**ARIC Manuscript Proposal #3692 (Amended)**

**PC Reviewed: 11/16/20** **Status: _____** **Priority: 2**
**SC Reviewed: _____** **Status: _____** **Priority: ____**

**1.a. Full Title**: Polygenic risk score and incident heart failure

**b. Abbreviated Title (Length 26 characters)**: PRS and HF

**2. Writing Group**:

Writing group members: Taryn Alkis, Christie Ballantyne, Eric Boerwinkle, Patricia Chang, Joe Coresh, Xi Luo, Amil, Shah, Katherine Wall and Bing Yu (Last name alphabetic order). Others are welcome.

I, the first author, confirm that all the coauthors have given their approval for this manuscript proposal. __TA___ **[please confirm with your initials electronically or in writing]**

**First author**: **Taryn Alkis**
Address: 1200 Pressler St, Suite E411, Houston, TX, 77030

Phone: Fax:
E-mail: Taryn.Alkis@uth.tmc.edu

**ARIC author** to be contacted if there are questions about the manuscript and the first author does not respond or cannot be located (this must be an ARIC investigator).
Name: **Bing Yu**
Address: 1200 Pressler St, Suite E407, Houston, TX, 77030

Phone: 713-500-9285 Fax:
E-mail: Bing.Yu@uth.tmc.edu

**3. Timeline**:

We will use existing genetic data, risk factors and incident heart failure follow-up information, and there is no other data collection work needed. When the proposal is approved, data analysis process will start. A manuscript is expected to be submitted within 6 months after the approval.

**4. Rationale**:

Between 2013 and 2016, 6.2 million Americans ages 20 and older had heart failure (HF).[1] By 2030, projections from the American Heart Association estimate HF will affect over 8 million in the US.[2] As the leading cause of morbidity and mortality worldwide, affecting over 30

million in the global population, heart failure is a complex disorder with known comorbidities and well-studied lifestyle risk factors.[3]

One 2018 Swedish adoption study characterized HF genetic heritability (h2) at 26%.[4] In the search for genetic variation that may account for disease susceptibility, autosomal single-nucleotide polymorphisms (SNPs) at several loci have been found important. An analysis using four prospective cohort studies identified two SNPs that reached genome-wide significance in incident HF cases within 60 kb of known genes, in European-ancestry near USP3 and CA12 and in African-ancestry participants near LRIG3.[5] Genome-wide association studies (GWAS) of HF in the UK Biobank dataset identified three loci associated with HF. One was an intergenic SNP nearest to PITX2, a susceptibility locus for atrial fibrillation, and two other SNPs were nearest to CDKN2B-AS1 and MAP3K7CL, coronary artery disease loci.6 Subsequent analysis of 26 studies from the Heart Failure Molecular Epidemiology for Therapeutic Targets (HERMES) Consortium found 12 independent genetic SNPs significantly associated with HF, and majority of the SNPs were associated with HF risk factors, including: atrial fibrillation (AF), coronary artery disease (CHD), low-density lipoprotein cholesterol, type 2 diabetes (T2D), body mass index (BMI), systolic and diastolic blood pressure. [3]

Common complex diseases are often polygenic, and a single SNP is not informative enough to assess the risk of the disease. Polygenic risk score (PRS), a genetic loading by combining a set of risk SNPs, is a powerful approach to identify people at high risk. PRS has been successful in predicting multiple phenotypes, such as CHD and stroke,[6,7] but such methods have yet to be applied to HF, and limited information about how PRS affecting HF development is currently available.

In this proposal, we seek to identify the associations between PRS and incident HF in the ARIC study. We will test whether a HF PRS derived from case-control GWAS is associated with incident HF. Additionally, we will expand the scope to test whether PRSs derived from clinical HF risk factors, including BMI, CHD, AF, T2D and etc., and from cardiac structure and function, are associated with incident HF.

