American Journal of Epidemiology
Published by Oxford University Press on behalf of the Johns Hopkins Bloomberg School of Public Health 2018.

Vol. 187, No. 11
DOI: 10.1093/aje/kwy139
Advance Access publication:
June 30, 2018

Study Design

Harmonization of Respiratory Data From 9 US Population-Based Cohorts

The NHLBI Pooled Cohorts Study

Elizabeth C. Oelsner*, Pallavi P. Balte, Patricia A. Cassano, David Couper, Paul L. Enright, Aaron R. Folsom, John Hankinson, David R. Jacobs, Jr., Ravi Kalhan, Robert Kaplan, Richard Kronmal, Leslie Lange, Laura R. Loehr, Stephanie J. London, Ana Navas Acien, Anne B. Newman, George T. O'Connor, Joseph E. Schwartz, Lewis J. Smith, Fawn Yeh, Yiyi Zhang, Andrew E. Moran, Stanford Mwasongwe, Wendy B. White, Sachin Yende, and R. Graham Barr

* Correspondence to Dr. Elizabeth C. Oelsner, Division of General Medicine, Department of Medicine, Columbia University Medical Center, 630 West 168th Street, PH9-105, New York, NY 10032 (e-mail: eco7@cumc.columbia.edu).

Initially submitted November 6, 2017; accepted for publication April 17, 2018.

Chronic lower respiratory diseases (CLRDs) are the fourth leading cause of death in the United States. To support investigations into CLRD risk determinants and new approaches to primary prevention, we aimed to harmonize and pool respiratory data from US general population-based cohorts. Data were obtained from prospective cohorts that performed prebronchodilator spirometry and were harmonized following 2005 ATS/ERS standards. In cohorts conducting follow-up for noncardiovascular events, CLRD events were defined as hospitalizations/deaths adjudicated as CLRD-related or assigned relevant administrative codes. Coding and variable names were applied uniformly. The pooled sample included 65,251 adults in 9 cohorts followed-up for CLRD-related mortality over 653,380 person-years during 1983–2016. Average baseline age was 52 years; 56% were female; 49% were never-smokers; and racial/ethnic composition was 44% white, 22% black, 28% Hispanic/Latino, and 5% American Indian. Over 96% had complete data on smoking, clinical CLRD diagnoses, and dyspnea. After excluding invalid spirometry examinations (13%), there were 105,696 valid examinations (median, 2 per participant). Of 29,351 participants followed for CLRD hospitalizations, median follow-up was 14 years; only 5% were lost to follow-up at 10 years. The NHLBI Pooled Cohorts Study provides a harmonization standard applied to a large, US population-based sample that may be used to advance epidemiologic research on CLRD.

asthma; cohort studies; COPD; harmonization; spirometry

Abbreviations: ARIC, Atherosclerosis Risk in Communities; ATS, American Thoracic Society; CARDIA, Cardiovascular Risk Development in Young Adults; CHS, Cardiovascular Health Study; CLRD, chronic lower respiratory disease; COPD, chronic obstructive pulmonary disease; ERS, European Respiratory Society; FEV1, forced expiratory volume in 1 second; FHS-O, Framingham Heart Study—Offspring Cohort; FVC, forced vital capacity; HCHS/SOL, Hispanic Community Health Study/Study of Latinos; HABC, Health, Aging and Body Composition; ICD-9, *International Classification of Diseases, Ninth Revision*; ICD-10, *International Classification of Diseases, Tenth Revision*; JHS, Jackson Heart Study; MESA, Multi-Ethnic Study of Atherosclerosis; NHLBI, National Heart, Lung, and Blood Institute; QC, quality control; SHS, Strong Heart Study.

Chronic lower respiratory diseases (CLRDs)—defined by the *International Classification of Diseases, Tenth Revision* (ICD-10) as chronic obstructive pulmonary disease (COPD), emphysema, chronic bronchitis, bronchiectasis, and asthma (1, 2)—are the fourth leading cause of death in the United States and globally (2–4). Of the CLRDs, COPD, which is defined by airflow limitation that does not fully reverse, is the most deadly, accounting for 6% of deaths worldwide in 2016 (5–7). Asthma, characterized by intermittent airflow limitation, is the most prevalent, affecting 16% of the world's population (5, 8). An estimated 15%–45% of adults with CLRD have features of both COPD and asthma (9, 10). Acute exacerbations of CLRD caused over 2

million US Emergency Department visits in 2014 (8) and are the main driver of US CLRD costs, which are projected to exceed $100 billion annually (11, 12).

There remain important knowledge gaps regarding risk determinants for CLRD and its progression. While smoking is the major known risk factor for COPD, further investigation is needed regarding the large minority of COPD that occurs in never-smokers (13–15), the risks of light and nondaily smoking (more prevalent in contemporary, multiethnic populations (16, 17)), the significance of maximally attained lung function in early adulthood (18), the relevance of developmental and early-life factors to lifetime CLRD risk (19), and the occurrence of CLRD symptoms and clinical events in persons who do not meet standard diagnostic criteria for COPD or asthma (20–23). In addition, many prior studies were conducted in relatively modest-sized and mainly non-Hispanic white samples, limiting statistical power and generalizability to the multiethnic US population, in which race/ethnicity, geography, and socioeconomics are known to affect lung function and CLRD risk (24–31).

Population-based cohorts remain fundamental to understanding the natural history of CLRD and determinants of disease incidence, which are particularly relevant to developing and targeting primary prevention strategies (32). Since the 1970s, numerous US cohorts have collected data relevant to CLRD epidemiology, including spirometry, CLRD hospitalizations and mortality, and time-varying smoking exposures—measures that are lacking from National Health and Nutrition Examination Survey and administrative data sets (27, 33). While data collection has been highly standardized, data management has varied across studies, and there is, to our knowledge, no standard coding taxonomy for these data.

The potential benefits of harmonizing and pooling US cohort data include sufficient samples to enhance statistical precision for subgroup analyses and adequate follow-up for analyses of incident CLRD-related clinical events (34, 35). However, the need for systematic validation and reconciliation of previously collected data was recognized as a potential barrier to pooling (35) and a limitation to meta-analytical approaches (36, 37), motivating contemporary interest in phenotype harmonization across cohorts (38, 39). In this work, we describe our approach to harmonization of data on lung function, respiratory events, and other relevant respiratory covariates across 9 US prospective cohort studies in the NHLBI Pooled Cohorts Study.

METHODS

Cohorts

The NHLBI Pooled Cohorts Study aimed to include all large National Heart, Lung, and Blood Institute (NHLBI)-funded prospective cohorts that measured spirometry (Web Figure 1, available at https://academic.oup.com/aje) (40–48). Most studies were initially funded to study cardiovascular epidemiology and were designed to capture target age ranges and racial/ethnic groups, as summarized in Table 1.

All studies were approved by institutional review boards at participating institutions, and all participants provided written informed consent. Participants who did not consent to having their data analyzed for noncardiovascular research were excluded from the present work.

Ancillary study and/or data analysis approvals, as well as data use agreements, were obtained from each cohort, and data were centralized at Columbia University. Investigators from all cohorts—in particular, those chairing pulmonary working groups and spirometry reading centers—were invited to collaborate and participate in regular teleconferences and in-person meetings.

Harmonization

All available data and data dictionaries were requested from each cohort for the main respiratory measures (spirometry, events, symptoms, diagnoses, medications), inhalational exposures (smoking, occupational, environmental), and standard sociodemographic and anthropometric variables. Variables available in 2 or more cohorts were considered potentially suitable for harmonization and pooling.

Consistent with phenotype harmonization approaches in the Trans-Omics for Precision Medicine (TopMED) Project (38), which is performing whole genome sequencing and collecting other "-omics" data in a subset of the NHLBI Pooled Cohorts, potentially harmonizable variables were first reviewed qualitatively by review of data dictionaries and study protocols, with cohort-specific investigator and data analyst input. They were next evaluated quantitatively, with comparison of means, variances, outliers, and missing data.

Within-individual data were used to minimize missing data and identify inconsistencies. Logic rules were applied (e.g., current smokers could not subsequently be classified as never-smokers; details available at the study website (49)). Outlier values were checked against repeated measurements in the same subject and reviewed by 3 coauthors (E.C.O., P.P.B., R.G.B.) to determine which extreme values should be recoded to last-value-carried-forward or missing. All recodings were catalogued.

A subset of the data (sociodemographic factors, anthropometry, smoking variables) was independently reharmonized by 2 investigators (Y.Z., A.E.M.) and results were compared. Any inconsistencies were investigated and corrected.

Straightforward harmonized variable names were developed and standardized coding rubrics (e.g., "0" = "no," "1" = "yes") were applied (Web Table 1). Categories were collapsed to align with the cohort(s) providing the fewest categories (least precision) for a given variable.

Variable- and cohort-specific harmonization protocols are provided at the study website (49). Additional participant-level quality control (QC) data are available on request, with permission from the relevant cohorts.

