SUPPLEMENTAL MATERIAL

Genetic admixture and survival in diverse populations with pulmonary arterial hypertension

Pulmonary arterial hypertension in diverse populations

Jason H. Karnes PharmD, PhD[a], Howard W. Weiner, PhD[b], Tae-Hwi Schwantes-An, PhD[c], Balaji Natarajan, MD[d], Andrew J. Sweatt, MD[e,f], Abhishek Chaturvedi, MD[g], Amit Arora, MD, MPH[h], Ken Batai, PhD[i], Vineet Nair, MD[j], Heidi E. Steiner, PSM[a], Jason B. Giles, MS[a], Jeffrey Yu, BS[a], Maryam Hosseini, MD[j], Michael W. Pauciulo, BS, MBA[k], Katie A. Lutz BS[k], Anna W. Coleman, BS[k], Jeremy Feldman, MD[l], Rebecca Vanderpool, PhD[j], Haiyang Tang, PhD[j], Joe GN Garcia, MD[j], Jason X-J. Yuan, MD, PhD[m], Rick Kittles, PhD[n], Vinicio de Jesus Perez, MD[e,f], Roham T. Zamanian, MD[e,f], Franz Rischard, DO[j], Hemant K. Tiwari PhD[b], William C. Nichols, PhD[k]*, Raymond L. Benza, MD[g]*, Ankit A. Desai, MD[o]*

*Authors supervised this work equally

[a]Department of Pharmacy Practice and Science, University of Arizona, Tucson, AZ
[b]Department of Biostatistics, University of Alabama-Birmingham, Birmingham, AL
[c]Department of Medical and Molecular Genetics, Indiana University, Indianapolis, IN
[d]Department of Cardiology, University of California Riverside School of Medicine, CA
[e]Division of Pulmonary, Allergy, and Critical Care Medicine, Stanford University School of Medicine, Stanford, CA
[f]Vera Moulton Wall Center for Pulmonary Vascular Disease, Stanford, CA
[g]Allegheny General Hospital, Pittsburgh, PA, USA
[h]Department of Epidemiology and Biostatistics, University of Arizona, Tucson, AZ
[i]Department of Surgery, University of Arizona, Tucson, AZ
[j]Department of Medicine, University of Arizona, Tucson, AZ
[k]Division of Human Genetics, Cincinnati Children's Hospital Medical Center, Department of Pediatrics, University of Cincinnati College of Medicine, Cincinnati, OH
[l]Arizona Pulmonary Specialists, Tucson, AZ
[m]Department of Medicine, University of California-San Diego, La Jolla, CA
[n]Department of Population Sciences, City of Hope Comprehensive Cancer Center, Duarte, CA
[o]Department of Medicine, Indiana University, Indianapolis, IN

Address for Correspondence:
Ankit A. Desai, MD
Department of Medicine
Krannert Institute of Cardiology
Indiana University
1800 N Capitol Ave, E371
Indianapolis, IN 46202
Email: AnkDesai@iu.edu

eMethods

Study design and patient populations.

PAH Biobank. Clinical data for participants represented cases defined as idiopathic or heritable PAH (I/H PAH) and were extracted from the NIH-funded National Biological Sample and Data Repository for Pulmonary Arterial Hypertension (PAH Biobank) from October 2012 to February 2018. In this cohort, only adults above 18 years of age were included. Patients were recruited prospectively from 36 Enrolling Centers and approved by each of the individual institution's institutional review board. All participants provided written informed consent on enrollment. Enrollment criteria included consecutively screened patients with World Symposium of Pulmonary Hypertension (WSPH) Group 1 PAH who met expanded hemodynamic criteria of mean pulmonary artery pressure (mPAP) $\geq$ 25mm Hg at rest (30 mmHg with exercise), pulmonary capillary wedge pressure (PCWP) $\leq$18 mmHg (as defined by REVEAL registry),(1) and pulmonary vascular resistance (PVR) $\geq$ 3 wood units (WU). Baseline demographic, clinical, and hemodynamic data were extracted from the available clinical data at the time of right heart catheterization (RHC) as previously described.(2) The PAH Biobank captured information on medication history use at the time of consent, including the use of PAH medications such as prostacyclin infusion analogues. Samples banked on each patient include total genomic lymphocyte DNA.

Allegheny Health Network (AHN) Cohort. Adult Group 1 PAH patients treated with endothelin receptor antagonist (ERAs) (n=1,198) were enrolled prospectively between 2007 and 2010 from 45 US and Canadian Pulmonary Hypertension (PH) centers or retrospectively from global sites participating in the Sitaxsentan To Relieve Impaired Exercise (STRIDE) trials.(3, 4)

Participants from larger AHN cohort were included in the present study only if sufficient survival data was available and if they did not overlap with the PAH Biobank samples as described in identity–by-descent (IBD) analysis in the main manuscript. PAH was defined by predefined clinical characteristics in the AHN trials. Sites enrolling patients prospectively were instructed to conform to the standard definition of PAH described at that time by the 2007 Venice classification clinical guidelines. Detailed descriptions of this cohort have been previously described.(5)

Blood samples were collected from the prospective PAH patients from various PH centers in the United States treated with bosentan and ambrisentan (these being the only available ERAs at the time of the initial study enrollment). Assessments, including vital status, a 6-minute walk distance (6MWD) test, Borg dyspnea score, functional class (FC), and standard laboratory tests were completed at baseline, prior to the initiation of ERAs, and repeated serially at pre-specified times according to protocol (STRIDE enrollees) or local clinical practice. Baseline hemodynamic data were extracted from the available clinical data at the time of RHC done as diagnostic in the STRIDE trials and at the time of enrolment, for sites that enrolled prospectively. AHN captured information on medication history use at the time of consent, including the use of PAH medications such as prostacyclin infusion analogues.

University of Arizona (UA) Cohort. The UA study protocol was in compliance with and approved by the UA Institutional Review Board (#1502660424). All participants were at least the age of 18 or older, prospectively enrolled and provided written informed consent on enrollment. Patients in this cohort included all WSPH Group 1 PAH [based on RHC with all patients exhibiting mPAP $\geq$ 25mm Hg, wedge pressure $<$ 15 mm Hg, and PVR $\geq$ 3 WU, and a diagnosis further confirmed by a dedicated PAH clinical provider] cases at the UA pulmonary hypertension clinic from 12/2011 to 01/2016. Demographic data including age, sex, and self-reported race and

ethnicity, and use of prostacyclin infusion medication (at the time of enrolment), including use of prostacyclin analogues, were extracted from the electronic medical records. Clinical testing results for PAH including six-minute walk distance (6MWD) studies and RHC were also acquired from those closest to enrolment date.

Stanford University Cohort. Adults with newly diagnosed WSPH Group 1 PAH were prospectively enrolled at Stanford University, as part of the Vera Moulton Wall Center (VMWC) Pulmonary Hypertension Database initiative (Stanford University IRB #12338). Established in 2000, the VMWC PH Database is a structured query language relational database which captures nearly 500 unique variables including demographic features, clinical data, and outcomes in a longitudinal manner. The VMWC PH Database was screened for the present study, to identify incident cases of WSPH Group 1 PAH established at Stanford University between 2008 and 2017 (treatment naïve and no formal PAH diagnosis preceding referral). All Group 1 PAH subtypes were included. PAH was hemodynamically confirmed in all cases with a mean pulmonary arterial pressure ≥25 mmHg, pulmonary arterial wedge pressure ≤15 mmHg, and pulmonary vascular resistance >3 Wood units. Each subject provided written informed consent. For the present study, patients without available self-reported race/ethnicity data were excluded. We also excluded subjects with left ventricular systolic dysfunction (ejection fraction <55%), aortic or mitral valve disease (≥moderate stenosis or regurgitation by echocardiogram), any radiographic evidence of interstitial lung disease, total lung capacity <70% of predicted (or forced vital capacity <70% when total lung capacity not available), chronic obstructive pulmonary disease/emphysema, known chronic thromboembolic disease, or active/recent malignancy (within preceding 10 years). For each subject, baseline was their index diagnostic RHC at Stanford, which corresponded with their enrolment date for this study. Hemodynamic measures were extracted for analyses. Baseline

characteristics, including demographic features (age, sex, self-reported race/ethnicity) and other clinical data (PAH subtype, six-minute walk distance, background medications), were extracted from the VMWC database when available within one month of the index right heart catheterization. The outcome of time to death or lung transplantation was captured.

NIS Database. The Healthcare Utilization Project (HCUP) is a family of databases developed through a Federal-State-Industry partnership and is sponsored by the Agency for Healthcare Research and Quality. We used the National Inpatient Sample (NIS) database, an HCUP database that is the largest, publicly available, all-payer administrative claims database of inpatient hospitalizations in the United States. The NIS database represents a random, 20% stratified sample of all inpatient hospitalizations from approximately 1000 non-federal hospitals in 46 states (representing >97% of the total US population) and includes approximately 7 to 8 million hospitalizations per year.(6)

The NIS database includes de-identified information on patient demographics, and clinical data including primary and secondary discharge diagnoses, comorbidities, and outcomes for each sampled hospitalization. All diagnoses and comorbidities are available in the NIS database as International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM) codes until September 30, 2015 and as International Classification of Diseases, Tenth Revision, Clinical Modification/Procedure Coding System (ICD-10-CM/PCS). HCUP coding includes race and ethnicity in one data element. If the source hospital supplied race and ethnicity in separate data elements, ethnicity takes precedence over race in setting the HCUP value for race including Whites, African Americans, Hispanics, Asians, Native Americans and other. Race/ethnicity in the NIS comes from the source hospital. While many hospitals gather race and ethnicity information directly from patients, hospitals use a variety of methods to record

race/ethnicity and thus NIS race/ethnicity is not considered self-reported. Some potential for misclassification of race/ethnicity exists in the NIS dataset. To be noted, not all data sources provide information on race. Since this is an analysis of publicly available de-identified data, the AHN's, Indiana University's, and Cincinnati Children's Hospital Institutional Review Board guidelines stipulate that board approval of the study and informed consent are waived.

