1.a. Full Title: Physical activity and Epigenetic Age Acceleration

b. Abbreviated Title (Length 26 characters): Exercise and Epigenetics

2. Writing Group:**
   Writing group members:
   Elizabeth Hibler, MPH, PhD, Northwestern University Feinberg School of Medicine, Chicago, IL
   Yinan Zheng, Northwestern University Feinberg School of Medicine, Chicago, IL, USA, vinanzheng2011@u.northwestern.edu
   Stella Aslibekyan, University of Alabama-Birmingham, Birmingham, AL, USA, saslibek@uab.edu
   SERGI BAIXERAS, University of Alabama-Birmingham, Birmingham, AL, US, Birmssayols@imim.es
   Qi Dai, Vanderbilt University School of Medicine, Nashville, TN, USA, qi.dai@vanderbilt.edu
   Nick Roetker, University of Minnesota, Minneapolis, MN, USA, roetk001@umn.edu
   Jim Pankow, University of Minnesota, Minneapolis, MN, USA, panko001@umn.edu
   Jan Bressler, University of Texas Health Science Center, Houston, TX, USA
   Jan.Bressler@uth.tmc.edu
   Andrea Baccarelli, Columbia University, New York, NY, USA, ab4303@columbia.edu
   Kiarrri Kershaw, Northwestern University Feinberg School of Medicine, Chicago, IL, USA, kershaw@northwestern.edu
   Lifang Hou, Northwestern University Feinberg School of Medicine, Chicago, IL, USA, l-hou@northwestern.edu

**Please note, as described below, this study also incorporates data from the Women’s Health Initiative (WHI), Atherosclerosis Risk in Communities (ARIC), Normative Aging Study (NAS), and a growing list of cohorts. Additional co-authors from the studies are TBD and will be added as co-authors once the manuscript is ready for review.

I, the first author, confirm that all the coauthors have given their approval for this manuscript proposal. _EH____ [please confirm with your initials electronically or in writing]
3. **Timeline:** We are in the process of conducting the data analysis in individual cohorts. Then, we expect the meta-analysis to be completed by May 1, 2017 and the manuscript to be out to co-authors by summer 2017 at the latest.

4. **Rationale:**

   Studies demonstrate that low levels of physical activity are associated with increased mortality and risk of common diseases \(^1,^2\). However, the biological mechanisms driving population-level associations between activity and disease risk are unclear. Changes in patterns of DNA methylation are suggested as a promising mechanism. We proposed another ARIC study to examine whether physical activity is associated with patterns of DNA methylation at cytosine-phosphate-guanine (CpG) loci on an epigenome-wide scale. In contrast, the current study will evaluate associations between physical activity and DNA methylation at a specific set of CpG loci in genes related to chronological age \(^3,^4\).

   Studies recently identified sets of CpG loci that predict chronological age and deviations from the expected are described as epigenetic age acceleration \(^3,^4\). Hannum et al. and Horvath identified sets of 71 and 353 CpG loci associated with chronological age in models trained using only blood samples\(^3\) and using multiple tissues and cell types,\(^4\) respectively. When the epigenetic age of DNA is different from expected (based on chronological age), these differences can potentially influence biological processes affecting health and normal aging. Recent studies demonstrate that epigenetic age is associated with overall mortality, as well as chronic disease risk \(^4-^7\). Two recent studies have identified lifestyle factors, such as obesity and physical cognitive functioning, may influence epigenetic age \(^5,^7\). However, no studies to-date evaluated the relationship between physical activity and epigenetic age. Furthermore, we found very promising preliminary results between high levels of physical activity and epigenetic age among participants of the NAS cohort. However, due to the limited sample size, we propose to evaluate this relationship across additional cohorts, as described below.
In summary, for this paper, we will evaluate the relationship between physical activity and epigenetic age acceleration. As described below, the methods for harmonizing physical activity data across studies will be the biggest challenge. However, we have developed a plan with alternative approaches that will be employed as needed. Furthermore, no studies have evaluated whether the relationship between physical activity and epigenetic age acceleration differs by sex or race, though there is growing evidence to suggest that patterns of DNA methylation may differ by these factors. This will be the first study to evaluate the relationship between physical activity and epigenetic age across multiple, large-scale cohort studies.

### 5. Main Hypothesis/Study Questions:

1. Determine if physical activity is associated with the DNA methylation clock and epigenetic age acceleration. We hypothesize that low physical activity will be associated with greater epigenetic age acceleration.

### 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).

**Overview.** The primary aims of the CHARGE consortium is to facilitate meta-analyses of epigenome-wide studies among an international group of large population-based cohort studies. The proposed study will be a large-scale study and novel in that we will evaluate associations between healthy behaviors and a specific set of CpG loci associated with epigenetic aging. We will examine if low physical activity is associated with accelerated epigenetic age. To complete this aims, we will develop detailed data analysis plans and R code for each study, with study specific instructions as necessary. After each cohort reports the results of their multivariate regression analysis to us, we will run a meta-analysis across studies to estimate summary measures of association. This unique study will potentially identify biomarkers that vary with self-reported physical activity level, which may have future applications towards studies of chronic disease and cancer risk.