Spirometry

Lung function was measured using water-seal, dry-rolling-seal, or one model of flow-sensing spirometers. Many cohorts used similar or identical equipment, spirometry reading centers, and protocols. One investigator (P.L.E.) ran the spirometry reading centers and designed the protocols for Atherosclerosis Risk in Communities (ARIC) Examination 5, Cardiovascular Health Study (CHS), Hispanic Community Health Study/Study of Latinos (HCHS/SOL), Jackson Heart Study (JHS), and Multi-Ethnic Study of Atherosclerosis (MESA), in collaboration with 2 others (J.H., R.G.B.) for ARIC Examination 5, CHS Year 18, HCHS/

Table 1. Design Features of Cohorts Included in the NHLBI Pooled Cohorts Study, United States, 1983–2016

Cohort	Site	Recruitment Period	Age at Recruitment, years	Sample Size	Race/Ethnicity				
					White, %	Black, %	Hispanic/ Latino, %	Asian, %	American Indian, %
ARIC	Winston-Salem, North Carolina	1987–1989	45–64	15,368[a]	73	27			
	Jackson, Mississippi								
	Minneapolis, Minnesota								
	Washington County, Maryland								
CARDIA	Birmingham, Alabama	1985–1986	18–30	5,114[b]	48	52			
	Chicago, Illinois								
	Minneapolis, Minnesota								
	Oakland, California								
CHS	Pittsburgh, Pennsylvania	1989–1990	≥65	5,888	84	16			
	Winston-Salem, North Carolina								
	Sacramento, California								
	Baltimore, Maryland								
FHS-O[c]	Framingham, Massachusetts	1971–1975	≥18	5,124	100				
HABC[d]	Pittsburgh, Pennsylvania	1997–1998	70–79	3,075	58	42			
	Memphis, Tennessee								
	San Francisco, California								
HCHS-SOL	San Diego, California	2008–2011	18–74	16,415			100		
	Chicago, Illinois								
	Bronx, New York								
	Miami, Florida								
JHS	Jackson, Mississippi	2000–2004	20–95	5,306[e]		100			
MESA[f]	Winston-Salem, North Carolina	2000–2002	45–84	7,071	39	27	23	11	
	Upper Manhattan/Bronx, New York								
	Los Angeles, California								
	Baltimore, Maryland								
	Chicago, Illinois								
	Minneapolis, Minnesota								
SHS[g]	Phoenix, Arizona	1989–1991	45–74	3,516					100
	Southwestern Oklahoma								
	Western and central North and South Dakota								
NHLBI Pooled Cohorts Study	17 sites	1971–2011	≥18	65,251[h]	44	22	28	1	5

Abbreviations: ARIC, Atherosclerosis Risk in Communities; CARDIA, Coronary Artery Risk Development in Young Adults; CHS, Cardiovascular Health Study; FHS-O, Framingham Heart Study—Offspring Cohort; HABC, Health, Aging and Body Composition; HCHS/SOL, Hispanic Community Health Study/Study of Latinos; JHS, Jackson Heart Study; MESA, Multi-Ethnic Study of Atherosclerosis; SHS, Strong Heart Study.

[a] In ARIC, out of 15,792 participants, 424 gave restricted consent.

[b] In CARDIA, out of 5,115 participants, 1 withdrew consent.

[c] Children of original Framingham Heart Study participants.

[d] Participants were required to have no major disabilities or functional limitations.

[e] In JHS, out of 5,306 participants, 1,626 were ARIC corecruits.

[f] Participants were required to be free of clinical cardiovascular disease. Sample includes 6,814 participants in MESA plus 257 participants recruited into the MESA Air Pollution Study in 2004–2006 under the same inclusion/exclusion criteria and followed in the same way.

[g] Participants recruited from 13 tribes and communities.

[h] Total of NHLBI Pooled Cohorts Study excludes 1,626 JHS participants who were ARIC corecruits.

SOL, and MESA. Bronchodilators were not administered in most cohorts; however, Framingham Heart Study—Offspring Cohort (FHS-O) Examination 9, HCHS/SOL, and MESA Examinations 5–6 attempted postbronchodilator spirometry in those with airflow limitation, defined as prebronchodilator forced expiratory volume in 1 second (FEV1)/forced

Table 2. Spirometry Examinations, Methods, and Harmonized Quality Grades, the NHLBI Pooled Cohorts Study, 1983–2016

Cohort Examination	Year	Spirometer[a]	ATS Guideline	Tests (n = 120,933; 100%)	Valid, No.		Invalid, No.			Valid Measurements (n = 105,696; 87.4%)	
					A (n = 65,294; 54.0%)	B (n = 40,402; 33.4%)	C (n = 4,939; 4.1%)	D (n = 6,643; 5.5%)	F (n = 3,655; 3.0%)		
ARIC											
1	1987–1989	WS (Collins)	1979	15,230		13,459		1,771		13,459	88.4
2	1990–1992	WS (Collins)	1979	13,533		12,345		1,188		12,345	91.2
3	2011–2013	DRS (SM/OMI)	2005	4,393	2,838	1,095	343	96	21	3,933	89.5
CARDIA											
0	1985–1986	WS (Collins)	1979	4,860	3,993	21	229	158	459	4,014	82.6
2	1987–1988	WS (Collins)	1979	4,466	3,900	13	220	154	179	3,193	87.6
5	1990–1991	WS (Collins)	1987	4,267	3,957	3	115	72	120	3,960	92.8
10	1995–1996	WS (Collins)	1987	3,753	3,602	5	73	38	35	3,607	96.1
20	2005–2006	DRS (SM/OMI)	1994	3,430	2,483	654	154	90	49	3,137	91.5
30[b]	2015–2016	FS (ndd)	2005	2,749							
CHS											
2	1989–1990	WS (Collins)	1979	5,111	2,295	1,310	637	347	522	3,605	70.5
6	1993–1994	WS (Collins)	1979	4,044	1,922	1,230	459	323	110	3,152	77.9
9	1996–1997	WS (Collins)	1979	2,836	2,431	273	73	36	23	2,704	95.3
18	2005–2006	FS (ndd)	2005	995	709	170	36	45	35	879	88.3
FHS-O											
3	1983–1987	WS (Collins)	1979	2,380		1,536		844		1,536	64.5
5	1991–1995	WS (Collins)	1979	3,271	1,847	661	29	21	713	2,508	76.7
6	1995–1998	WS (Collins)	1979	2,863	1,940	703	23	9	188	2,643	92.3
7	1998–2001	WS (Collins)	1994	2,609	1,962	494	26	11	116	2,456	94.1
8	2005–2008	WS (Collins)	1994	2,574	2,292	71		160	51	2,363	91.8
9	2011–2014	WS (Collins)	2005	1,884	1,757	45		41	41	1,802	95.6
HABC											
1	1997–1998	DRS (SM/OMI)	1994	2,863	2,047	430		305	81	2,477	86.5
5	2001–2002	DRS (SM/OMI)	1994	2,096	1,525	270		245	56	1,795	85.6
8	2004–2005	FS (ndd)	1994	1,648	1,081	276		229	62	1,357	82.3
10	2006–2007	FS (ndd)	2005	1,456	955	308		140	53	1,263	86.7
HCHS/SOL	2008–2011	DRS (SM/OMI)	2005	15,576	11,470	2,733	885	410	78	14,203	91.2
JHS											
JHS only	2000–2004	DRS (SM/OMI)	1994	3,501	2,370	539	227	99	266	2,909	83.1
ARIC corecruits	2000–2004	DRS (SM/OMI)	1994	1,505	1,090	177	69	41	128	1,267	84.2

Table continues

Table 2. Continued

Cohort Examination	Year	Spirometer[a]	ATS Guideline	Tests (*n* = 120,933; 100%)	Valid, No.		Invalid, No.			Valid Measurements (*n* = 105,696; 87.4%)	
					A (*n* = 65,294; 54.0%)	B (*n* = 40,402; 33.4%)	C (*n* = 4,939; 4.1%)	D (*n* = 6,643; 5.5%)	F (*n* = 3,655; 3.0%)		
MESA											
3 or 4	2004–2007	DRS (SM/OMI)	1994	3,953	3,371	377	141	64	0	3,748	94.8
5	2010–2011	DRS (SM/OMI)	2005	3,199	2,013	716	334	136	0	2,729	85.3
6[b]	2016–	DRS (SM/OMI)	2005	2,850							
SHS	1993–1995	DRS (Mijnhardt)	1987	2,625	1,444	488	198	238	257	1,932	73.6

Abbreviations: ARIC, Atherosclerosis Risk in Communities; ATS, American Thoracic Society; CARDIA, Coronary Artery Risk Development in Young Adults; CHS, Cardiovascular Health Study; DRS, dry-rolling-seal; FHS-O, Framingham Heart Study—Offspring Cohort; FS, flow-sensing; HABC, Health, Aging and Body Composition; HCHS/SOL, Hispanic Community Health Study/Study of Latinos; JHS, Jackson Heart Study; MESA, Multi-Ethnic Study of Atherosclerosis; SHS, Strong Heart Study; WS, water-seal.