We identified all patients in the NIS database from January 1, 2007, through December 31, 2016, with a primary discharge diagnosis of idiopathic pulmonary arterial hypertension (PAH) using an ICD-9-CM primary diagnosis code of 416.0 and ICD-10-CM/PCS primary diagnosis code of I27.0. This ICD definition has been previously used to reduce misclassification.(7, 8) These codes are specific to IPAH and do not capture patients with PAH-associated with other conditions like HIV or congenital heart disease. For each discharge record with primary diagnosis of PAH, we obtained the following variables: patient demographics, including age, sex and race; outcome measures, including inpatient length of hospitalization and inpatient-specific mortality; and Elixhauser comorbidities and comorbidity index that have been previously validated in administrative databases.(9) One Elixhauser comorbidity i.e; pulmonary circulation disorders was not included in the comparison analysis as PAH is a part of this comorbidity. We also compared distribution of few other specific inpatient comorbidities/complications pertaining to PAH patients, namely coronary artery disease, atrial fibrillation, cardiogenic shock, chronic obstructive pulmonary disease & bronchiectasis, pneumonia, acute respiratory failure, acute cerebrovascular disease.(7) These comorbidities were identified using valid ICD-9-CM and ICD-10-CM/PCS codes.(7, 10)

National estimates for total number of discharges with primary diagnosis of pulmonary arterial hypertension in 10 years (from 2007 through 2016) were generated from the NIS sample

using pre-specified weights and published HCUP methods, accounting for the complex survey design and clustering.(11) Nationally-weighted PAH population data was used to compare race-wise distribution of patients in terms of their age, sex, frequency of comorbidities, length of hospital stay and all-cause inpatient-specific mortality. Continuous variables are presented as mean ± standard error for parametric and as median (interquartile range) for non-parametric data. Dichotomous variables are presented as frequencies (%). Intergroup differences were assessed using analysis of variance test for parametric continuous variables, Kruskal-Wallis equality-of-populations rank test for non-parametric variables and Pearson's chi-square test for dichotomous variables. Univariate and multi-variable logistic regression analysis was performed to determine association between racial background and inpatient-specific mortality. All statistical analysis was completed using STATA 15.0 (StataCorp, College Station, TX, USA). P value <0.05 was considered for statistical significance.

Analysis of Race/Ethnicity and Admixture Proportions.

Self-reported Race/Ethnicity. Primary analyses related to PAH outcomes and clinical data used self-reported race/ethnicity. Participants reporting race as White with no Hispanic ethnicity or Black with no Hispanic ethnicity were considered to be Non-Hispanic White or Non-Hispanic African American (AA) patients, respectively. Patients reporting Hispanic ethnicity were considered to be Hispanic regardless of reported race. Definitions for Non-Hispanic Whites, AAs and Hispanics using self-reported race/ethnicity were consistent across the PAH Biobank, the AHN cohort, the UA cohort, and the Stanford University cohort. In the UA cohort, the lack of AA patients precluded our ability to analyze PAH data specific to non-Hispanic AAs. Due to small numbers, we were unable to investigate associations with Asian race in the AHN cohort, the PAH

Biobank, and the UA cohort. We were also unable to investigate associations with AA patients in the UA cohort. Similarly, the relatively small numbers of non-Hispanic AAs, Asians, and Hawaiian/Pacific Islanders precluded analysis of these subgroups in the Stanford University cohort.

PCA-based race/ethnicity and estimation of admixture proportions. Since high density genotyping was available for the AHN cohort and PAH Biobank, modeling of genetic ancestry was performed to exclude any misclassification inherent in self-reported race/ethnicity. In the UA Cohort, DNA samples were not collected and thus we were unable to determine STRUCTURE-defined admixture proportions and PCA-based race/ethnicity.

PAH Biobank. Primary analyses related to PAH outcomes and clinical data used self-reported race/ethnicity. Definitions for White, AA, and Hispanic using self-reported race/ethnicity were consistent with the AHN cohort. A sensitivity analyses was also performed using PCA-based race/ethnicity. Samples were genotyped using the Illumina® HumanOmni5-QUAD BeadChip array. Genotype quality control was performed as previously reported on the subset of PAH biobank patients for whom a DNA sample was available.(12) In brief, we excluded samples with a high degree of relatedness, low sample call rates (<0.95), and sex discordance between recorded clinical data and genomic data. Single nucleotide polymorphisms (SNPs) with low call rates (<0.97), 10% GenCall score ≤0.3, Het Excess >0.2, and significant deviation from Hardy Weinberg equilibrium ($p<0.001$) were excluded. Principal components were estimated using the PC_AiR function in GENESIS version 3.8 package (github.com/UW-GAC/GENESIS). Unlike regular PCA, this approach takes recent ancestry into consideration to make principal components (PCs) less biased than traditional methods. In primary analyses of PCA-based race/ethnicity, individuals belonging to African, European, Asian, and admixed ancestry groups were manually

identified using the first two PCs alongside 1000 Genome Project Reference samples.(13) As Hispanic ethnicity represent a combination of genetic, environmental, and socio-cultural factors, individuals who were considered admixed though PCA and also self-reported Hispanic ethnicity were considered Hispanic in PCA-based race/ethnicity analyses. Proportions of African, European, and Native American/East Asian ancestry were generated using 1,381 ancestry informative markers (AIMs) input into STRUCTURE (k=3) using Hapmap reference populations.(14)

AHN Cohort. Genotyping was performed on the Illumina® HumanOmniExpressExome BeadChip. Genotype quality control for genome-wide arrays was performed as previously reported.(5) For each dataset, we excluded samples with a high degree of relatedness, low sample call rates (<0.985), and sex discordance between recorded clinical data and genomic data. SNPs with low call rates (<0.95) or significant deviation from Hardy Weinberg equilibrium ($p<0.001$) were excluded. Principal components were estimated using the PC_AiR function in GENESIS version 3.8 package (github.com/UW-GAC/GENESIS). Unlike regular PCA, this approach takes recent ancestry into consideration to make PCs less biased than traditional methods. In primary analyses of PCA-based race/ethnicity, individuals belonging to African, European, Asian, and admixed ancestry groups were manually identified using the first two PCs alongside 1000 Genome Project Reference samples.(13) Individuals who were considered admixed though PCA and also self-reported Hispanic ethnicity were considered Hispanic in PCA-based race/ethnicity analyses. Samples from the AHN cohort were genotyped using 1,050 ancestry informative markers (AIMs) from the Illumina® Omni Express genome-wide platform. Proportions of African, European, and Native American/East Asian ancestry were generated using AIMs input into STRUCTURE (k=3) using reference samples from the 1000 Genome Project.(13)

Statistical analysis.

Data analysis was performed using SAS version 9.4 (SAS Institute, Inc., Cary, NC), STATA 15.0 (StataCorp, College Station, TX, USA), and R statistical software version 3.3.0.(15) Three separate measures of race/ethnicity were analysed: 1) PCA-based race/ethnicity, 2) self-reported race/ethnicity, and 3) STRUCTURE-based percent African ancestry, European ancestry, and East Asian/Native American ancestry estimates. Data analysis was performed using SAS version 9.4 (SAS Institute, Inc., Cary, NC) and R statistical software version 3.3.0.(15) Descriptive statistics by self-reported race/ethnicity are presented as mean and standard deviation. Categorical variables are presented as proportions and percentages. P values for descriptive analyses were derived from Mann-Whitney U test and Kruskal-Wallis tests. Fisher's Exact test was used for categorical variables. Multi-variable linear regression analyses were performed to test for association of self-reported race/ethnicity, PCA-based race/ethnicity, and percent African ancestry, European ancestry, and Native American ancestry estimates (STRUCTURE-based) with disease characteristics (pulmonary vascular resistance [PVR], right ventricular [RV] power, six minute walk distance [6MWD], and mean right atrial pressure [mRAP]). Ternary plots included in Figure 1 were generated using R with confidence intervals (50, 95, 99) calculated using the Mahalnobis Distance and Log-Ratio Transformation (https://cran.r-project.org/web/packages/ggtern/ggtern.pdf).

Linear regression models for all cohorts were adjusted for age, sex, use of prostacyclin analogues, and log(PVR). Log(PVR) was used in regressions to ensure normality of the PVR adjustment variable. PVR was used in all models (except models where PVR was the dependent variable). For multi-variable regressions, PVR was transformed to its natural logarithm before use

in the regression models. We adjusted for prostacyclin analogue use based on both, differences in use in univariate analyses across the three populations and its significant effects on hemodynamics, functional status, and mortality and previously reported data on differences on medication use across minority populations.(2) Use of prostacyclin analogues were assessed at time of enrolment and hemodynamic characteristics, including log(PVR), were assessed at time of RHC. In analyses of influence of race/ethnicity on hemodynamic characteristics, Hispanics and AAs were compared to whites as the baseline comparator group. $P \leq 0.05$ was set as threshold for statistical significance. No adjustment for principal components was employed since regressions tested the influence of race/ethnicity rather than the influence of genetic variation independent of race/ethnicity.

All-cause, transplant-free mortality was used as the primary endpoint in survival analyses using Kaplan-Meier estimates, Logrank tests, and Cox proportional hazards regression. The proportional hazards assumption was tested for Cox modesl using plotting of Martingale residuals using proc phreg in SAS (SAS Institute, Inc., Cary, NC). Survival was calculated from enrolment in the PAH biobank or AHN cohort to date of death or censoring. If a patient received a lung transplant, this was considered an event with date of lung transplantation as the end survival time. In the PAH Biobank, January 1, 2018 was used as the date of censoring. In the AHN cohort, if a specific date for last contact was not recorded, the date 12/19/2014 was used as the censoring date as previously reported.(5) In the UA cohort, the censor date was patient-specific and reflected the date of acquisition of mortality status. In all UA cohort patients, the censor data was between 9/26/18 and 10/22/18. In the Stanford University cohort, survival was calculated from the date of PAH diagnosis (index right heart catheterization) to the date of death, lung transplantation, or censoring (five years from PAH diagnosis, or last known clinical encounter prior to 9/1/19 among subjects without five-year follow-up). As in linear regression models, age, sex, use of prostacyclin

analogues, and log(PVR) were included as covariates for hazard ratio (HR) estimates in Cox regressions. Stanford University cohort analyses did not adjust for prostacyclin use, as these newly diagnosed patients were treatment naïve.