**Study Population:** Within the ARIC study, there are roughly n = 2,771 participants with available DNA methylation (n=2,393 from visit 2 and n=378 from visit 3) and activity data. In addition, to-date, other cohorts included in this study include the Women’s Health Initiative (WHI), Coronary Artery Risk Development in Young Adults (CARDIA), Normative Aging Study (NAS), Genetics of Lipid Lowering Drugs and Diet Network (GOLDN), and Registre Gironi del COR, the Girona Heart Registry (REGICOR).

**Exclusions:** Participants affected by leukemia and on chemotherapy will be excluded because it is possible that these factors may alter DNA methylation in whole blood cells.

**Primary Outcomes.** Epigenetic age acceleration, defined as the residuals resulting from regressing epigenetic age on chronological age. Various versions of epigenetic age acceleration will be used,
Main Exposure. Physical activity will be evaluated as a dichotomous variable (Yes/No) for meeting the American Cancer Society guidelines for 150 minutes of moderate-vigorous physical activity per week. Further, metabolic-equivalent (MET) minutes per week will be evaluated as a continuous variable.

Co-variables. Additional covariates include: body mass intake, alcohol intake, smoking status, and history of cancer. The technical variables of plate and chip number will adjust for any variation that may have been introduced in the lab. We will also adjust for Houseman estimates of cell type proportions (neutrophils; eosinophils; monocytes; lymphocytes; basophils). Covariates were selected based upon prior evidence in the literature and/or a biological basis for confounding the relationship between activity and DNA methylation.

Outcome Measurement:
Epigenetic Age Acceleration: Methylation status was measured from DNA extracted from whole blood white cells using the Illumina HM450 chip. Degree of methylation was determined using Illumina GenomeStudio 2011.1, Methylation module 1.9.0 software. The methylation score for each CpG was represented as a beta (β) value calculated by dividing the fluorescence intensity of the methylated allele by the sum of the intensities of the methylated allele and unmethylated allele. Background subtraction was conducted with the GenomeStudio software using built-in negative control bead types on the array. An average normalization was applied to minimize scanner-to-scanner variation. We will use the online Age Calculator software to perform additional normalization and imputation for missing beta values, as well as calculation of the Horvath and Hannum versions of epigenetic age and epigenetic age acceleration. AAHAAAdjCellCounts is the main age acceleration outcome variable of interest and uses the Houseman method to adjust for blood leukocyte cell type distribution and any confounding that may occur related to variation in proportions of cell types found in whole blood.

STATISTICAL CONSIDERATIONS

Data analysis plan: Following data harmonization, participating cohorts will run the provided R script using their study data. Multivariate linear regression will be the primary method utilized, with epigenetic age acceleration as the outcome and physical activity as the exposure of interest. In the first model, we will adjust for age, sex, race, plate/chip number, and estimated cell type distribution. A second model will further adjust for body mass index, education, alcohol use, smoking, and history of cancer.

Meta-Analyses
We will then conduct a meta-analysis of cohort-specific effects, accounting for heterogeneity between studies, to determine the overall associations between physical activity and epigenetic age acceleration. All analysis will be conducted using R software.

7.a. Will the data be used for non-CVD analysis in this manuscript? _____ Yes ___X___ No
b. If Yes, is the author aware that the file ICTDER03 must be used to exclude persons with a value RES_OTH = “CVD Research” for non-DNA analysis, and for DNA analysis RES_DNA = “CVD Research” would be used? ____ Yes    ____ No
(This 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 file ICTDER03 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/search.php

____ 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)?

Note: the present manuscript proposal is using specific sets of methylation markers that predict age as an exposure as related to physical activity, making its focus different than the genome-wide methylation proposal by Bressler et al. (which lists physical activity as a risk factor of interest)

MS#1928 Bressler J et al. Genome-wide methylation analysis of cardiovascular disease (CVD) and its risk factors

MS#2345 Roetker N et al. A prospective study of the association of DNA methylation age with lung function and type 2 diabetes in the Atherosclerosis Risk in Communities Study

MS#2827 Roetker N et al. A prospective study of DNA methylation age acceleration and incidence of coronary heart disease, heart failure, and peripheral arterial disease in the Atherosclerosis Risk in Communities (ARIC) Study

11.a. Is this manuscript proposal associated with any ARIC ancillary studies or use any ancillary study data? ____ Yes    __X__ 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)* __________ __________ __________)
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.cscck.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.

13. Per Data Use Agreement Addendum, approved manuscripts using CMS data shall be submitted by the Coordinating Center to CMS for informational purposes prior to publication. Approved manuscripts should be sent to Pingping Wu at CC, at pingping_wu@unc.edu. I will be using CMS data in my manuscript ____ Yes __X__ No.