[a] The water-seal spirometer was the Collins Survey II Spirometer (Warren E. Collins, Inc., Braintree, Massachusetts). The dry-rolling-seal spirometer was from SensorMedics (Viasys Corp., Yorba Linda, California; OMI, Houston, Texas), except in the Strong Heart Study (Mijnhardt B.V., Bunnik, the Netherlands). The flow-sensing spirometer was the EasyOne (ndd Medical Technologies, Inc., Andover, Massachusetts).

[b] Examinations incomplete at time of harmonization; data not yet harmonized.

vital capacity (FVC) ratio <0.70 or less than the lower limit of normal (50, 51).

Spirometry protocols were designed based upon American Thoracic Society (ATS)/European Respiratory Society (ERS) guidelines. Because the cohorts were examined from 1971 to the present—and ATS/ERS standards were issued and revised in 1979, 1987, 1994, and 2005 (50–54)—there was modest heterogeneity in protocols, QC, and reporting standards across cohorts and, in some cases, among repeated examinations within cohorts.

We therefore developed a spirometry quality grading rubric based upon current (2005) ATS/ERS standards (50) (Table 2). Valid spirometry was defined as acquisition of ≥2 curves meeting acceptability criteria (50–54), with the 2 largest lung volumes reproducible within 150 mL (50). Spirometry not meeting this standard was defined as invalid. In sensitivity analyses, the 1994 reproducibility standard of <200 mL was used (54).

Reproducibility of the 2 largest volumes was further used to classify valid spirometry into grades A (<100 mL) and B (<150 mL), which met 2005 criteria (50), and C (<200 mL), which met 1994 but not 2005 repeatability criteria. Grade D was defined by nonreproducibility (>200 mL) or only 1 acceptable curve, and grade F was defined by nonreproducibility (>250 mL) or failure to obtain 1 acceptable curve.

FEV1 and FVC were graded independently. Best FEV1 and best FVC were used to calculate FEV1/FVC, which was classified as valid if both FEV1 and FVC measurements were valid.

This grading system was previously applied by 3 coauthors (P.L.E., J.H., R.G.B.) in 4 cohorts (ARIC Examination 5, CHS, HCHS/SOL, MESA) and also applied in the Strong Heart Study (SHS). For the remaining cohorts and examinations, the grading rubric was adapted based upon the data available, as summarized in Table 2.

Events

All-cause mortality was ascertained in all 9 cohorts. Five cohorts (Cardiovascular Risk Development in Young Adults (CARDIA), CHS, Health, Aging and Body Composition (HABC), HCHS/SOL, SHS) adjudicated noncardiovascular causes of death, including CRLD mortality, via protocolized medical record review by a clinical events committee. Two cohorts that did not adjudicate respiratory and CLRD mortality (ARIC, MESA) nonetheless collected ICD data for all deaths.

Only 2 cohorts (HABC, HCHS/SOL) were designed to prospectively ascertain and adjudicate CLRD hospitalizations (55). A subset of MESA deaths and hospitalizations was retrospectively adjudicated for CLRD (22, 56). Four cohorts (ARIC, CHS, HCHS/SOL, MESA) collected ICD data for all hospitalizations occurring over follow-up. CARDIA collected only self-reported CLRD hospitalization data, which appeared to be underreported (cumulative incidence of reported CLRD hospitalizations <1%); hence, these data were not harmonized. Noncardiovascular hospitalization data were not available in SHS, and neither CLRD mortality nor CLRD hospitalization data were available in FHS-O or JHS at the time of publication (August 2018).

To supplement adjudicated respiratory endpoints in cohorts collecting diagnosis-code data for deaths and hospitalizations, we selected all events assigned diagnostic codes for asthma (*International Classification of Diseases, Ninth Revision* (ICD-9): 493,

ICD-10: J45–J46), COPD (ICD-9: 496, ICD-10: J44), chronic bronchitis (ICD-9: 490–491, ICD-10: J40–J42), and/or emphysema (ICD-9: 492, ICD-10: J43).

According to an algorithm we previously developed in HCHS/SOL and validated in MESA (56), severe obstructive lung events (SOLE) were defined as hospitalizations or deaths adjudicated as primarily attributable to CLRD or, if adjudication was lacking, those with CLRD coded as the primary discharge diagnosis or as the underlying cause of death. CLRD-related events were defined as hospitalizations or deaths adjudicated as primarily or secondarily attributable to CLRD, or, if adjudication was lacking, those with CLRD listed in any ICD code position.

Clinical lung disease and symptoms

Participants in all cohorts were asked to report prior physician diagnoses of asthma, COPD, chronic bronchitis, or emphysema. Because the term COPD was not well-known to the general public prior to the 21st Century, self-reported chronic bronchitis and emphysema were coded as self-reported COPD.

Utilization of inhaled bronchodilators and inhaled corticosteroids was assessed by self-report or medication inventory in all cohorts at each examination (57, 58).

Dyspnea was assessed in all cohorts. In ARIC, CARDIA, CHS, MESA, and SHS, it was classified using the modified Medical Research Council (mMRC) scale (59), additionally allowing definition of mMRC-classified chronic bronchitis.

Smoking

Smoking status was assessed by standard questionnaire items in all cohorts and all examinations (60). Pack-years were self-reported at baseline examinations and updated based upon time-variant cigarettes-per-day as described on the study website. Secondhand smoke exposure was self-reported in selected cohorts.

Covariates

In all cohorts, race/ethnicity, sex, and educational attainment were self-reported. Race/ethnicity was defined using the 2000 US Census approach (61), which is comparable to the proposed 2020 Census approach (62). Body mass index was calculated from height and weight. Cohort-specific procedures are described on the study website.

Statistical analysis

The baseline characteristics of the pooled sample and subsamples with valid spirometry and CLRD events follow-up were tabulated and compared.

Within- and between-subject variability in lung function was compared before and after exclusion of invalid spirometry using within- and between-subject variances and their ratio, the intraclass correlation, in mixed models including adjustment for age, sex, height, and race/ethnicity. The number of lung function outliers, defined by values ≥2.5 standard deviations from the mean, was also assessed, as was the proportion of the population with ≥15% improved lung function over time, as this is not consistent with long-accepted physiologic declines in lung function in middle and older ages. Results were compared using 2005 versus 1994 reproducibility standards (50, 54).

Statistical analyses were performed in SAS, version 9.4 (SAS Institute, Inc., Cary, North Carolina).

RESULTS

The NHLBI Pooled Cohorts Study included 65,251 participants from all 9 large US population-based prospective cohort studies that measured spirometry in adults (Table 3). The mean age at baseline examination was 52 ± 16 years; there were 17,005 (26%) participants who were 18–45 years old, the age range during which adults typically attain maximum lung function (63, 64). Fifty-six percent were female. Compared with the current US population, nonwhites were oversampled: 44% of participants were white, 22% were black, 28% were Hispanic/Latino, 5% were American Indian, and 1% were Asian.

After between- and within-individual QC and harmonization, missing data for demographic factors and self-reported lung disease were infrequent or nonexistent (Web Table 2). Smoking status was missing for 195 participants (0.3%), and pack-years were missing for 1,713 (2.6%). Among data undergoing independent reharmonization, one incongruence related to selection of a single variable was identified in one cohort and reconciled; otherwise, the harmonization was fully replicated.

Spirometry completion

All cohorts selected all participants for spirometry at baseline, except for MESA. Spirometry was performed in MESA as part of an ancillary study that randomly selected 4,483 participants in Examination 3 or 4 (65) in addition to all 257 new recruits in the MESA Air Pollution Study (66). Of 65,251 participants in the NHLBI Pooled Cohort Study, 2,331 (the remainder of MESA participants; 4%) were consequently not selected, and 3,408 (5%) additional participants from all studies declined spirometry (Figure 1).

Of 59,512 participants attempting at least 1 spirometry examination, 46,440 (78%) had valid spirometry at all attempted examinations, 4,499 (8%) participants had no valid spirometry, and 8,573 (14%) had valid spirometry as some but not all examinations.

Among 55,013 participants with at least 1 valid spirometry measurement, the median number of valid spirometry measurements was 2 (interquartile range, 1–3), yielding 105,696 spirometry examinations over a median of 2.80 (interquartile range, 0–8.93) years. Fifty percent ($n = 27,328$) had at least 1 subsequent valid measurement of spirometry, and 25% ($n = 13,767$) had 3 or more. Four or more valid measurements were available in 6,493 participants, all from ARIC/JHS corecruits, CARDIA, or FHS-O.

Eighty-four percent ($n = 53,191$) of participants had both valid spirometry and complete sociodemographic, anthropometric, and smoking data; of these, 26,222 (49%) had more than 1 valid measurement of spirometry.