Multiple sensitivity analyses were performed, including analyses to account for left truncation time. Left truncation time was defined as time from PAH diagnosis to study enrolment. This time was used for entry in to the risk set for Cox models (left truncation analysis) and Cox models adjusted for left truncation time in addition to other covariates. Cox models using time from PAH diagnosis to death/censoring were also used. Sensitivity analyses were also performed to test the influence of lung transplantation on results, given the potential for disparities with lung transplantation. These analyses included Cox models that completely excluded patients receiving a lung transplant and Cox competing risk model for mortality and lung transplant. In addition, Cox models were run in IPAH patients alone and excluding patients with *BMPR2* mutation(s). In order to test consistent effects of age, cohorts were stratified into upper and lower 50^{th} percentiles of age based on the median age and separate Cox regressions were performed in each group.

Meta-analysis of cox proportional hazards regression was performed using the rmeta package version 2.16 in R. A fixed effects model was used when a non-significant test for heterogeneity was observed. Inverse standard error was used as a weight for each study. Meta-analysis did not include NIS database results due to differences in inclusion criteria, follow-up time, statistical model, and use of logistic rather than Cox proportional hazards regression.

eRESULTS

Genetic data was available for all AHN cohort patients and 994 PAH Biobank patients. In the AHN cohort, NHWs had a high average European ancestry (96%) and AAs had high average African ancestry (90%) (Figure 1A and eTable S3). Hispanics were admixed with average 50% European, 2% African, and 47% Native American ancestry. Ancestry results were similar in the PAH biobank, with NHWs having a high average European ancestry (97%), AAs having high average African ancestry (82%), and Hispanics being admixed with 58% European, 7% African, and 36% Native American ancestry (Figure 1B and eTable S2). AAs did not have significantly different survival versus NHWs in the PAH Biobank (HR 1.19[0.76-1.86], p=0.454) and the AHN cohort (HR 0.91[0.59-1.38, p=0.650).

Sensitivity Analyses.

In sensitivity analyses with PCA-defined race/ethnicity (eFigures S3-S6), we observed non-significant indications of mortality benefit with Hispanics in the PAH Biobank (HR 0.44 [0.18-1.09], p=0.077) and the AHN cohort (HR 0.55 [0.29-1.06], p=0.074). Meta-analysis of these results indicated a significant effect of PCA-defined Hispanic ethnicity (HR 0.51 [0.30-0.87], p=0.013) (eFigure S2). Meta-analysis indicated a lack of heterogeneity for both self-reported race/ethnicity (heterogeneity 0.39, p=0.824) and PCA-based race/ethnicity (heterogeneity 0.15, p=0.697). In the PAH Biobank, the use of RHC diagnostic date rather than enrollment date to account for left truncations times resulted in a significant effect in the PAH Biobank (HR 0.46 [0.21-0.99], p=0.047). Meta-analysis of all cohorts using diagnostic RHC date also resulted in a significant survival benefit for Hispanics (HR 0.59 [0.40-0.87], p=0.007). When patients with *BMPR2* mutation were excluded from PAH Biobank, a consistent mortality benefit for Hispanics was observed versus NHWs (HR 0.43 [0.19-0.99], p=0.49).

Survival Analyses for Proportions of ancestry.

We observed a non-significant indication of mortality benefit with increasing Native American ancestry in the AHN cohort (HR 0.46 [0.18-1.18], p=0.104) and non-significant results from the PAH biobank seemed to corroborate this relationship (HR 0.51 [0.15-1.77], p=0.289). Meta-analysis suggested that the observed mortality benefit in Hispanics was partly due to decreased mortality with increasing Native American ancestry (HR 0.48 [0.23-1.01], p=0.053) (eFigure S2). No significant associations were observed between European or African ancestry and mortality.

Influence of Race/Ethnicity and Admixture on Hemodynamics and Functional Outcomes in PAH Registries

While a significant association was not consistently observed across all cohorts, self-reported Hispanic ethnicity was associated with significantly increased PVR and decreased 6MWD (eTable S6). AAs had significantly reduced 6MWD in both the PAH Biobank and AHN Cohort. Percent African ancestry was strongly associated with reduced 6MWD in the AHN cohort and PAH Biobank (eTable S7). Increasing African ancestry was also strongly associated with reductions in PVR. Conversely, increasing Native American ancestry was strongly and consistently associated with increased PVR. A strong positive association between European ancestry and 6MWD was observed in the AHN cohort with a consistent trend in the PAH Biobank.

Table E1. Comparison of survival characteristics by cohort

Cohort	Trial Duration	PAH Etiology[a]	Survival Time Start Date (primary analysis)	Censor date(s)	Outcome	Statistical Model (primary analysis)	Race Definition(s)
PAH Biobank	2012-2018	IPAH/ HPAH	Enrolment	Date of acquisition of mortality status (1/1/2018)	all cause death or lung transplantation	age, sex, use of prostacyclin analogues, and log(PVR)	Self-reported; PCA-based[b]
AHN	2007-2010	Group 1 PAH patients treated ERAs	Enrolment	Last contact or 12/19/2014	all cause death or lung transplantation	age, sex, use of prostacyclin analogues, and log(PVR)	Self-reported; PCA-based[b]
UA	2011-2016	All Group 1 PAH	Enrolment	Date of acquisition of mortality status (patient-specific)	all cause death or lung transplantation[c]	age, sex, use of prostacyclin analogues, and log(PVR)	Self-reported
Stanford	2008-2017	Newly diagnosed Group 1 PAH	Enrolment / date of PAH diagnosis (index right heart catheterization)	5 years from PAH diagnosis or last known clinical encounter prior to 9/1/2019	all cause death or lung transplantation	age, sex, and log(PVR)[d]	Self-reported

AHN indicates Allegheny Health Network; ERA endothelin receptor agonist; PAH, pulmonary arterial hypertension; PCA, principal components analysis; PVR, pulmonary vascular resistance; UA, University of Arizona.

[a]All Group 1 PAH classifications based on World Symposium of Pulmonary Hypertension (WSPH)

[b]Principal components were estimated using the PC_AiR function in GENESIS version 3.8 package. Individuals belonging to African, European, Asian, and admixed ancestry groups were manually identified using the first two PCs alongside 1000 Genome Project Reference samples. Individuals who were considered admixed though PCA and also self-reported Hispanic ethnicity were considered Hispanic in PCA-based race/ethnicity analyses.

[c]No lung transplantation events occurred in the UA cohort.

[d]The Stanford cohort was treatment-naïve and were not treated with prostacyclin analogues at baseline

Table E2: Demographic, clinical and hemodynamic characteristics of the PAH Biobank cohort by self-reported race/ethnicity

Characteristic (units)[a]	Non-Hispanic Whites (n=1039)	Non-Hispanic AAs (n=158)	Hispanics (n=129)	P value[b]
Age (years)	53 ± 18	52 ± 17	42 ± 16	<0.001
Female sex (n)	786 (76%)	134 (85%)	111 (86%)	0.002
Body Surface Area (m^2)	1.91 ± 1.52	1.88 ± 0.34	1.78 ± 0.36	0.06
Prostacylin analogs (n)	284 (27%)	34 (22%)	24 (19%)	0.04
PAH Etiology (n)				<0.001
Idiopathic	899 (87%)	152 (96%)	119 (92%)	-
Heritable	140 (13%)	6 (4%)	10 (8%)	-
Ancestry estimates[c]				
African	0.01	0.82	0.07	<0.001
European	0.97	0.16	0.58	<0.001
Native American	0.02	0.02	0.36	<0.001
mRAP (mm Hg)	9.27 ± 5.65	10.2 ± 5.91	9.87 ± 7.97	0.29
mPAP (mm Hg)	52.2 ± 14.1	47.7 ± 13.6	56.1 ± 15.8	<0.001
PCWP (mm Hg)	10.1 ± 3.99	10.3 ± 3.79	9.34 ± 3.34	0.76
CI ($l/min/m^2$)	2.52 ± 1.26	2.34 ± 0.93	2.12 ± 0.72	0.19
PVR (WU)	10.93 ± 6.54	10.2 ± 6.40	13.6 ± 8.12	0.001
6MWD (m)	356 ± 138	331 ± 114	378 ± 99	0.13
Time from diagnosis to enrolment (years)	6.90 ± 6.05	6.35 ± 6.53	7.14 ± 6.22	0.47

6MWD indicates six minute walk distance; AA, African American; CI, cardiac index; mPAP, mean pulmonary arterial pressure; mRAP, mean right atrial pressure; PAH, pulmonary arterial hypertension; PCWP, pulmonary capillary wedge pressure; PVR, pulmonary vascular resistance; RV, right ventricular; WU, Wood units.

[a]Continuous data are presented by self-reported race/ethnicity as mean and standard deviation. Categorical variables are presented as numbers and percentages.

[b]Differences in baseline demographic and PAH clinical traits between the racial/ethnic groups were compared using non-parametric Kruskal Wallis test and Fisher's Exact test for frequency differences in categorical traits.

