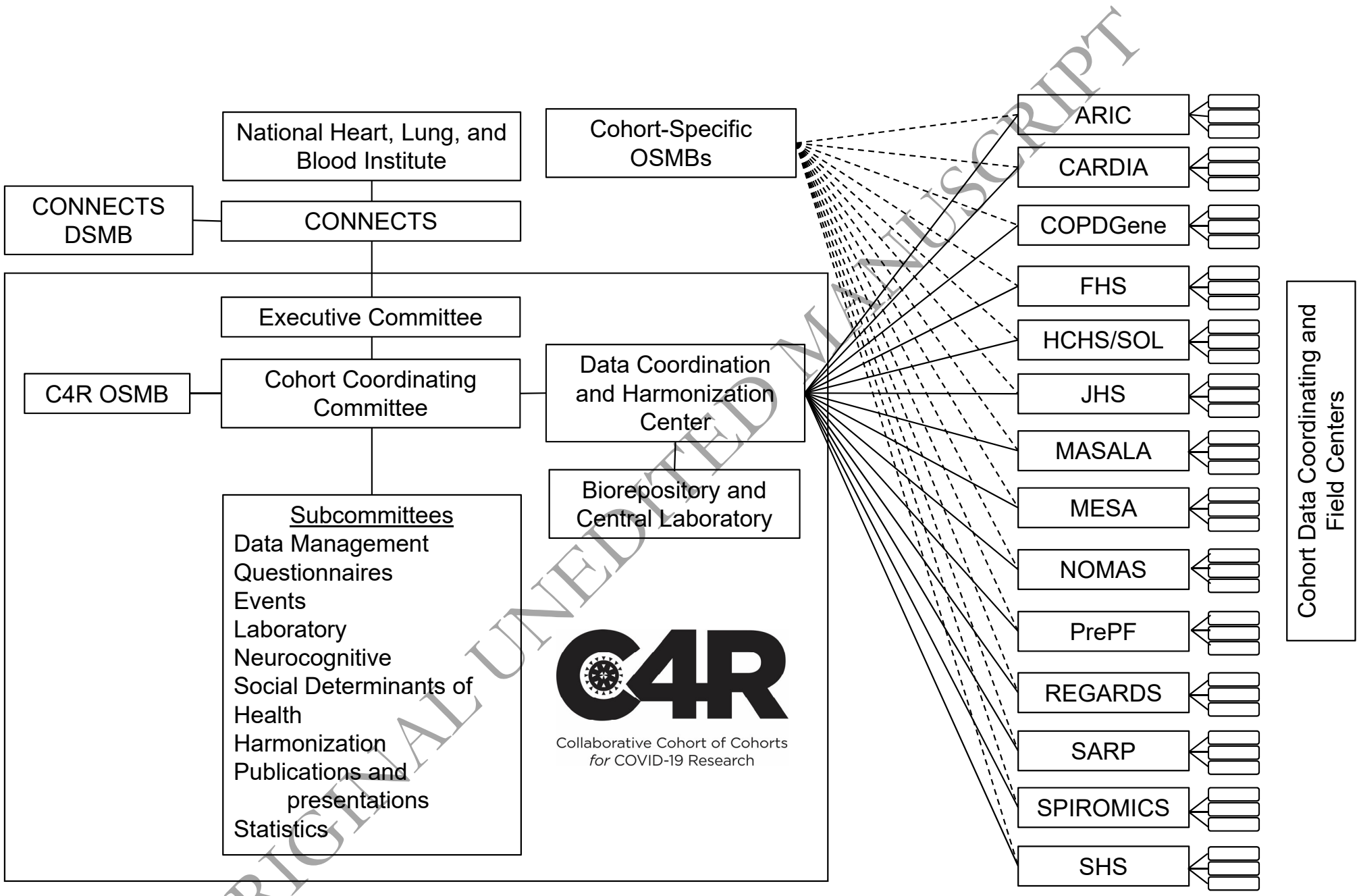
ORIGINAL UNEDITED MANUSCRIPT



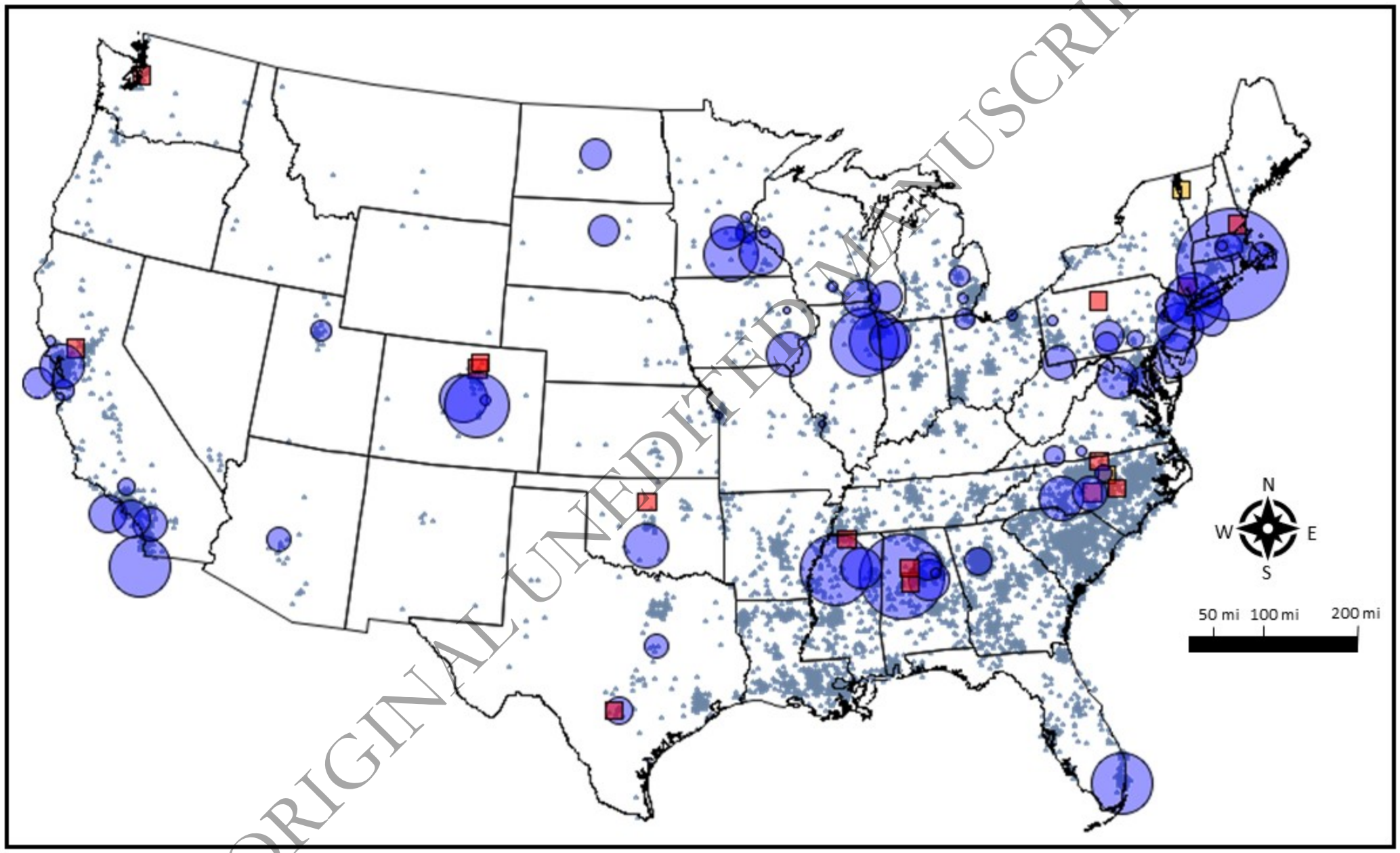
ORIGINAL UNEDITED MANUSCRIPT



ORIGINAL UNEDITED MANUSCRIPT



ORIGINAL UNEDITED MANUSCRIPT



ORIGINAL UNEDITED MANUSCRIPT

Lessons Learned

- Unlike case registries and electronic health record-based studies, C4R's repeated examinations and cognitive assessments before and after COVID-19 provide important opportunities to estimate the biological, clinical, social, and behavioral impact of the COVID-19 pandemic on changes in long-term physical and mental health across multiple domains.
- Success of the C4R serosurvey suggests that blood collection by self-collected dried blood spots is feasible and acceptable for remote and, potentially, repeated biosampling and can be integrated into existing and new population-based cohort studies. Traditional cohort participant-contact methods and procedures work for in-home biosample collection and are critical to success.
- C4R provides a model for cross-cohort collaboration, harmonization, 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.

ORIGINAL UNEDITED MANUSCRIPT