Spirometry quality

Fifty-four percent of spirometry examinations ($n = 65,294$) were of the highest quality (grade A), while absence of acceptable

Table 3. Harmonized Spirometry Quality Grading Rubric for Forced Expiratory Volume in 1 Second and Forced Vital Capacity, NHLBI Pooled Cohorts Study, 1983–2016

Cohort	Valid Spirometry Examination ≥2 Acceptable Curves and Largest 2 Values Reproducible Within 150 mL				Invalid Spirometry Examination <2 Acceptable Curves or Largest 2 Values >150 mL Apart					
	A[a]		B[a]		C[a]		D[a]		F[a]	
	No. of Acceptable Curves	Reproducibility	No. of Acceptable Curves	Reproducibility	No. of Acceptable Curves	Reproducibility	No. of Acceptable Curves	Reproducibility	No. of Acceptable Curves	Reproducibility
ARIC[b]										
1–2			3				0–2			
5	3	≤100 mL	2	≤150 mL	2	≤200 mL	1	≥200 mL	0	
CARDIA[c]	3	≤150 mL	2	≤150 mL	2	≤200 mL	2	≤250 mL	0–1	>250 mL
CHS[d]	3	≤100 mL or ≤5%	2	≤150 mL	2	≤200 mL	2	≤250 mL	0–1	>250 mL
FHS-O[e]										
3			2				0–1			
5–7	3	≤150 mL	2	≤150 mL	2	≤200 mL	2	≤250 mL	0–1	>250 mL
8–9	3	≤5%	2	≤5%			1	>5%	0	
HABC[f]		≤100 mL		≤200 mL				≤300 mL		>300 mL
HCHS/SOL	2	≤100 mL	2	≤150 mL	2	≤200 mL	1	≤250 mL	0	>250 mL
JHS[g]	3	≤100 mL	2	≤150 mL	2	≤200 mL	1	≤250 mL	0	
MESA	2	≤100 mL	2	≤150 mL	2	≤200 mL	1	≤250 mL	0	>250 mL
SHS	2	≤100 mL	2	≤150 mL	2	≤200 mL	1	≤250 mL	0	>250 mL

Abbreviations: ARIC, Atherosclerosis Risk in Communities; CARDIA, Coronary Artery Risk Development in Young Adults; CHS, Cardiovascular Health Study; FEV1, forced expiratory volume in 1 second; FHS-O, Framingham Heart Study—Offspring Cohort; FVC, forced vital capacity; HABC, Health, Aging and Body Composition; HCHS/SOL, Hispanic Community Health Study/Study of Latinos; JHS, Jackson Heart Study; MESA, Multi-Ethnic Study of Atherosclerosis; QC: quality control; SHS, Strong Heart Study.

[a] Reproducibility pertains to largest 2 values of FEV1 or FVC. For grades A, B, and C, both acceptability and reproducibility criteria must be met or exceeded. For grades D and F, either acceptability or reproducibility criteria could be met in order to qualify for the grade (e.g., "D" spirometry in HCHS/SOL includes examinations with 2 FEV1 measurements that are 225 mL apart as well as examinations with only 1 acceptable curve).

[b] In ARIC Examinations 1–2, considerable QC information was available, but it corresponded poorly with the QC approach used in the main grading rubric. Based upon prior QC efforts applied in these ARIC data (43), 2005 American Thoracic Society/European Respiratory Society acceptability criteria, and expert opinion, we classified as valid those spirometry examinations with 3 or more maneuvers attempted and none of the following technical errors: no flow-volume loop recorded or the computer started after the beginning of the forced exhalation; breath-hold leak >5% detected; submaximal participant effort; no plateau during forced exhalation; or incorrect spirometer calibration.

[c] CARDIA used a similar grading approach to ours in its year-20 examination, with only minor discrepancies that did not affect the distinction between valid and invalid spirometry. However, CARDIA had not applied this standard to the prior 4 CARDIA spirometry examinations (years 0, 2, 5, and 10). We therefore obtained full spirometry data from these examinations, including all available curves, and consistently applied CARDIA's own year-20 approach.

[d] FVC was not measured in CHS Examination 18; FEV6 (6 seconds) was therefore interpreted as FVC.

[e] FHS-O Examinations 1 and 2, which were performed prior to the 1979 publication of American Thoracic Society/European Respiratory Society spirometry standards, were excluded. For FHS-O Examination 3, only the number of acceptable curves was available for QC review; to correspond best with our standardized rubric, we therefore dichotomized spirometry examinations into valid (≥2 acceptable curves) versus invalid (<2 acceptable curves). FHS-O provided data on lung volumes for all acceptable curves obtained in Examinations 5–7, and we therefore applied the CARDIA year-20 grading rubric to these data. In FHS-O Examinations 8–9, only the number of acceptable curves and their reproducibility within 5% were available; hence, these data were used to classify spirometry provisionally into grades A, B, D, and F.

[f] The HABC grading system defined grade A as <100 mL and B as <200 mL. Experience in other elderly cohorts (e.g., CHS) indicated that, among spirometry repeatable between 100 mL and 200 mL, repeatability <150 mL was much more frequent than 150–200 mL. Hence, HABC grade B was treated as grade B in our rubric.

[g] JHS provided data on lung volumes for all acceptable curves obtained in Examination 1, and we therefore applied the NHLBI Pooled Cohorts Study grading rubric to these data.

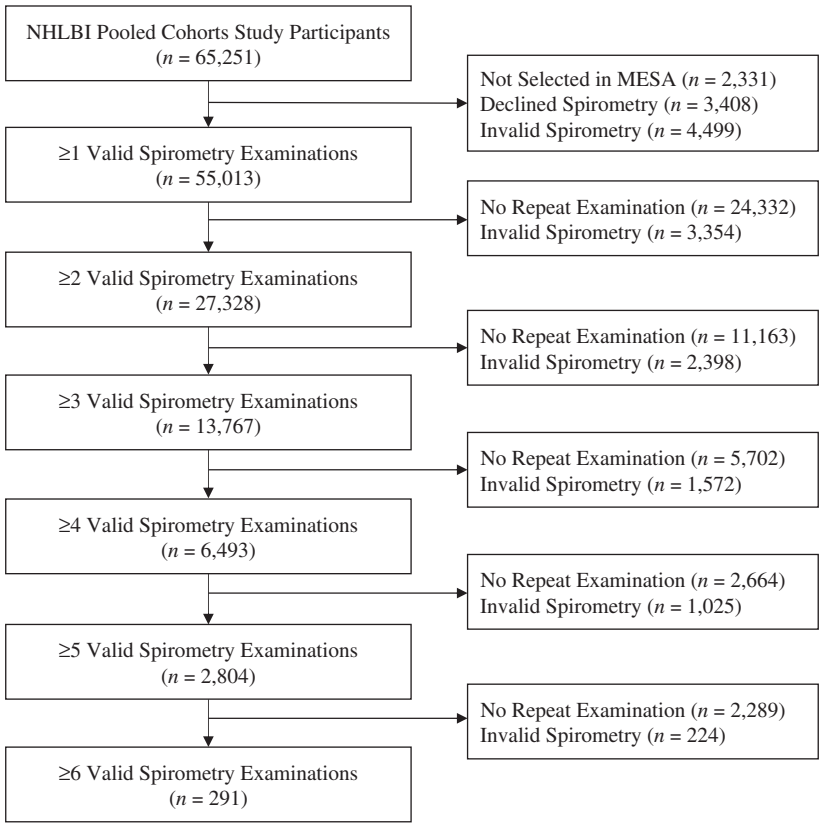

Figure 1. Flow chart of longitudinal lung function data in the NHLBI Pooled Cohorts Study, United States, 1983–2016. MESA, Multi-Ethnic Study of Atherosclerosis.

curves (grade F) was infrequent ($n = 3{,}655$, 3%) (Web Table 3; Web Figures 2–7). Examinations completed earlier in calendar time had a lower proportion of valid spirometry, and quality mainly improved over subsequent examinations within cohorts. For example, FHS-O Examination 3, for which original QC data was very limited (Table 2), had the lowest proportion of valid results (64.5%). FHS-O Examination 5, for which there was much more data available for QC purposes, also demonstrated relatively low proportion of valid spirometry (76.7%). This was not due to the grading rubric; exactly the same approach was used for FHS-O Examinations 6 and 7, in which valid proportions were 92% and 94%, respectively.

Higher spirometry quality was also more frequently observed in younger participants, white participants, women, and never-smokers without airflow limitation. Nonetheless, due to the relatively high quality of spirometry measurements overall, exclusion of participants with invalid spirometry yielded a sample with similar baseline characteristics (Table 3).

As expected, within- and between-subject variability in FEV1 and FVC were lower among valid versus invalid spirometry measurements. Compared with valid FEV1 measurements, invalid measurements demonstrated higher variance (0.62 versus 0.57) and significantly lower intraclass correlations (0.73 versus 0.84, $P < 0.0001$). The number of outliers (>2.5 standard deviations) was higher (2.1% versus 1.2%), as was the proportion of participants showing an annual increase of ≥15% (1.4% versus 0.08%) (details provided in Web Table 3).