[c]Ancestry estimates were generated using ancestry informative markers (AIMs) input into STRUCTURE using Hapmap reference populations

Table E3: Demographic, clinical and hemodynamic characteristics for the AHN cohort by self-reported race/ethnicity

Characteristic (units)[a]	Non-Hispanic White (n=656)	Non-Hispanic AAs (n=97)	Hispanic (n=69)	P value[b]
Age (years)	52 ± 16	51 ± 14	36 ± 11	<0.001
Female sex (n)	523 (79%)	90 (93%)	60 (87%)	<0.001
Body Surface Area (m^2)	1.86 ± 0.26	1.92 ± 0.27	1.68 ± 0.19	<0.001
Prostacylin analogs (n)	87 (13%)	19 (20%)	3 (4%)	0.01
PAH Etiology (n)				0.64
Idiopathic/Heritable	346 (53%)	46 (48%)	36 (53%)	
Associated	305 (47%)	50 (52%)	32 (47%)	
Ancestry estimates[c]				
African	0.02	0.82	0.02	<0.001
European	0.97	0.17	0.50	<0.001
Native American	0.02	0.01	0.47	<0.001
mRAP (mm Hg)	8.66 ± 5.29	10.41 ± 8.04	10.09 ± 12.34	0.25
mPAP (mm Hg)	51.2 ± 16.9	45.8 ± 15.2	66.0 ± 20.0	<0.001
PCWP (mm Hg)	10.52 ± 5.22	10.91 ± 4.48	8.94 ± 4.16	0.02
CI ($l/min/m^2$)	2.61 ± 1.34	2.39 ± 0.69	3.17 ± 1.41	0.003
PVR (WU)	10.91 ± 6.80	8.58 ± 5.25	14.32 ± 9.55	<0.001
6MWD (m)	353 ± 128	297 ± 107	311 ± 96	<0.001
Time from diagnosis to enrolment (years)	5.58 ± 7.23	3.32 ± 3.32	5.08 ± 5.52	<0.001

6MWD indicates six minute walk distance; AA, African American; AHN, Allegheny Health Network; CI, cardiac index; mPAP, mean pulmonary arterial pressure; mRAP, mean right atrial pressure; PAH, pulmonary arterial hypertension; PCWP, pulmonary capillary wedge pressure; PVR, pulmonary vascular resistance; RV, right ventricular; WU, Wood units.

[a]Continuous data are presented by self-reported race/ethnicity as mean and standard deviation. Categorical variables are presented as numbers and percentages.

[b]Differences in baseline demographic and PAH clinical traits between the racial/ethnic groups were compared using non-parametric Kruskal Wallis test and Fisher's Exact test for frequency differences in categorical traits.

[c]Ancestry estimates were generated using ancestry informative markers (AIMs) input into STRUCTURE using Hapmap reference populations

Table E4. Demographic, clinical and hemodynamic characteristics of the UA Cohort by self-reported race/ethnicity

Characteristic (units)[a]	Non-Hispanic Whites (n=135)	Hispanics (n=29)	P value[b]
Age (years)	61 ± 13	51 ± 10	<0.001
Female sex (n)	97 (72)	22 (76)	0.82
Body Surface Area (m^2)	1.89 ± 0.27	1.95 ± 0.36	0.74
Prostacyclin Analogues (n)	46 (35%)	8 (31%)	0.82
PAH Etiology (n)			0.12
Idiopathic	61 (47%)	8 (30%)	
Heritable	1 (1%)	1 (4%)	
Associated	69 (53%)	18 (67%)	
mRAP (mm Hg)	7.12 ± 4.44	10.9 ± 7.25	0.01
mPAP (mm Hg)	37.7 ± 15.4	44.2 ± 14.9	0.03
PVR (WU)	5.39 ± 3.34	6.81 ± 4.29	0.05
PCWP (mm Hg)	9.50 ± 3.68	10.5 ± 3.20	0.14
CI ($l/min/m^2$)	3.08 ± 0.91	3.20 ± 1.26	0.94
6MWD (m)	333 ± 138	331 ± 107	0.83

6MWD indicates six minute walk distance; BP, blood pressure; CI, cardiac index; mPAP, mean pulmonary arterial pressure; mRAP, mean right atrial pressure; PAH, pulmonary arterial hypertension; PCWP, pulmonary capillary wedge pressure; PVR, pulmonary vascular resistance; RV, right ventricular; UA, University of Arizona; WU, Wood units.

[a]Continuous data are presented by self-reported race/ethnicity as mean and standard deviation. Categorical variables are presented as numbers and percentages.

[b]Differences in baseline demographic and PAH clinical traits between the racial/ethnic groups were compared using non-parametric Wilcoxon Rank Sum test and Fisher's Exact test for frequency differences in categorical traits.

Table E5. Demographic, clinical and hemodynamic characteristics of the Stanford PAH cohort (newly diagnosed) by self-reported race/ethnicity

Characteristic (units)[a]	Non-Hispanic Whites (n=140)	Hispanics (n=63)	P value[b]
Age (years)	53.0 ± 14.4	44.8 ± 12.9	0.0002
Female sex (n)	105 (75.0%)	45 (71.4%)	0.607
Body Surface Area (m^2)	1.90 ± 0.23	1.82 ± 0.24	0.024
PAH Etiology (n)			0.225
Idiopathic	21 (15%)	13 (21%)	
Heritable	5 (4%)	0 (0%)	
Associated	114 (81%)	50 (79%)	
mRAP (mm Hg)	9.26 ± 5.11	9.92 ± 6.19	0.849
mPAP (mm Hg)	50.7 ± 13.9	53.3 ± 13.9	0.209
PVR (Wood units)	12.4 ± 6.3	12.1 ± 6.2	0.752
PCWP (mm Hg)	10.1 ± 4.1	11.2 ± 4.3	0.088
CI ($L/min/m^2$)	2.04 ± 0.77	2.18 ± 0.67	0.063
6MWD (m)	405.0 ± 147.1	371.4 ± 144.8	0.049

6MWD indicates six-minute walk distance; BP, blood pressure; mPAP, mean pulmonary arterial pressure; mRAP, mean right atrial pressure; PCWP, pulmonary capillary wedge pressure; PVR, pulmonary vascular resistance; RV, right ventricular; WU, Wood units.

[a]Continuous data are presented by self-reported race/ethnicity as mean and standard deviation. Categorical variables are presented as numbers and percentages.

[b]Differences in baseline demographic and PAH clinical traits between the racial/ethnic groups were compared using non-parametric Wilcoxon Rank Sum test and Fisher's Exact test for frequency differences in categorical traits.

Table E6: Association between self-reported race/ethnicity and PVR and functional outcomes by study cohort

	Hispanic		Non-Hispanic AA	
Characteristic[a]	Beta (SE)	P value	Beta (SE)	P value
AHN Cohort				
PVR (WU)	0.96 (0.97)	0.324	-2.84 (0.78)	0.0003
6MWD (m)	-82.88 (15.34)	<0.0001	-61.30 (13.38)	<0.0001
PAH Biobank				
PVR (WU)	1.25 (0.62)	0.046	-0.72 (0.55)	0.184
6MWD (m)	-9.93 (18.64)	0.595	-32.30 (15.67)	0.039
UA Cohort				
PVR (WU)	1.52 (0.77)	0.05	-	-
6MWD (m)	-20.0 (32.6)	0.54	-	-
Stanford Cohort				
PVR (WU)	0.996 (0.971)	0.306		
6MWD	-59.93 (22.02)	0.007		

6MWD indicates six minute walk distance; AA, African American, PAH, pulmonary arterial hypertension; PVR, pulmonary vascular resistance; SE, standard error; WU, wood units.
[a]Linear regressions for self-identified Hispanic and Non-Hispanic African American patients were compared to patients of European descent as a reference group. Models included adjustment variables age, sex, prostacyclin use, and log(PVR). Regressions for PVR did not include adjustments for log(PVR).

Table E7: Association between proportions of admixture and PVR and functional outcomes by study cohort

	African ancestry (%)[b]		European ancestry (%)[b]		Native American ancestry (%)[b]	
PAH severity index[a]	Beta (SE)	P value	Beta (SE)	P value	Beta (SE)	P value
AHN Cohort						
PVR (WU)	-3.02 (0.95)	0.002	1.13 (0.88)	0.199	5.82 (1.77)	0.001
6MWD (m)	-64.39 (15.90)	<0.0001	90.58 (14.33)	<0.0001	-152.27 (29.60)	<0.0001
PAH Biobank						
PVR (WU)	-2.05 (0.85)	0.016	0.40 (0.74)	0.585	3.19 (1.24)	0.010
6MWD (m)	-51.19 (22.36)	0.022	33.73 (19.56)	0.085	17.08 (35.12)	0.627

6MWD indicates six minute walk distance; PAH, pulmonary arterial hypertension; PVR, pulmonary vascular resistance; SE, standard error; WU, Wood units.

[a]Linear regressions for effect of estimated percent ancestry on hemodynamic parameters and functional outcomes were performed in the entire genotyped cohort for the AHN cohort and the PAH Biobank. Models included adjustment variables age, sex, prostacyclin use, and log(PVR). Regressions for PVR did not include adjustments for log(PVR).

[b]Proportions of African, European, and Native American ancestry were generated using ancestry informative markers input into STRUCTURE (k=3) using Hapmap reference populations.