## 5. Main Hypothesis/Study Questions:

1. To test whether HF PRS derived from large-scale HF case-control GWAS is associated with incident HF in ARIC European and African Americans;

2. To test whether PRSs of clinical HF risk factors, including BMI, CHD, AF, T2D and etc., are associated with incident HF in ARIC European and African Americans;

3. To test whether PRSs of cardiac structure and function are associated with incident HF in ARIC European and African Americans.

**6. Design and analysis (study design, inclusion/exclusion, outcome and other variables of interest with specific reference to the time of their collection, summary of data analysis, and any anticipated methodologic limitations or challenges if present).**

**Study design:**
This is a prospective study design using ARIC HRC/1000G imputed genetic data and incident HF information followed by 2018 in European and African Americans.

**Exclusion criteria:**
The individuals will be excluded from this study, whose HF follow-up information are missing, as well as other covariates. Participants with HF history at Visit 1 will also be excluded from the analysis.

**Variables:**
Outcome variables: incident HF followed by 2018 (primary)
Incident HF is defined as the first occurrence of a hospitalization with a HF diagnosis according to the International Classification of Diseases- 9th Revision (ICD-9) code 428 (428.0 to 428.9) in any position, or a death certificate with an ICD-9 code of 428 or with an ICD-10 code of I50 among any of the listed diagnoses or underlying causes of death.
If an association is found, we will test post-2005 HFpEF and HFrEF and echo measures at visit 5 secondarily.

Exposure variables: PRS for HF, BMI, AF, CHD, T2D and echo measures

1. HF PRS: top snps weighted sum methods, as well as other Bayesian prediction methods (i.e. LDpred[10], PRS-CS[11]), using HERMS[3] or UKBB[8] HF GWAS summary

PRS will be generated by race group based on their genotype dosages from the selected SNPs that met genome-wide significance in large HF GWAS.[3, 8]

2. Risk factor PRS: pre-derived genome-wide snps weights

PRS will be generated by race group using the per-allele weights derived from previous PRS studies.[6, 9]

3. Echo measure PRS: top snps weighted sum methods, as well as other Bayesian prediction methods (i.e. LDpred, PRS-CS), using UKBB[12] echo GWAS summary.

Covariates: age, sex, study center, BMI, smoking, SBP, blood pressure lowering medication, prevalent CHD, heart rate, prevalent T2D, prevalent AF, and first 10 principal components (PCs).

Exploratory variables: echo measures at visit 5, proteome measures at visits 3 and 5, and metabolome measures at visits 1 and 5.

**Statistical analysis:**
We will construct PRSs and perform a series of analysis by race groups. PRS will be standardized (mean = 0 and SD = 1) and categorized (low risk, median risk, high risk) prior to the analyses. We will run Cox proportional hazards regressions for several models to account for

potential confounding factors. Model 1 will be adjusted for age, sex, center and PCs. Model 2 will additionally adjust for heart rate, smoking, BMI, SBP and blood pressure lowering medication. Model 3 will further adjust for prevalent diabetes and CHD. We will examine the potential effect modification by sex using stratified analysis and interaction term, and investigate the incremental predictability of PRS beyond traditional risk factors using C-stat. We will further test the PRS predictability across lifetime by stratifying the population with a mean age at 55 years old (visit 1), 65 years old (visit 4) and 75 years old (visit 5). If an association is found, we will secondarily test PRSs 1) with post-2005 incident HFpEF vs HFrEF setting Jan 1 2005 as start of follow-up and using the AFU data for covariate adjustment and carrying forward BMI from visit 4; 2) with echo measures at visit 5; and 3) with proteome and metabolome measures at visits 3 and 5. Significance will be defined as $p < 0.01$ to correct for the five PRSs to be tested.

**7.a. Will the data be used for non-ARIC analysis or by a for-profit organization in this manuscript? ____ Yes __X__ No**

**b. If Yes, is the author aware that the current derived consent file ICTDER05 must be used to exclude persons with a value RES_OTH and/or RES_DNA = "ARIC only" and/or "Not for Profit" ? ____ Yes ____ No**
(The file ICTDER has been distributed to ARIC PIs, and contains the responses to consent updates related to stored sample use for research.)