Application of the 1994 reproducibility standard permitted the inclusion of an additional 4,699 participants with grade C spirometry (Web Table 3). Compared with the intraclass correlation for grades A and B (0.89 and 0.87, respectively), the intraclass correlation for grade C was lower (0.85), but it was substantially higher than that for D and F (0.73 and 0.78, respectively). The spirometric characteristics of the sample were similar whether the 2005 or 1994 reproducibility standards were applied.

Events follow-up

Among 6 cohorts (ARIC, CARDIA, CHS, HABC, MESA, SHS) with CLRD mortality data available at the time of manuscript preparation (August 2018), there were 37,982 participants with a median of 16.4 (interquartile range, 11.9–24.4) years of follow-up, yielding 653,380 person-years of observation (Table 4). A subset of 29,356 participants in 4 cohorts (ARIC, CHS, HABC, MESA) were additionally followed for CLRD hospitalizations over a median of 13.9 (interquartile range, 10.2–20.7) years, providing 410,320 person-years of observation for severe obstructive lung events and CLRD-related events. Of these, complete data for standard covariates and smoking were available for 28,398 (96.8%), and 26,935 (94.8%) had complete follow-up at 10 years. Only 19,880

Table 4. Baseline Characteristics of the Total Pooled Population and of Subsamples With Valid Spirometry and Follow-up for Chronic Lower Respiratory Disease Events, NHLBI Pooled Cohorts Study, United States, 1983–2016

Covariate	Total (n = 65,251; 100.0%)		Valid Spirometry[a] (n = 55,013; 84.3%)		Follow-up for CLRD Events			
					CLRD Mortality (n = 37,982; 58.2%)		CLRD Hospitalizations (n = 29,352; 45.0%)	
	No.	%	No.	%	No.	%	No.	%
Cohort								
ARIC + JHS corecruits[b]	15,368	23.6	14,966	27.2	13,323	35.1	13,323	45.4
CARDIA[c]	5,114	7.8	5,033	9.2	5,114	13.5		
CHS	5,888	9.0	4,983	9.1	5,888	15.5	5,888	20.1
FHS-O	5,124	7.9	3,934	7.2				
HABC	3,075	4.7	2,833	5.2	3,075	8.1	3,075	10.5
HCHS/SOL	16,415	25.2	14,203	25.8	–[d]	–[d]	–[d]	–[d]
JHS only[e]	3,680	5.6	2,909	5.3				
MESA[f]	7,071	10.8	4,220	7.7	7,066	18.6	7,066	24.1
SHS	3,516	5.7	1,932	3.5	3,516	9.3		
Age, years[g]	51.9 (16.0)		53.1 (15.8)		56.8 (15.9)		62.4 (10.4)	
Sex								
Female	36,735	56.3	31,003	56.4	20,695	54.5	15,852	54.0
Male	28,516	43.7	24,010	43.6	17,287	45.5	13,500	46.0
Race/ethnicity								
White	28,396	43.5	25,087	45.6	21,700	57.1	19,223	65.5
Black	14,486	22.2	12,202	22.2	10,341	27.2	7,704	26.2
Hispanic/Latino	17,962	27.5	15,098	27.4	1,546	4.1	1,546	5.3
Asian	842	1.3	654	1.2	833	2.2	833	2.8
American Indian	3,545	5.4	1,957	3.6	3,542	9.3	26	0.1
Other	20	0.03	15	0.03	20	0.1	20	0.1
Body mass index[h]	28.1 (5.9)		28.1 (5.8)		27.5 (5.5)		27.7 (5.3)	
Education								
No schooling	2,178	3.3	1,795	3.3	102	0.3	89	0.3
Some schooling	14,216	21.8	11,250	20.5	8,982	23.7	7,011	23.9
High school	17,011	26.1	14,850	27.0	10,207	26.9	7,701	26.2
Some college	10,134	15.5	8,444	15.4	6,731	17.7	4,341	14.8
Bachelor's	20,294	31.1	17,940	32.6	11,910	31.4	10,165	34.6
Smoking status								
Current	14,792	22.7	12,242	22.3	8,469	22.3	5,591	19.1
Former	18,047	27.7	15,229	27.7	12,601	33.2	10,765	36.7
Never	32,217	49.4	27,455	49.9	16,836	44.3	12,928	44.1
Pack years of smoking (years) at baseline[i]	16.0 (28.8)		15.8 (28.6)		20 (31.5)		25 (31.3)	

Abbreviations: ARIC, Atherosclerosis Risk in Communities; CARDIA, Coronary Artery Risk Development in Young Adults; CHS, Cardiovascular Health Study; CLRD, chronic lower respiratory disease; FHS-O, Framingham Heart Study—Offspring Cohort; HABC, Health, Aging and Body Composition; HCHS/SOL, Hispanic Community Health Study/Study of Latinos; JHS, Jackson Heart Study; MESA, Multi-Ethnic Study of Atherosclerosis; SHS, Strong Heart Study.

[a] Valid spirometry examinations defined by ≥2 acceptable curves reproducible within 150 mL, as per 2005 American Thoracic Society/European Respiratory Society standards (i.e., grades A and B by NHLBI Pooled Cohorts Study grading rubric).

[b] In ARIC, 424 gave restricted consent; sample includes 1,622 participants who were corecruits in JHS.

[c] Withdrawal of consent by 1 participant.

[d] CLRD mortality and hospitalizations are being ascertained in HCHS/SOL but were not available to investigators at the time of manuscript preparation (August 2018).

[e] Excludes 1,626 ARIC corecruits.

[f] MESA + 257 new recruits into the MESA Air Pollution Study. In MESA CLRD events follow-up, 5 participants were excluded because of baseline diagnosis of cardiovascular event.

[g] Values are expressed as mean (standard deviation).

[h] Weight (kg)/height $(m)^2$. Values are expressed as mean (standard deviation).

[i] Smoking pack-years in ever smokers. Values expressed as median (interquartile range).

(70.0%) and 15,563 (54.8%) had complete follow-up at 15 and 20 years, respectively, due in part to the fact that MESA is currently reporting a maximum of 14 years of follow-up.

Self-reported lung disease and symptoms

Self-reported CLRD was complete for 96.2% of participants (Web Table 2). Eighty-nine percent ($n = 54{,}387$) had data on chronic bronchitis as classified by the modified Medical Research Council scale.

DISCUSSION

The NHLBI Pooled Cohorts Study harmonized and pooled respiratory data from 9 US prospective cohort studies, yielding a large, population-based sample that ranges from young adulthood to old age, spans over 50 years of observation, and reflects the racial/ethnic, socioeconomic, and geographic diversity in the United States. This work leverages 5 decades of research investment, highly standardized protocols, gold-standard measures, and prospective events surveillance with very high follow-up rates to apply, for the first time, contemporary spirometry standards as well as to define clinical CLRD endpoints using standardized methodology to all available US cohorts. The NHLBI Pooled Cohorts Study thereby provides a unique sample of US adults that may be used to advance epidemiologic research on CLRD, especially among population subgroups (e.g., women, racial/ethnic minorities, and never-smokers) underrepresented in the CLRD literature.

While the importance of data harmonization is drawing increasing attention from the research community (67–70)—driven, at least in part, by the growing availability of heterogeneous "big data"—a current search of PubMed (https://www.ncbi.nlm.nih.gov/pubmed/) for articles on "harmonization AND spirometry" yields zero records. This is despite the fact that standardization of spirometry measures, which are effort-dependent, has been the subject of considerable attention from the clinical community, resulting in a series of evolving guidelines in recent decades (50–54). In this work, exclusion of invalid spirometry reduced between- and within-individual variability, outliers, and lung-function trend irregularities, consistent with decreased measurement error. This was achieved without sacrificing the diversity or scale of the component cohorts.

While meta-analysis is frequently used to address differences in study designs and measurements (36, 71–73), there are well-recognized limitations to this approach, especially in the context of observational studies (74, 75). In this work, we aimed to minimize within-study measurement error and between-study heterogeneity by standardized, longitudinal QC and harmonization, yielding data suitable for meta-analyses as well as for pooled analyses that may be more appropriate for epidemiologic analyses for which multiple sensitivity analyses are often required, stratification is of particular interest, and multivariate methods are indicated (76). Indeed, in the context of increasing interest in harmonization and pooling (77–79), NHLBI Pooled Cohorts Study investigators are collaborating actively with the TransOmics for Precision Medicine Project and the Cross-Cohort Collaboration to share the protocols and data described in this report with the shared goal of promoting precision epidemiology for CLRD as well as other diseases (38, 39).

Strengths of the current work include the inclusion of 9 US epidemiologic cohorts, the systematic harmonization approach, and the expertise of leading pulmonologists and epidemiologists who collaborated in the development of the NHLBI Pooled Cohorts Study, most of whom were involved with the collection the original data. There are nonetheless several limitations and areas requiring further investigation and refinement.