Table E8: Inpatient Characteristics of Patients with Primary Diagnosis of Pulmonary Arterial Hypertension Based on Race/ethnicity

Characteristic, units[a]	Non-Hispanic White	Non-Hispanic AA	Hispanic	Native American[b]	Non-Hispanic Asian	P value
PAH discharges, n	8829	2628	1524	185	403	**<0.0001**
Age, years	55.7 ± 0.7	51.7 ± 1.0	42.4 ± 1.4	43.4	41.5 ± 3.0	**<0.0001**
Female Sex, %	74.3	78.9	78.2	80.5	67.4	**<0.0001**
Outcome Measures						
All-Cause inpatient-specific mortality, %	6.9	8.7	4.4	2.7	9.9	**<0.0001**
Length of stay, days	5 (3-9)	6 (3-11)	5 (3-9)	4 (3-8)	6 (3-11)	**0.0001**
Elixhauser Comorbidities[c]						
Congestive heart failure, %	53.4	60.4	44.8	57.8	49.5	**<0.0001**
Valvular heart disease, %	20.8	19.6	21.1	13.6	26.1	**0.008**
Cardiac arrhythmia, %	31.6	25.3	19.4	26.6	34.4	**<0.0001**
Peripheral vascular disease, %	4.7	4.5	1.9	2.7	1.2	**<0.0001**
Essential hypertension, %	24.7	30.8	20.8	24.4	12.1	**<0.0001**
Complicated/Secondary hypertension, %	16.4	24.5	10.2	5.2	21.5	**<0.0001**
Diabetes without complications, %	20.2	20.8	10.1	27.2	7	**<0.0001**
Diabetes with complications, %	5.9	9	4.5	2.7	5.7	**<0.0001**
Chronic pulmonary disease, %	37.2	38.9	24.4	26.5	27.9	**<0.0001**
Renal failure, %	20	24.8	12.4	13.6	21.5	**<0.0001**
Fluid and electrolytes disorders, %	31.4	33.9	28.8	27.6	37.3	**0.001**
Coagulopathy, %	12.3	8.3	13.2	8.1	15.9	**<0.0001**
Blood loss anemia, %	0.8	0.7	1	0	0	0.156
Deficiency anemia, %	5.8	10.1	8.3	5.4	6.2	**<0.0001**
Obesity, %	18.5	22.9	14.7	21.2	8.5	**<0.0001**
Weight loss, %	5	5.9	4.8	0	3.6	**0.004**
Hypothyroidism, %	19.3	9.4	14.5	18.7	18.3	**<0.0001**

Table E8: Inpatient Characteristics of Patients with Primary Diagnosis of Pulmonary Arterial Hypertension Based on Race/ethnicity (continued)

Characteristic[a]	White	African American	Hispanic	Native American[b]	Asian	P value
Liver disease, %	7.6	7.5	10.4	11.2	4.7	**<0.0001**
Peptic ulcer disease excluding bleeding, %	0.6	0.7	0	0	1.2	**0.006**
HIV and AIDS, %	0.3	0.6	0.7	0	0	**0.019**
Paralysis, %	0.4	0.9	0	0	0	**<0.0001**
Other neurological disorder, %	4.3	4	2.7	0	1.2	**<0.0001**
Lymphoma, %	0.8	0.9	0.4	0	0	0.079
Metastatic cancer, %	0.5	0.5	0.7	0	1.2	0.199
Solid tumor without metastasis, %	1	1.3	0.3	2.9	1.2	**0.006**
Rheumatoid arthritis/Collagen vascular disease, %	8.8	11.8	7.4	2.7	6.1	**<0.0001**
Alcohol abuse, %	2.6	1.3	3.5	5.4	0	**<0.0001**
Drug abuse, %	3.3	3.1	5.5	5.1	2.5	**<0.0001**
Psychoses, %	0.6	0.9	1. 0	0	0	**0.037**
Depression, %	14.7	10.1	13.8	13.7	7.5	**<0.0001**
Elixhauser Comorbidity Index[c]	5 (3-6)	5 (3-6)	4 (2-6)	4 (3-6)	4 (2-6)	**0.0001**
Other Comorbidities						
Coronary artery disease, %	21.6	20.5	8.5	0	8	**<0.0001**
Atrial fibrillation , %	20	12.2	11.2	13.4	18.5	**<0.0001**
Cardiogenic shock, %	2.8	2.8	0.9	0	3.7	**<0.0001**
Chronic obstructive pulmonary disease-Bronchiectasis, %	24.1	23.9	7.3	10.7	9.7	**<0.0001**
Pneumonia, %	6.5	6.2	7.2	0	9.6	**<0.0001**
Acute respiratory failure, %	18.3	19.3	10.3	13.1	13.1	**<0.0001**
Acute cerebrovascular disease, %	0.5	0.8	0	0	0	**0.005**

IQR indicates interquartile range; PAH, Pulmonary arterial hypertension; SE, Standard error.

[a]Total number of PAH discharges are expressed as N. Continuous variables are expressed as mean ± SE and median IQR for parametric and non-parametric variables respectively. Categorical variables are expressed as frequencies (%). Group-wise comparisons were made using analysis of variance test for parametric continuous variables, Kruskal-Wallis equality-of-populations rank test for non-parametric variables and chi-square test for categorical variables.

[b]Missing standard error because of stratum with single sampling unit. PAH discharges in all 5 races do not add to make a total of 16151 due to race being unknown or missing data.
[c]Pulmonary circulation disorder is an Elixhauser comorbidity that was not included in the analysis because PAH comes under this broad category of comorbidity.

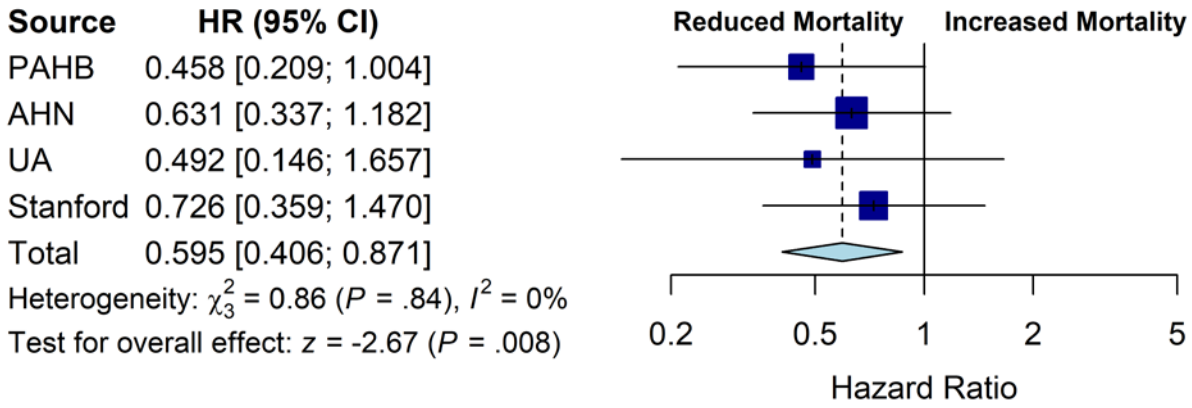


Figure E1. Global meta-analysis of all-cause mortality hazard ratios for self-reported race/ethnicity in the PAHB, AHN, UA, and Stanford cohorts. Cox proportional hazards regression for all-cause mortality was performed in Group 1 PAH patients adjusted for age, sex, prostacyclin use, and log(PVR). Meta-analysis of cox proportional hazards regression used a fixed effects model and inverse standard error was used as a weight for each study. 95% CI indicates 95% confidence interval; AHN, Allegheny Health Network; HR, hazard ratio; NHW, Non-Hispanic Whites; PAH, pulmonary arterial hypertension; PAHB, PAH Biobank; PCA, principal components analysis; PVR, pulmonary vascular resistance; UA, University of Arizona.

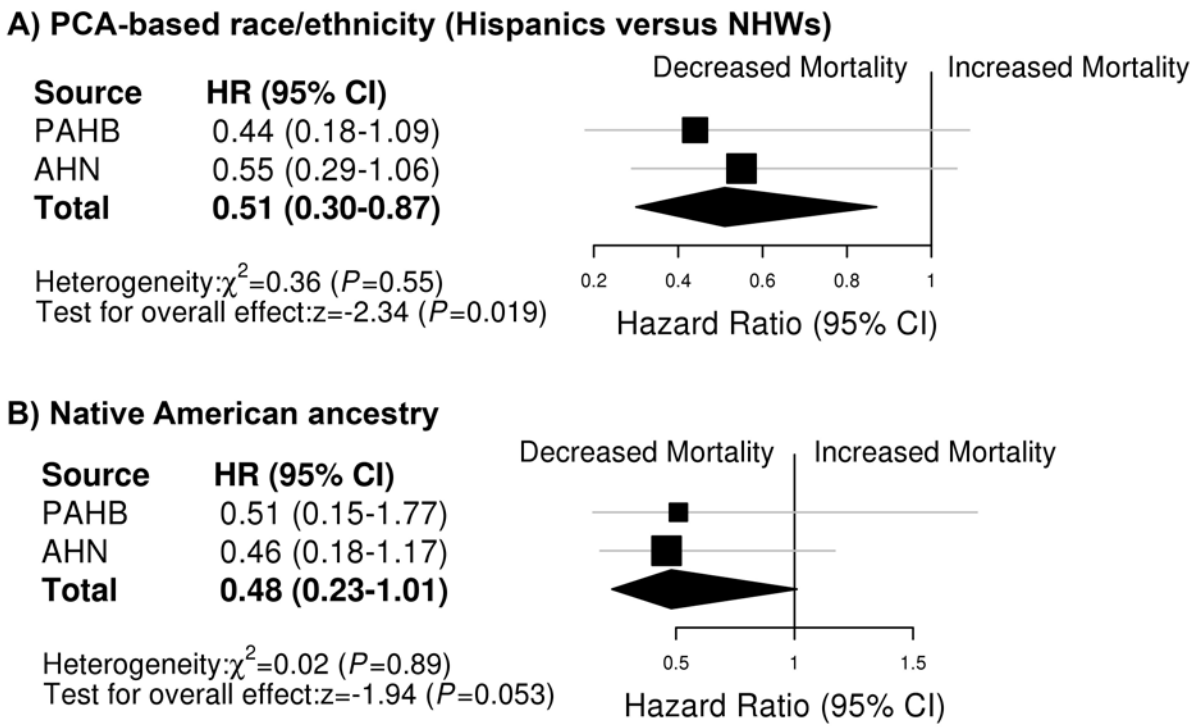


Figure E2. All-cause mortality hazard ratios for A) PCA-based race/ethnicity and B) Native American ancestry. Cox proportional hazards regression for all-cause mortality was performed in Group 1 PAH patients adjusted for age, sex, prostacyclin use, and log(PVR). For genetically-defined race/ethnicity, PCA was used to determine race/ethnicity as stated in the methods section. Meta-analysis of cox proportional hazards regression used a fixed effects model and inverse standard error was used as a weight for each study. Proportions of Native American ancestry were generated using ancestry informative markers input into STRUCTURE (k=3) using 1000 Genomes reference populations. 95% CI indicates 95% confidence interval; AHN, Allegheny Health Network; HR, hazard ratio; NHW, Non-Hispanic Whites; PAH, pulmonary arterial hypertension; PCA, principal components analysis; PVR, pulmonary vascular resistance.