**8.a. Will the DNA data be used in this manuscript? __X__ Yes ____ No**

**8.b. If yes, is the author aware that either DNA data distributed by the Coordinating Center must be used, or the current derived consent file ICTDER05 must be used to exclude those with value RES_DNA = "No use/storage DNA"? __X__ Yes ____ No**

**9. The lead author of this manuscript proposal has reviewed the list of existing ARIC Study manuscript proposals and has found no overlap between this proposal and previously approved manuscript proposals either published or still in active status.**
ARIC Investigators have access to the publications lists under the Study Members Area of the web site at: http://www.cscc.unc.edu/aric/mantrack/maintain/search/dtSearch.html

__X____ Yes _____ No

**10. What are the most related manuscript proposals in ARIC (authors are encouraged to contact lead authors of these proposals for comments on the new proposal or collaboration)?**

**11.a. Is this manuscript proposal associated with any ARIC ancillary studies or use any ancillary study data? __ X __ Yes ____ No**

**11.b. If yes, is the proposal**
**A. primarily the result of an ancillary study (list number* _____)**

**B. primarily based on ARIC data with ancillary data playing a minor role (usually control variables; list number(s)* _____ _____ _____)**

*ancillary studies are listed by number https://sites.cscc.unc.edu/aric/approved-ancillary-studies

**12a. Manuscript preparation is expected to be completed in one to three years. If a manuscript is not submitted for ARIC review at the end of the 3-years from the date of the approval, the manuscript proposal will expire.**

**12b. The NIH instituted a Public Access Policy in April, 2008** which ensures that the public has access to the published results of NIH funded research. It is **your responsibility to upload manuscripts to PubMed Central** whenever the journal does not and be in compliance with this policy. Four files about the public access policy from http://publicaccess.nih.gov/ are posted in http://www.cscc.unc.edu/aric/index.php, under Publications, Policies & Forms. http://publicaccess.nih.gov/submit_process_journals.htm shows you which journals automatically upload articles to PubMed central.

References:


1. Benjamin EJ, et al. Heart disease and stroke statistics—2019 update: a report from the American Heart Association. Circulation. 2019;139(10):e56–528.
2. Heidenreich PA, et al. Forecasting the impact of heart failure in the United States: a policy statement from the American Heart Association. Circ Heart Fail. 2013;6(3):606-619.
3. Shah S, et al. Genome-wide association and Mendelian randomisation analysis provide insights into the pathogenesis of heart failure. Nat Commun 2020, 11, 163.
4. Lindgren MP, et al. A Swedish Nationwide Adoption Study of the heritability of heart failure. JAMA Cardiol. 2018, 3, 703–710.
5. Smith NL, et al. Association of genome-wide variation with the risk of incident heart failure in adults of European and African ancestry: a prospective meta-analysis from the cohorts for heart and aging research in genomic epidemiology (CHARGE) consortium. Circ. Cardiovasc. Genet. 2010, 3, 256–266.
6. Khera AV, et al. Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. Nat Genet. 2018, 50(9):1219-1224.
7. Abraham G, et al. Genomic risk score offers predictive performance comparable to clinical risk factors for ischaemic stroke. Nat Commun. 2019; 10: 5819.
8. Aragam KG, et al. Phenotypic refinement of heart failure in a National Biobank facilitates genetic discovery. Circulation 2019, 139, 489–501.
9. Khera AV, et al. Polygenic Prediction of Weight and Obesity Trajectories from Birth to Adulthood. Cell. 2019,177(3):587-596.e9.
10. Vilhjálmsson, B. et al. Modeling linkage disequilibrium increases accuracy of polygenic risk scores. Am. J. Hum. Genet. 97, 576–592 (2015).

11. Ge T, et al. Polygenic prediction via Bayesian regression and continuous shrinkage priors. Nat Commun 10, 1776 (2019).
12. Pirruccello JP, et al. Analysis of cardiac magnetic resonance imaging in 36,000 individuals yields genetic insights into dilated cardiomyopathy. Nat Commun 11, 2254 (2020).