The 9 cohorts included in this work collected high-quality data using highly standardized and often identical protocols; nonetheless, there were certainly distinct differences in measurement across cohorts, not to mention birth cohort and historical differences. This situation necessitated assumptions based upon a combination of empirical analyses, published standards, prior literature, and expert opinion, yet these were sometimes unverifiable. To mitigate these unavoidable uncertainties and to promote ongoing improvement, the present analysis and its supplemental materials describe and justify the current protocol in detail, and even more granular data on participant-level QC was recorded so that it may be made available to collaborators.

While excluding invalid spirometry is expected to minimize misclassification, applying reproducibility standards may also select out individuals with more severe lung disease (80). Hence, beyond contemporary validity standards, we have provided more precise grading for consideration by investigators as they determine which measures to use for testing specific hypotheses. With respect to the potential application of reference equations to estimate percent-predicted lung function, recent work has raised concerns regarding misclassification contingent on age and race/ethnicity (17, 51, 64); thus measured lung function values may be more suitable for epidemiological analyses, with relevant adjustment.

Most cohorts did not attempt representative sampling, so the NHLBI Pooled Cohorts Study is not directly representative of the US population. Nonetheless, all cohorts sample community-dwelling adults, and all major US racial/ethnic groups are represented in substantial proportions.

Postbronchodilator spirometry is important to clinical definition of COPD and asthma (7), yet postbronchodilator spirometry was available only for a limited number of participants in a few cohorts. Prebronchodilator spirometry remains nonetheless highly prognostic of health outcomes and is highly correlated with postbronchodilator measurements in the general population (81).

In conclusion, the NHLBI Pooled Cohorts Study has harmonized and pooled data from 9 gold-standard NIH-funded epidemiologic cohorts in order to promote research on common and increasingly prevalent respiratory diseases that, especially in the case of COPD, lack effective medical therapies or preventive strategies beyond smoking cessation and avoidance.

ACKNOWLEDGMENTS

Author affiliations: Division of General Medicine, Department of Medicine, Vagelos College of Physicians and Surgeons, Columbia University, New York, New York (Elizabeth C. Oelsner, Pallavi P. Balte, Yiyi Zhang, Andrew E. Moran, R. Graham Barr); Department of Epidemiology, Mailman School of Public Health, Columbia University,

New York, New York (Elizabeth C. Oelsner, R. Graham Barr); Division of Nutritional Sciences, Weill Cornell Medical College, Ithaca, New York (Patricia A. Cassano); Collaborative Studies Coordinating Center, Gillings School of Global Public Health, University of North Carolina, Chapel Hill, Chapel Hill, North Carolina (David Couper); Department of Medicine, College of Medicine, University of Arizona, Tucson, Arizona (Paul L. Enright); Division of Epidemiology and Community Health, School of Public Health, University of Minnesota, Minneapolis, Minnesota (Aaron R. Folsom, David R. Jacobs, Jr.); Hankinson Consulting, Inc., Athens, Georgia (John Hankinson); Department of Medicine (Pulmonary and Critical Care), Feinberg School of Medicine, Northwestern University, Chicago, Illinois (Ravi Kalhan, Lewis J. Smith); Department of Epidemiology and Population Health, Albert Einstein College of Medicine, New York, New York (Robert Kaplan); Department of Biostatistics, School of Public Health, University of Washington, Seattle, Washington (Richard Kronmal); Division of Biomedical Informatics and Personalized Medicine, Department of Medicine, University of Colorado, Denver, Colorado (Leslie Lange); Department of Epidemiology, Gillings School of Global Public Health, University of North Carolina, Chapel Hill, Chapel Hill, North Carolina (Laura R. Loehr); National Institute of Environmental Health Sciences, National Institutes of Health, Department of Health and Human Services, Research Triangle Park, North Carolina (Stephanie J. London); Department of Environmental Health Sciences, Mailman School of Public Health, Columbia University, New York, New York (Ana Navas Acien); Department of Epidemiology, Pitt Public Health, University of Pittsburgh, Pittsburgh, Pennsylvania (Anne B. Newman); Department of Medicine, School of Medicine, Boston University, Boston, Massachusetts (George T. O'Connor); Division of Cardiology, Department of Medicine, Vagelos College of Physicians and Surgeons, Columbia University, New York, New York (Joseph E. Schwartz); Department of Psychiatry and Behavioral Sciences, School of Medicine, Stony Brook University, Stony Brook, New York (Joseph E. Schwartz); Biostatistics and Epidemiology, College of Public Health, University of Oklahoma Health Sciences Center, Oklahoma City, Oklahoma (Fawn Yeh); Jackson Heart Study, Jackson State University, Jackson, Mississippi (Stanford Mwasongwe); Jackson Heart Study, Undergraduate Training and Education Center, Tougaloo College, Tougaloo, Mississippi (Wendy B. White); and Division of Pulmonary and Critical Care, Department of Medicine, University of Pittsburgh, Pittsburgh, Pennsylvania (Sachin Yende).

This work was funded by numerous grants from the National Institutes of Health. The NHLBI Pooled Cohorts Study effort is funded by the National Heart, Lung, and Blood Institute (grants K23-HL-130627, R21-HL-129924, and R21-HL121457). The Atherosclerosis Risk in Communities Study has been funded by the National Heart, Lung, and Blood Institute (contracts HHSN268201700001I, HHSN268201700003I, HHSN268201700005I, HHSN268201700004I, and HHSN2682017000021). The Coronary Artery Risk Development in Young Adults Study has been funded by the National Heart, Lung, and Blood Institute (contracts HHSN268201300025C, HHSN268201300026C, HHSN268201300027C, HHSN268201300028C, HHSN268201300029C, and HHSN268200900041C) and the Intramural Research Program of the National Institute on Aging. The Cardiovascular Health Study has been funded by the National Heart, Lung, and Blood Institute (contracts HHSN268201200036C, HHSN268200800007C, N01HC55222, N01HC85079, N01HC85080, N01HC85081, N01HC85082, N01HC85083, and N01HC85086 and grant U01HL080295), with additional contribution from the National Institute of Neurological Disorders and Stroke. Additional support was provided by the National Institute on Aging (grant R01AG023629). The Framingham Heart Study—Offspring Cohort has been funded by the National Heart, Lung, and Blood Institute (contracts N01-HC-25195 and HHSN268201500001I). The Health, Aging and Body Composition Study has been funded by the National Institute on Aging (contracts N01-AG-6-2101, N01-AG-6-2103, and N01-AG-6-2106 and grant R01-AG028050) and the National Institute of Nursing Research (grant R01-NR012459). The Hispanic Community Health Study/Study of Latinos has been funded by the National Heart, Lung, and Blood Institute (contracts HHSN268201300001I/N01-HC-65233 to the University of North Carolina, HHSN268201300004I/N01-HC-65234 to the University of Miami, HHSN268201300002I/N01-HC-65235 to the Albert Einstein College of Medicine, HHSN268201300003I to the University of Illinois at Chicago, N01-HC-65236 to Northwestern University, and HHSN268201300005I/N01-HC-65237 to San Diego State University). The following Institutes/Centers/Offices have contributed to the Study through a transfer of funds to the National Heart, Lung, and Blood Institute: 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 has been funded by National Heart, Lung, and Blood Institute and National Institute for Minority Health and Health Disparities (contracts HHSN268201300049C and HHSN268201300050C to Jackson State University, HHSN268201300048C to Tougaloo College, and HHSN268201300046C and HHSN268201300047C to the University of Mississippi Medical Center). The Multi-Ethnic Study of Atherosclerosis has been funded by the National Heart, Lung, and Blood Institute (contracts RC1-HL-100543, N01-HC-95159, N01-HC-95160, N01-HC-95161, N01-HC-95162, N01-HC-95163, N01-HC-95164, N01-HC-95165, N01-HC-95166, N01-HC-95167, N01-HC-95168, and N01-HC-95169 and grants R01-HL-077612 and R01-HL-093081). This publication was also developed under a STAR research assistance agreement (no. RD831697) (MESA Air), awarded by the US Environmental Protection Agency. The Strong Heart Study has been funded by the National Heart, Lung, and Blood Institute (cooperative agreement grants U01-HL41642, U01-HL41652, U01-HL41654, U01-HL65520,

and U01-HL65521 and research grants R01-HL109315, R01HL109301, R01HL109284, R01HL109282, and R01HL109319). This work was also funded by the Intramural Research Program of the NIH, National Institute of Environmental Health Sciences.

We thank the other investigators, the staff, and the participants of all 9 cohorts for their valuable contributions. In particular, we thank the following study personnel who were instrumental to the accomplishment of this work: Pramod Anugu, Karen Hansen, Nancy Heard-Costa, Hunter Holbrook, Erika Enright, Lucia Juarez, Karen Mutalnik, Lisa Reeves, Linda Sellers, Greta Lee Splansky, Karen D. Hinckley Stukovsky, Kayleen Williams, Anthony Wilsdon, David Vu, Jiayi Xu, Marston Youngblood, and Ying Zhang.

Preliminary data were presented as an abstract at the American Thoracic Society 2017 International Conference in May 19–24, 2017, Washington, DC, and published in abstract form (*Am J Respir Crit Care Med*. 2017;195: A3670).