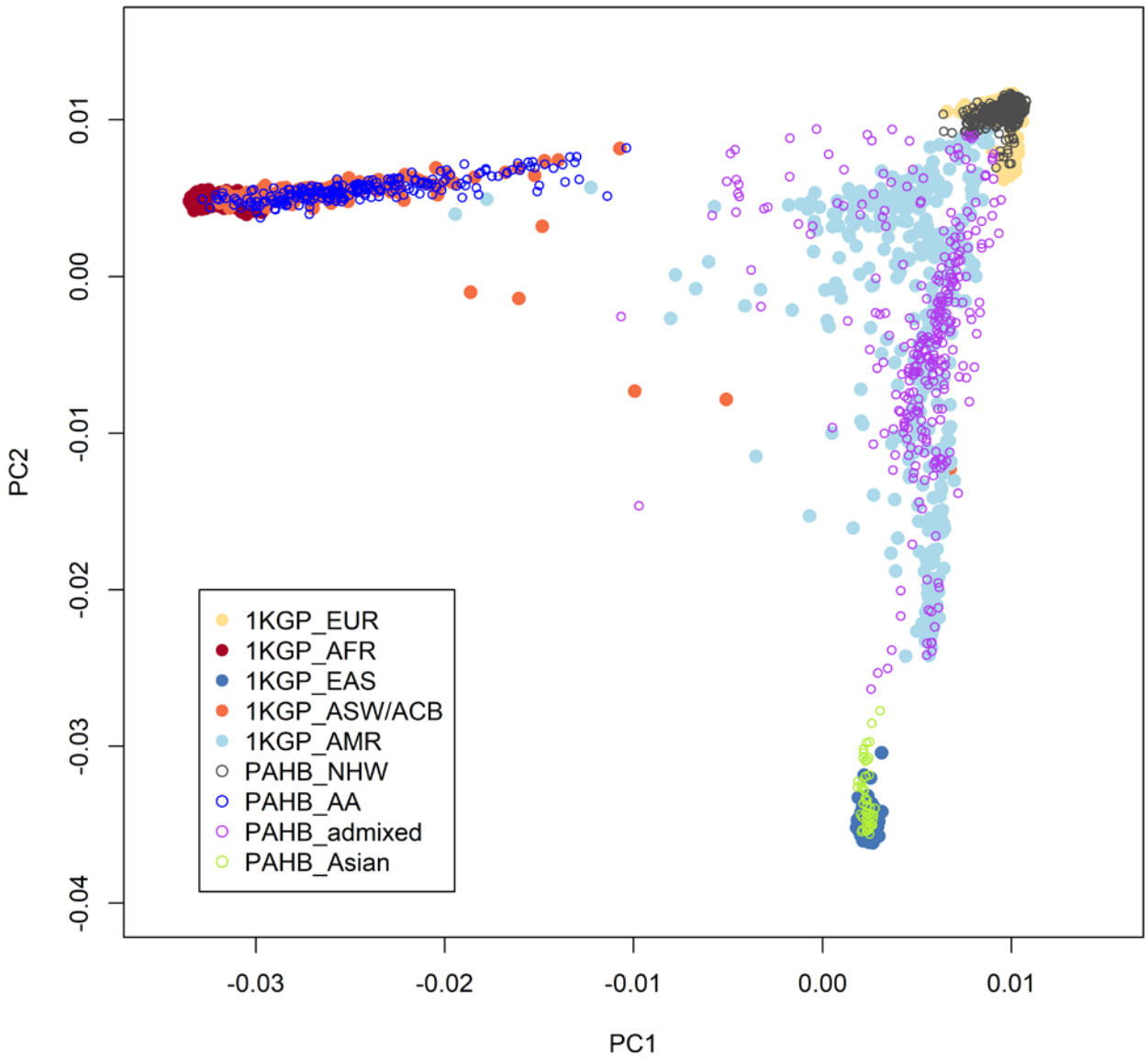


Figure E3. Principle components 1 and 2 for PAH Biobank patients by PCA-defined race/ethnicity alongside 1000 Genomes Project Reference samples. Samples were genotyped using the Illumina® HumanOmni5-QUAD BeadChip array. Principal components were estimated using the PC_AiR function in GENESIS version 3.8 package (github.com/UW-GAC/GENESIS). Individuals belonging to African, European, Asian, and admixed ancestry groups were manually identified using the first two PCs alongside 1000 Genome Project Reference samples.(13) Individuals who were considered admixed though PCA and also self-reported Hispanic ethnicity were considered Hispanic in PCA-based race/ethnicity analyses. 1KGP indicates 1000 Genomes Project Reference samples; AA, African American; ACB, African Caribbean in Barbados; AFR, African; AMR, American; ASW, African Ancestry in Southwest US; EAS, East Asian; EUR, European; NHW, Non-Hispanic White; PAH, pulmonary arterial hypertension; PAHB, PAH Biobank.

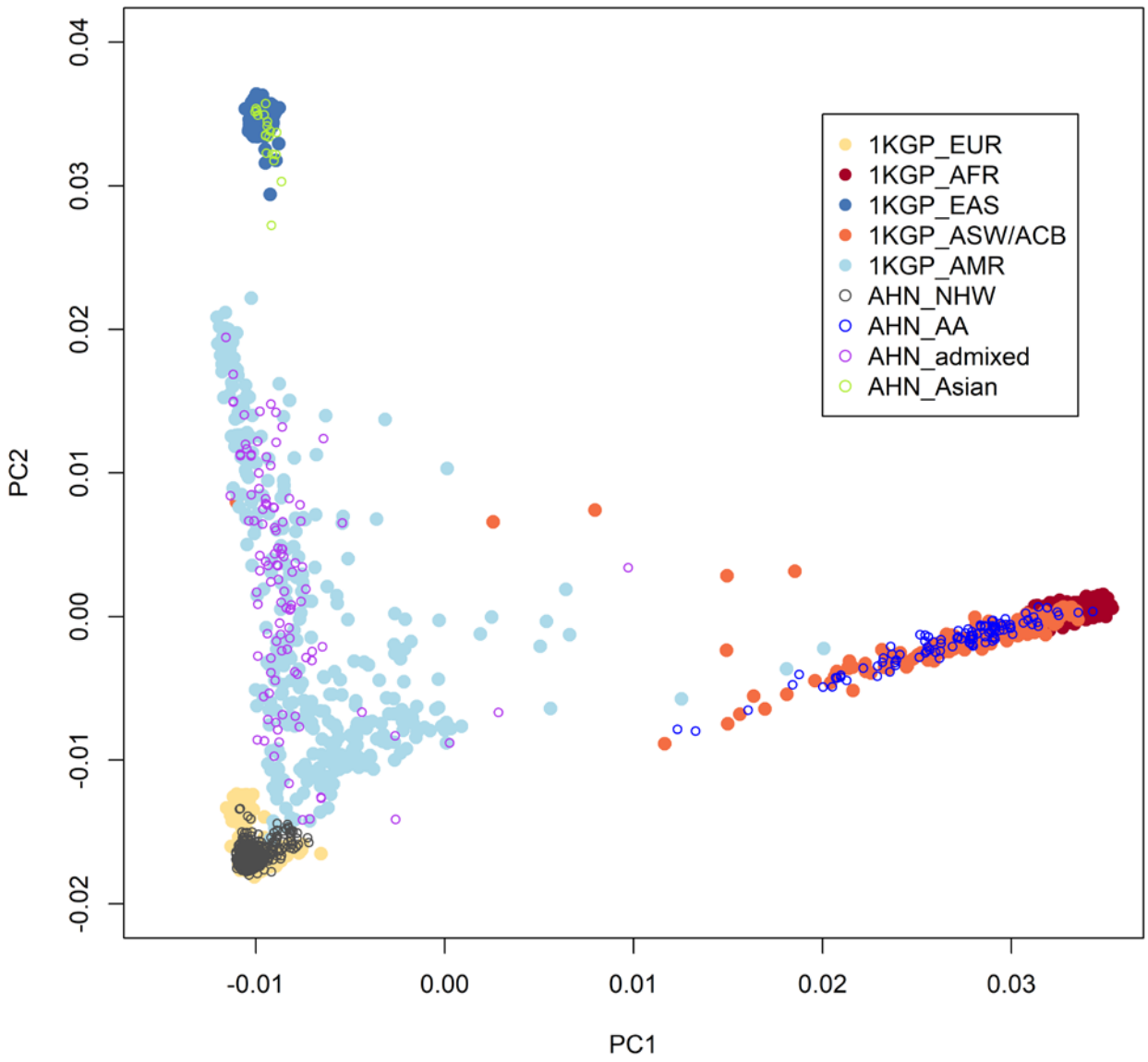
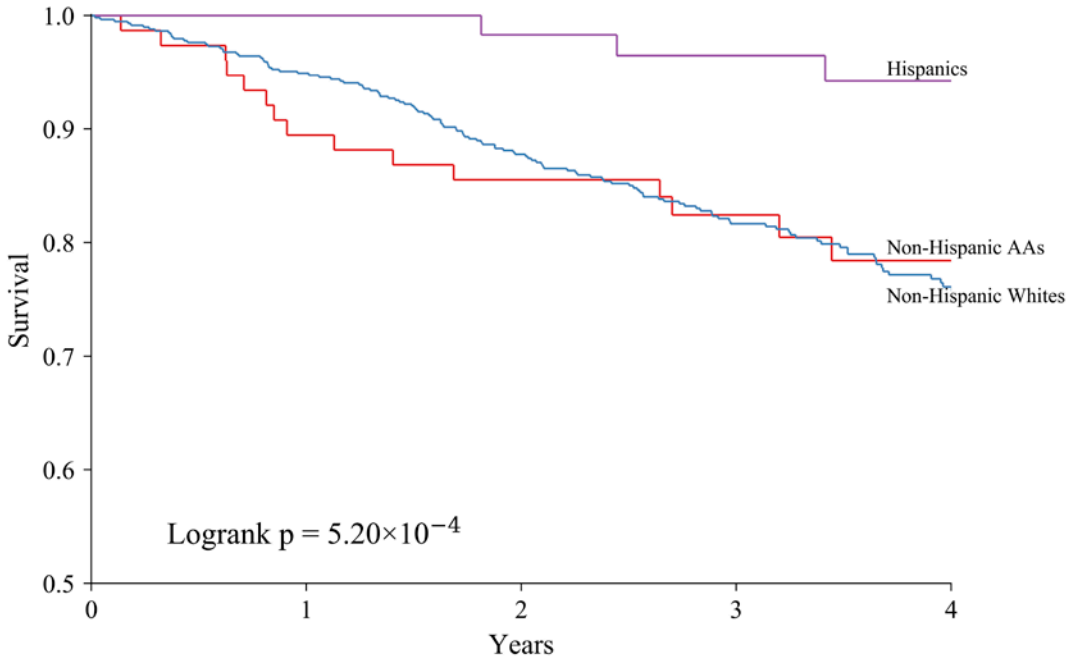
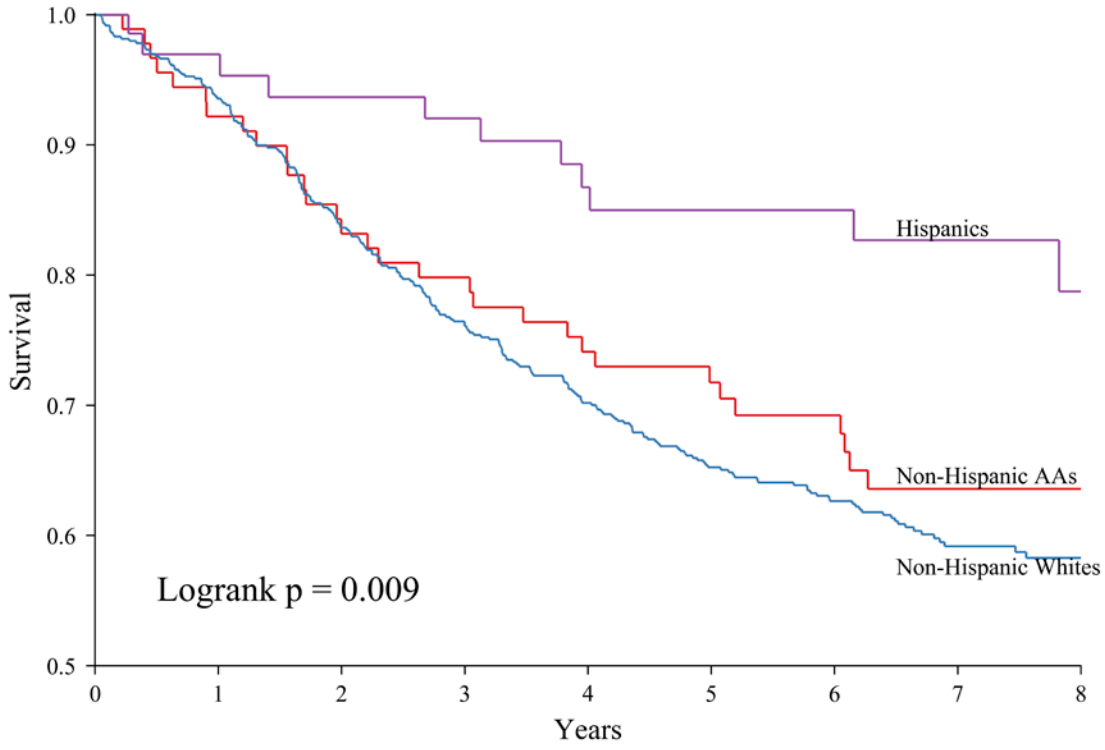


Figure E4. Principle components 1 and 2 for AHN Cohort patients by PCA-defined race/ethnicity alongside 1000 Genomes Project Reference samples. Genotyping was performed on the Illumina ® HumanOmniExpressExome BeadChip. Principal components were estimated using the PC_AiR function in GENESIS version 3.8 package (github.com/UW-GAC/GENESIS. Individuals belonging to African, European, Asian, and admixed ancestry groups were manually identified using the first two PCs alongside 1000 Genome Project Reference samples.(13) Individuals who were considered admixed though PCA and also self-reported Hispanic ethnicity were considered Hispanic in PCA-based race/ethnicity analyses. . 1KGP indicates 1000 Genomes Project Reference samples; AA, African American; ACB, African Caribbean in Barbados; AFR, African; AHN, Allegheny Health Network; AMR, American; ASW, African Ancestry in Southwest US; EAS, East Asian; EUR, European; NHW, Non-Hispanic White.



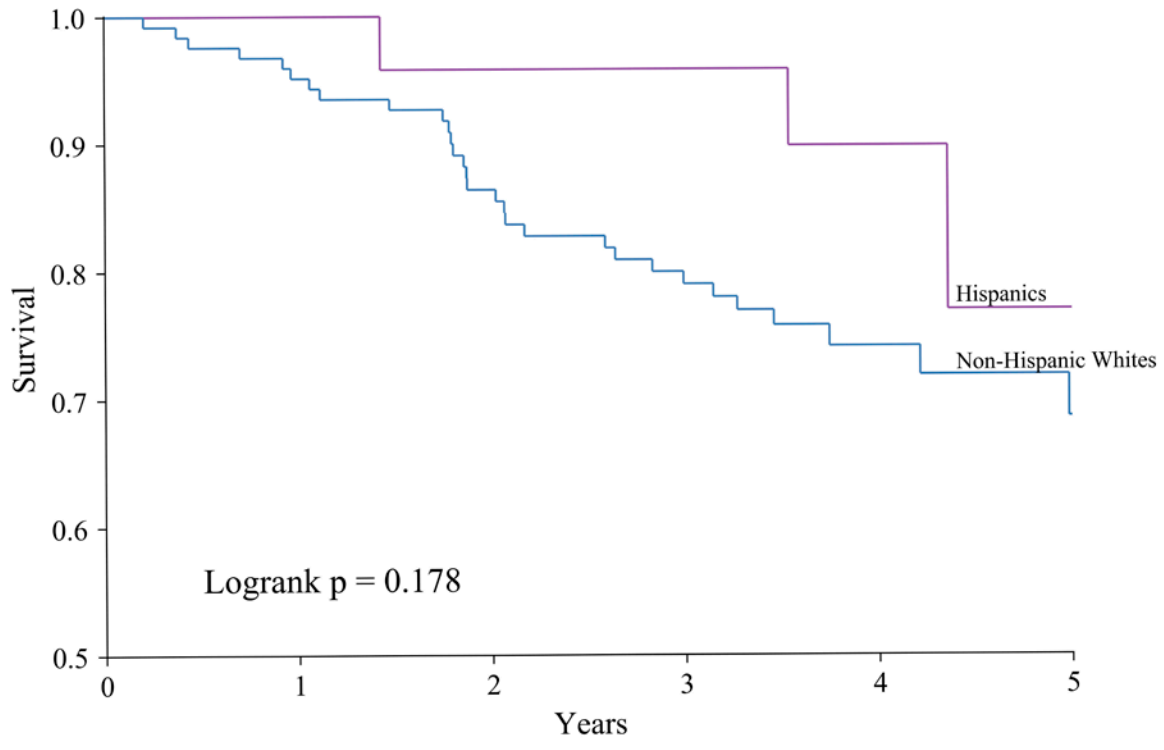
No. at risk					
Non-Hispanic AAs	76	68	60	45	27
Non-Hispanic Whites	590	560	498	356	213
Hispanics	60	60	57	45	37

Figure E5. Kaplan-Meier curves for all-cause mortality by PCA-based race/ethnicity in the PAH Biobank. Survival was calculated from enrolment in to date of death or censoring. In the PAH Biobank, January 1, 2018 was used as the date of censoring. Log-rank tests were performed for differential survival between race/ethnic groups. Kaplan-Meier curves were stopped when 10% of the original number at risk remained. Samples were genotyped using the Illumina® HumanOmni5-QUAD BeadChip array. Principal components were estimated using the PC_AiR function in GENESIS version 3.8 package (github.com/UW-GAC/GENESIS). Individuals belonging to African, European, Asian, and admixed ancestry groups were manually identified using the first two PCs alongside 1000 Genome Project Reference samples.(13) Individuals who were considered admixed though PCA and also self-reported Hispanic ethnicity were considered Hispanic in PCA-based race/ethnicity analyses. AA indicates African American; PAH, pulmonary arterial hypertension; PCA, principal components analysis.