To analyze the NHLBI Pooled Cohorts Study data, please contact the corresponding author. Harmonized data have been transmitted back to the originating cohorts. For cohorts collaborating with the Trans-Omics for Precision Medicine Project, these data are being uploaded for preliminary whole genome sequencing analyses relating to respiratory phenotypes.

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

Conflict of interest: none declared.

REFERENCES

1. World Health Organization. *International Classification of Diseases, 10th Edition*. Geneva, Switzerland: World Health Organization; 2010.
2. National Center for Health Statistics. *Health, United States, 2016: With Chartbook on Long-term Trends in Health*. Hyattsville, MD: National Center for Health Statistics; 2017.
3. World Health Organization. Top 10 causes of death. Fact Sheet. Geneva, Switzerland: World Health Organization. http://www.who.int/mediacentre/factsheets/fs310/en//. Updated January 2017. Accessed March 14, 2018.
4. Kochanek KD, Murphy SL, J Xu, et al. Mortality in the United States, 2016. *NCHS Data Brief No. 293*. 2017. https://www.cdc.gov/nchs/products/databriefs/db293.htm. Updated December 2017. Accessed July 24, 2018.
5. GBD 2015 Chronic Respiratory Disease Collaborators. Global, regional, and national deaths, prevalence, disability-adjusted life years, and years lived with disability for chronic obstructive pulmonary disease and asthma, 1990–2015: a systematic analysis for the Global Burden of Disease Study 2015. *Lancet Respir Med*. 2017;5(9):691–706.
6. GBD 2016 Causes of Death Collaborators. Global, regional, and national age-sex specific mortality for 264 causes of death, 1980–2016: a systematic analysis for the Global Burden of Disease Study 2016. *Lancet*. 2017;390(10100):1151–1210.
7. Vogelmeier CF, Criner GJ, Martinez FJ, et al. Global strategy for the diagnosis, management, and prevention of chronic obstructive lung disease 2017 report. GOLD executive summary. *Am J Respir Crit Care Med*. 2017; 195(5):557–582.
8. Rui P, Kang K. National Hospital Ambulatory Medical Care Survey: 2014 Emergency Department Summary Tables. 2014; https://www.cdc.gov/nchs/data/nhamcs/web_tables/2014_ed_web_tables.pdf. Accessed March 14, 2018.
9. Postma DS, Rabe KF. The asthma-COPD overlap syndrome. *N Engl J Med*. 2015;373(13):1241–1249.
10. Barnes PJ. Asthma-COPD overlap. *Chest*. 2016;149(1):7–8.
11. Ford ES, Murphy LB, Khavjou O, et al. Total and state-specific medical and absenteeism costs of COPD among adults aged ≥18 years in the United States for 2010 and projections through 2020. *Chest*. 2015;147(1):31–45.
12. Criner GJ, Bourbeau J, Diekemper RL, et al. Executive summary: prevention of acute exacerbation of chronic obstructive pulmonary disease: American College of Chest Physicians and Canadian Thoracic Society Guideline. *Chest*. 2015;147(4)883–893.
13. Lamprecht B, McBurnie MA, Vollmer WM, et al. COPD in never smokers: results from the population-based burden of obstructive lung disease study. *Chest*. 2011;139(4):752–763.
14. Tan WC, Sin DD, Bourbeau J, et al. Characteristics of COPD in never-smokers and ever-smokers in the general population: results from the CanCOLD study. *Thorax*. 2015;70(9): 822–829.
15. Thomsen M, Nordestgaard BG, Vestbo J, et al. Characteristics and outcomes of chronic obstructive pulmonary disease in never smokers in Denmark: a prospective population study. *Lancet Respir Med*. 2013;1(7):543–550.
16. Thun MJ, Carter BD, Feskanich D, et al. 50-year trends in smoking-related mortality in the United States. *N Engl J Med*. 2013;368(4):351–364.
17. Kaplan RC, Bangdiwala SI, Barnhart JM, et al. Smoking among US Hispanic/Latino adults: the Hispanic community health study/study of Latinos. *Am J Prev Med*. 2014;46(5): 496–506.
18. Lange P, Celli B, Agusti A, et al. Lung-function trajectories leading to chronic obstructive pulmonary disease. *N Engl J Med*. 2015;373(2):111–122.
19. Allinson JP, Hardy R, Donaldson GC, et al. Combined impact of smoking and early-life exposures on adult lung function trajectories. *Am J Respir Crit Care Med*. 2017;196(8): 1021–1030.
20. Tan WC, Bourbeau J, Hernandez P, et al. Exacerbation-like respiratory symptoms in individuals without chronic obstructive pulmonary disease: results from a population-based study. *Thorax*. 2014;69(8):709–717.
21. Oelsner EC, Hoffman EA, Folsom AR, et al. Association between emphysema-like lung on cardiac computed tomography and mortality in persons without airflow obstruction: a cohort study. *Ann Intern Med*. 2014;161(12): 863–873.
22. Oelsner EC, Carr JJ, Enright PL, et al. Per cent emphysema is associated with respiratory and lung cancer mortality in the general population: a cohort study. *Thorax*. 2016;71(7): 624–632.
23. Woodruff PG, Barr RG, Bleecker E, et al. Clinical significance of symptoms in smokers with preserved pulmonary function. *N Engl J Med*. 2016;374(19):1811–1821.
24. Eisner MD, Blanc PD, Omachi TA, et al. Socioeconomic status, race and COPD health outcomes. *J Epidemiol Community Health*. 2011;65(1):26–34.