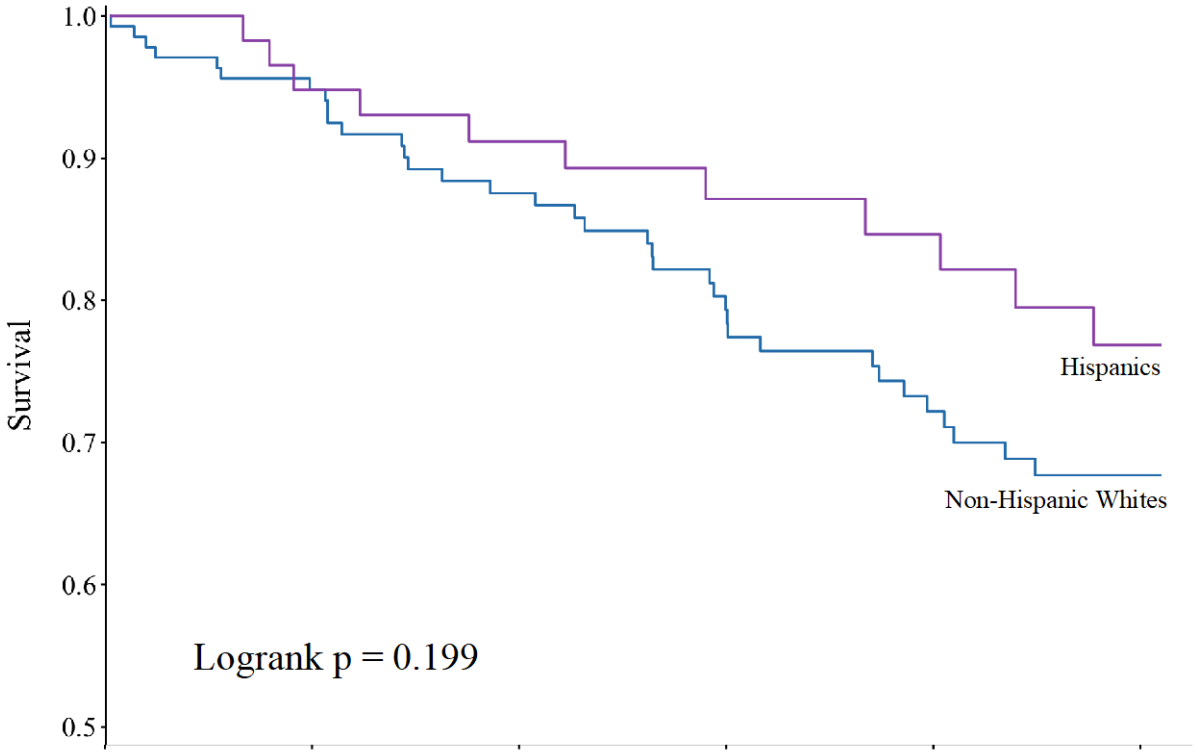
No. at risk									
Non-Hispanic AAs	90	82	74	70	65	59	52	37	20
Non-Hispanic Whites	597	548	489	440	403	355	305	179	104
Hispanics	70	59	57	53	49	45	37	26	20

Figure E6. Kaplan-Meier curves for all-cause mortality by PCA-based race/ethnicity in the AHN cohort. Survival was calculated from enrolment in to date of death or censoring. If a specific date for last contact was not recorded, the date 12/19/2014 was used as the censoring date. Log-rank tests were performed for differential survival between race/ethnic groups. Kaplan-Meier curves were stopped when 10% of the original number at risk remained. Genotyping was performed on the Illumina ® HumanOmniExpressExome BeadChip. Principal components were estimated using the PC_AiR function in GENESIS version 3.8 package (github.com/UW-GAC/GENESIS. Individuals belonging to African, European, Asian, and admixed ancestry groups were manually identified using the first two PCs alongside 1000 Genome Project Reference samples.(13) Individuals who were considered admixed though PCA and also self-reported Hispanic ethnicity were considered Hispanic in PCA-based race/ethnicity analyses. AA indicates African American; AHN, Allegheny Health Network; PAH, pulmonary arterial hypertension; PCA, principal components analysis.



No. at risk						
Non-Hispanic Whites	124	117	95	82	37	21
Hispanics	24	24	21	18	11	2

Figure E7. Kaplan-Meier curves for all-cause mortality by race/ethnicity in the UA cohort. Survival was calculated from enrolment up to date of death or censoring. The censor date was patient-specific, reflecting the date of acquisition of mortality status. Log-rank tests were performed for differential survival between race/ethnic groups. Kaplan-Meier curves were stopped when 10% of the original number at risk remained. PAH indicates pulmonary arterial hypertension; UA, University of Arizona.



No. at risk						
Non-Hispanic Whites	140	122	102	83	66	54
Hispanics	63	55	49	40	34	27

Figure E8. Kaplan-Meier curves for all-cause mortality by race/ethnicity in the Stanford cohort of newly diagnosed patients. Survival was calculated from the time of PAH diagnosis (index right heart catheterization) up to the date of death or censoring (loss to follow-up) during the five-year period following diagnosis. The number of patients remaining at risk each year is shown. The log-rank test was used to assess for differential survival between race/ethnic groups.

eReferences

1. Badesch DB, Raskob GE, Elliott CG, Krichman AM, Farber HW, Frost AE, Barst RJ, Benza RL, Liou TG, Turner M, Giles S, Feldkircher K, Miller DP, McGoon MD. Pulmonary arterial hypertension: baseline characteristics from the REVEAL Registry. *Chest* 2010; 137: 376-387.
2. Al-Naamani N, Paulus JK, Roberts KE, Pauciulo MW, Lutz K, Nichols WC, Kawut SM. Racial and ethnic differences in pulmonary arterial hypertension. *Pulm Circ* 2017; 7: 793-796.
3. Sandoval J, Torbicki A, Souza R, Ramirez A, Kurzyna M, Jardim C, Jerjes-Sanchez Diaz C, Teal SA, Hwang LJ, Pulido T. Safety and efficacy of sitaxsentan 50 and 100 mg in patients with pulmonary arterial hypertension. *Pulmonary pharmacology & therapeutics* 2012; 25: 33-39.
4. Benza RL, Barst RJ, Galie N, Frost A, Girgis RE, Highland KB, Strange C, Black CM, Badesch DB, Rubin L, Fleming TR, Naeije R. Sitaxsentan for the treatment of pulmonary arterial hypertension: a 1-year, prospective, open-label observation of outcome and survival. *Chest* 2008; 134: 775-782.
5. Benza RL, Gomberg-Maitland M, Demarco T, Frost AE, Torbicki A, Langleben D, Pulido T, Correa-Jaque P, Passineau MJ, Wiener HW, Tamari M, Hirota T, Kubo M, Tiwari HK. Endothelin-1 Pathway Polymorphisms and Outcomes in Pulmonary Arterial Hypertension. *Am J Respir Crit Care Med* 2015; 192: 1345-1354.
6. Healthcare Utilization Project. HCUP introduction. https://www.hcup-us.ahrq.gov/db/nation/nis/NIS_Introduction_2012.jsp.
7. Anand V, Roy SS, Archer SL, Weir EK, Garg SK, Duval S, Thenappan T. Trends and Outcomes of Pulmonary Arterial Hypertension-Related Hospitalizations in the United States: Analysis of the Nationwide Inpatient Sample Database From 2001 Through 2012. *JAMA Cardiol* 2016; 1: 1021-1029.
8. Mathai SC, Ryan Hemnes A, Manaker S, Anguiano RH, Dean BB, Saundankar V, Classi P, Nelsen AC, Gordon K, Ventetuolo CE. Identifying Patients with Pulmonary Arterial Hypertension (PAH) Using Administrative Claims Algorithms. *Ann Am Thorac Soc* 2019.
9. Quan H, Sundararajan V, Halfon P, Fong A, Burnand B, Luthi JC, Saunders LD, Beck CA, Feasby TE, Ghali WA. Coding algorithms for defining comorbidities in ICD-9-CM and ICD-10 administrative data. *Med Care* 2005; 43: 1130-1139.
10. . Available from: https://phewascatalog.org/.
11. Healthcare Utilization Project, US Department of Health and Human Services. HCUP method series. 2016. https://www.hcup-us.ahrq.gov/reports/methods/2006_05_NISTrendsReport_1988-2004.pdf.
12. Rhodes CJ, Batai K, Bleda M, Haimel M, Southgate L, Germain M, Pauciulo MW, Hadinnapola C, Aman J, Girerd B, Arora A, Knight J, Hanscombe KB, Karnes JH, Kaakinen M, Gall H, Ulrich A, Harbaum L, Cebola I, Ferrer J, Lutz K, Swietlik EM, Ahmad F, Amouyel P, Archer SL, Argula R, Austin ED, Badesch D, Bakshi S, Barnett C, Benza R, Bhatt N, Bogaard HJ, Burger CD, Chakinala M, Church C, Coghlan JG, Condliffe R, Corris PA, Danesino C, Debette S, Elliott CG, Elwing J, Eyries M, Fortin T, Franke A, Frantz RP, Frost A, Garcia JGN, Ghio S, Ghofrani HA, Gibbs JSR, Harley J, He H, Hill NS, Hirsch R, Houweling AC, Howard LS, Ivy D, Kiely DG, Klinger J, Kovacs G,

Lahm T, Laudes M, Machado RD, Ross RVM, Marsolo K, Martin LJ, Moledina S, Montani D, Nathan SD, Newnham M, Olschewski A, Olschewski H, Oudiz RJ, Ouwehand WH, Peacock AJ, Pepke-Zaba J, Rehman Z, Robbins I, Roden DM, Rosenzweig EB, Saydain G, Scelsi L, Schilz R, Seeger W, Shaffer CM, Simms RW, Simon M, Sitbon O, Suntharalingam J, Tang H, Tchourbanov AY, Thenappan T, Torres F, Toshner MR, Treacy CM, Noordegraaf AV, Waisfisz Q, Walsworth AK, Walter RE, Wharton J, White RJ, Wilt J, Wort SJ, Yung D, Lawrie A, Humbert M, Soubrier F, Tregouet DA, Prokopenko I, Kittles R, Graf S, Nichols WC, Trembath RC, Desai AA, Morrell NW, Wilkins MR, Consortium UNBRD, Consortium UPCS, Consortium UPB. Genetic determinants of risk in pulmonary arterial hypertension: international genome-wide association studies and meta-analysis. *Lancet Respir Med* 2018.

13. Abecasis GR, Altshuler D, Auton A, Brooks LD, Durbin RM, Gibbs RA, Hurles ME, McVean GA. A map of human genome variation from population-scale sequencing. *Nature* 2010; 467: 1061-1073.
14. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. *Genetics* 2000; 155: 945-959.
15. R Development Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2017.