25. Forno E, Celedón JC. Health disparities in asthma. *Am J Respir Crit Care Med*. 2012;185(10):1033–1035.
26. LaVange L, Davis SM, Hankinson J, et al. Spirometry reference equations from the HCHS/SOL (Hispanic Community Health Study/Study of Latinos). *Am J Respir Crit Care Med*. 2017;196(8):993–1003.
27. Abrams TE, Vaughan-Sarrazin M, Fan VS, et al. Geographic isolation and the risk for chronic obstructive pulmonary disease-related mortality: a cohort study. *Ann Intern Med*. 2011;155(2):80–86.
28. Holt JB, Zhang X, Presley-Cantrell L, et al. Geographic disparities in chronic obstructive pulmonary disease (COPD) hospitalization among Medicare beneficiaries in the United States. *Int J Chron Obstruct Pulmon Dis*. 2011;6:321–328.
29. National Heart, Lung, and Blood Institute. *Morbidity & Mortality: 2012 Chart Book on Cardiovascular, Lung, and Blood Diseases*. Bethesda, MD: National Institutes of Health; 2012.
30. Akinbami LJ, Moorman JE, Bailey C, et al. Trends in asthma prevalence, health care use, and mortality in the United States, 2001–2010. *NCHS Data Brief*. 2012;(94):1–8.
31. Akinbami LJ, Moorman JE, Liu X. Asthma prevalence, health care use, and mortality: United States, 2005–2009. *Natl Health Stat Report*. 2011(32):1–14.
32. Camargo CA, Jr., Budinger GR, Escobar GJ, et al. Promotion of lung health: NHLBI Workshop on the Primary Prevention of Chronic Lung Diseases. *Ann Am Thorac Soc*. 2014;11(suppl 3):S125–S138.
33. Schneeweiss S, Avorn J. A review of uses of health care utilization databases for epidemiologic research on therapeutics. *J Clin Epidemiol*. 2005;58(4):323–337.
34. Manolio TA, Weinmann GG, Buist AS, et al. Pulmonary function testing in population-based studies. *Am J Respir Crit Care Med*. 1997;156(3):1004–1010.
35. Kohansal R, Soriano JB, Agusti A. Investigating the natural history of lung function: facts, pitfalls, and opportunities. *Chest*. 2009;135(5):1330–1341.
36. Psaty BM, O'Donnell CJ, Gudnason V, et al. Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium: design of prospective meta-analyses of genome-wide association studies from 5 cohorts. *Circ Cardiovasc Genet*. 2009;2(1):73–80.
37. Musunuru K, Lettre G, Young T, et al. Candidate gene association resource (CARe): design, methods, and proof of concept. *Circ Cardiovasc Genet*. 2010;3(3):267–275.
38. NHLBI Trans-Omics for Precision Medicine Whole Genome Sequencing Program: Data Coordinating Center (DCC). https://www.nhlbiwgs.org/group/dcc. Accessed April 13, 2018.
39. Cross-Cohort Collaboration Consortium. https://chs-nhlbi.org/node/6539. Updated March 2, 2018. Accessed March 14, 2018.
40. Taylor HA, Jr., Wilson JG, Jones DW, et al. Toward resolution of cardiovascular health disparities in African Americans: design and methods of the Jackson Heart Study. *Ethn Dis*. 2005;15(4 suppl 6):S6-4–S6-17.
41. The Atherosclerosis Risk in Communities (ARIC) Study: design and objectives. The ARIC investigators. *Am J Epidemiol*. 1989;129(4):687–702.
42. Friedman GD, Cutter GR, Donahue RP, et al. CARDIA: study design, recruitment, and some characteristics of the examined subjects. *J Clin Epidemiol*. 1988;41(11):1105–1116.
43. Fried LP, Borhani NO, Enright P, et al. The Cardiovascular Health Study: design and rationale. *Ann Epidemiol*. 1991;1(3):263–276.
44. Feinleib M, Kannel WB, Garrison RJ, et al. The Framingham Offspring Study. Design and preliminary data. *Prev Med*. 1975;4(4):518–525.
45. Goodpaster BH, Carlson CL, Visser M, et al. Attenuation of skeletal muscle and strength in the elderly: the Health ABC Study. *J Appl Physiol (1985)*. 2001;90(6):2157–2165.
46. Sorlie PD, Avilés-Santa LM, Wassertheil-Smoller S, et al. Design and implementation of the Hispanic Community Health Study/Study of Latinos. *Ann Epidemiol*. 2010;20(8):629–641.
47. Bild DE, Bluemke DA, Burke GL, et al. Multi-Ethnic Study of Atherosclerosis: objectives and design. *Am J Epidemiol*. 2002;156(9):871–881.
48. Lee ET, Welty TK, Fabsitz R, et al. The Strong Heart Study. A study of cardiovascular disease in American Indians: design and methods. *Am J Epidemiol*. 1990;132(6):1141–1155.
49. NHLBI Pooled Cohorts Study, Columbia University Department of Medicine. Data Dictionary and Harmonization Protocol for the NHLBI Pooled Cohorts Study. http://columbiamedicine.org/pcs/files/Public_Web_Resource1.pdf. Published April 10, 2018. Accessed April 10, 2018.
50. Miller MR, Hankinson J, Brusasco V, et al. Standardisation of spirometry. *Eur Respir J*. 2005;26(2):319–338.
51. Hankinson JL, Kawut SM, Shahar E, et al. Performance of American Thoracic Society-recommended spirometry reference values in a multiethnic sample of adults: the multi-ethnic study of atherosclerosis (MESA) lung study. *Chest*. 2010;137(1):138–145.
52. ATS statement—Snowbird workshop on standardization of spirometry. *Am Rev Respir Dis*. 1979;119(5):831–838.
53. Gardner RM. Standardization of spirometry: a summary of recommendations from the American Thoracic Society. The 1987 update. *Ann Intern Med*. 1988;108(2):217–220.
54. Standardization of Spirometry, 1994 Update. American Thoracic Society. *Am J Respir Crit Care Med*. 1995;152(3):1107–1136.
55. Oelsner EC, Loehr LR, Henderson AG, et al. Adjudication protocol for chronic lower respiratory disease events in two population-based cohorts: the Hispanic Community Health Study/Study Of Latinos (HCHS/SOL) and the Multi-Ethnic Study of Atherosclerosis (MESA) [abstract]. *Am J Respir Crit Care Med*. 2014;189:A2941.
56. Oelsner EC, Loehr LR, Henderson AG, et al. Classifying chronic lower respiratory disease events in epidemiologic cohort studies. *Ann Am Thorac Soc*. 2016;13(7):1057–1066.
57. Smith NL, Psaty BM, Heckbert SR, et al. The reliability of medication inventory methods compared to serum levels of cardiovascular drugs in the elderly. *J Clin Epidemiol*. 1999;52(2):143–146.
58. Psaty BM, Lee M, Savage PJ, et al. Assessing the use of medications in the elderly: methods and initial experience in the Cardiovascular Health Study. The Cardiovascular Health Study Collaborative Research Group. *J Clin Epidemiol*. 1992;45(6):683–692.
59. Definition and classification of chronic bronchitis for clinical and epidemiological purposes. A report to the Medical Research Council by their Committee on the Aetiology of Chronic Bronchitis. *Lancet*. 1965;1(7389):775–779.
60. Ferris BG. Epidemiology Standardization Project (American Thoracic Society). *Am Rev Respir Dis*. 1978;118(6 Pt 2):1–120.
61. Ingram DD, Parker JD, Schenker N, et al. United States Census 2000 population with bridged race categories. National Center for Health Statistics. *Vital Health Stat 2*. 2003;(135):1–55.
62. Matthews K, Phelan J, Jones NA, et al. *2015 National Content Test Race and Ethnicity Analysis Report, Version 1.0*.

Washington, DC: US Census Bureau; 2017. https://www2.census.gov/programs-surveys/decennial/2020/program-management/final-analysis-reports/2015nct-race-ethnicity-analysis.pdf. Published February 28, 2017. Accessed March 14, 2018.
63. Tager IB, Segal MR, Speizer FE, et al. The natural history of forced expiratory volumes. Effect of cigarette smoking and respiratory symptoms. *Am Rev Respir Dis*. 1988;138(4):837–849.
64. Quanjer PH, Stanojevic S, Cole TJ, et al. Multi-ethnic reference values for spirometry for the 3-95-yr age range: the global lung function 2012 equations. *Eur Respir J*. 2012;40(6):1324–1343.
65. Rodriguez J, Jiang R, Johnson WC, et al. The association of pipe and cigar use with cotinine levels, lung function, and airflow obstruction: a cross-sectional study. *Ann Intern Med*. 2010;152(4):201–210.
66. Kaufman JD, Adar SD, Allen RW, et al. Prospective study of particulate air pollution exposures, subclinical atherosclerosis, and clinical cardiovascular disease: the Multi-Ethnic Study of Atherosclerosis and Air Pollution (MESA Air). *Am J Epidemiol*. 2012;176(9):825–837.
67. Mishra GD, Chung HF, Pandeya N, et al. The InterLACE study: design, data harmonization and characteristics across 20 studies on women's health. *Maturitas*. 2016;92:176–185.
68. Dinov ID, Heavner B, Tang M, et al. Predictive Big Data analytics: a study of Parkinson's disease using large, complex, heterogeneous, incongruent, multi-source and incomplete observations. *PLoS One*. 2016;11(8):e0157077.
69. Chandler RK, Kahana SY, Fletcher B, et al. Data collection and harmonization in HIV research: the Seek, Test, Treat, and Retain Initiative at the National Institute on Drug Abuse. *Am J Public Health*. 2015;105(12):2416–2422.
70. Elmore K, Nelson R, Gant Z, et al. Data harmonization process for creating the National Center for HIV/AIDS, Viral Hepatitis, STD, and TB Prevention Atlas. *Public Health Rep*. 2014;129(suppl 1):63–69.
71. Lewington S, Clarke R, Qizilbash N, et al. Age-specific relevance of usual blood pressure to vascular mortality: a meta-analysis of individual data for one million adults in 61 prospective studies. *Lancet*. 2002;360(9349):1903–1913.
72. Fibrinogen Studies Collaboration. Collaborative meta-analysis of prospective studies of plasma fibrinogen and cardiovascular disease. *Eur J Cardiovasc Prev Rehabil*. 2004;11(1):9–17.
73. Emerging Risk Factors Collaboration, Danesh J, Erqou S, et al. The Emerging Risk Factors Collaboration: analysis of individual data on lipid, inflammatory and other markers in over 1.1 million participants in 104 prospective studies of cardiovascular diseases. *Eur J Epidemiol*. 2007;22(12):839–869.
74. Egger M, Schneider M, Davey Smith G. Spurious precision? Meta-analysis of observational studies. *BMJ*. 1998;316(7125):140–144.
75. Stroup DF, Berlin JA, Morton SC, et al. Meta-analysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis Of Observational Studies in Epidemiology (MOOSE) group. *JAMA*. 2000;283(15):2008–2012.
76. Palla L, Higgins JP, Wareham NJ, et al. Challenges in the use of literature-based meta-analysis to examine gene-environment interactions. *Am J Epidemiol*. 2010;171(11):1225–1232.
77. Goff DC, Jr., Lloyd-Jones DM, Bennett G, et al. 2013 ACC/AHA guideline on the assessment of cardiovascular risk: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. *J Am Coll Cardiol*. 2014;63(25 Pt B):2935–2959.
78. Global Burden of Metabolic Risk Factors for Chronic Diseases Collaboration (BMI Mediated Effects), Lu Y, Hajifathalian K, et al. Metabolic mediators of the effects of body-mass index, overweight, and obesity on coronary heart disease and stroke: a pooled analysis of 97 prospective cohorts with 1.8 million participants. *Lancet*. 2014;383(9921):970–983.
79. Kivimäki M, Kuosma E, Ferrie JE, et al. Overweight, obesity, and risk of cardiometabolic multimorbidity: pooled analysis of individual-level data for 120 813 adults from 16 cohort studies from the USA and Europe. *Lancet Public Health*. 2017;2(6):e277–e285.
80. Eisen EA, Oliver LC, Christiani DC, et al. Effects of spirometry standards in two occupational cohorts. *Am Rev Respir Dis*. 1985;132(1):120–124.
81. Mannino DM, Diaz-Guzman E, Buist S. Pre- and post-bronchodilator lung function as predictors of mortality in the Lung Health Study. *Respir Res*. 2011;12:136.