ARIC Manuscript Proposal # 3016

PC Reviewed: 7/11/17  Status: _____  Priority: 2
SC Reviewed: _________  Status: _____  Priority: ____

1.a. **Full Title**: Duration, Severity and Distribution of Obesity and Subsequent CVD
   b. **Abbreviated Title (Length 26 characters)**: Obesity heterogeneity and CVD

2. **Writing Group**:
   Writing group members:

   Ellen Demerath (University of Minnesota), Bing Yu (University of Texas Health Sciences Center at Houston), Penny Gordon-Larsen (University of North Carolina), Danyu Lin (University of North Carolina), Annie Green Howard (University of North Carolina)

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

   **First author**: Christy Avery
   **Address**: 137 E. Franklin St., Suite 306
                Chapel Hill, NC 27514

   Phone: 919/966-2148  Fax: 919/966-9800
   E-mail: kari_north@unc.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**: Kari North
   **Address**: 137 E. Franklin St., Suite 306
                Chapel Hill, NC 27514

   Phone: 919/966-2148  Fax: 919/966-9800
   E-mail: kari_north@unc.edu

3. **Timeline**: Analyses will begin once study-specific permissions are obtained.

4. **Rationale**: Over the past 30 years, the prevalence of obesity has increased markedly in the U.S., doubling among adults and disproportionately affecting racial/ethnic minorities.\textsuperscript{1,2} The prevalence of overweight has also increased, with women aged >60 being at greater risk compared to other age groups.\textsuperscript{1} Increases in obesity are paralleled by rises in obesity-related
morbidity, mortality, and health care expenditures, most notably from cardiovascular diseases (CVD), reversing health and life expectancy gains achieved over the past century. Studies evaluating associations between overall obesity (assessed by body mass index [BMI]) and CVD and its risk factors (e.g. dyslipidemia, inflammation, elevated blood pressure, and metabolic impairment) have shown that not all obese individuals are at equal risk for CVD, underscoring the heterogeneous nature of obesity. One plausible, but largely unexplored source of heterogeneity is the impact of duration, which may better characterize obesity-associated CVD risk while minimizing reverse causation (i.e. weight loss from underlying disease). Furthermore, the health and economic burden at a BMI≥40 kg/m² is much higher than that at BMI≥30 kg/m², yet most studies categorize all individuals with BMI≥30 kg/m² identically, thus ignoring severity. Differences in fat distribution (e.g. abdominal vs overall obesity) also may underlie obesity heterogeneity, given the associations between abdominal obesity (e.g. assessed by waist circumference [WC]) with CVD independent of BMI. Finally, no studies, to the best of our knowledge, have jointly examined the influence of overall and abdominal obesity, severity, and duration on CVD.

We therefore propose to leverage novel statistical methods, longitudinal measures of obesity distribution, duration, and severity, and validated CVD events to interrogate heterogeneity in obesity distribution, duration and severity, operationalized as obesity subtypes, in the Atherosclerosis Risk in Communities (ARIC) study. Racial/ethnic- and sex-specific associations will be examined as secondary exploratory extensions to each aim.

5. Main Hypothesis/Study Questions:
Examine associations of obesity subtypes characterized by body fat distribution, duration, and severity with incident CVD - coronary heart disease (CHD), stroke, heart failure (HF) and overall mortality in ARIC. We will extend a joint latent class model to simultaneously relate obesity subtypes estimated from longitudinal obesity distribution (abdominal and overall), duration (years), and severity (degree of obesity) measures to CVD risk while accommodating longitudinal confounders, missing data, and bias induced by random measurement error, intermittent exposure measurements, and event occurrence.

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).

Obesity measurements come from anthropometry and capture overall body mass and body fat distribution (trunk, limb); a majority of participants have at least two assessments (Table 1). These measures will be used to construct obesity subtypes. Briefly, WC was assessed at the level of the umbilicus and hip circumference at the level of maximal protrusion of the gluteal muscles. Subscapular and triceps skinfold measures were taken twice using a Lange caliper on standardized right-side locations. Calf circumference was assessed at the maximum circumference over the calf muscle. Standard height and weight were collected, with weight measured using a scale zeroed daily and calibrated quarterly. Inter-technician reliability coefficients for tricep skinfold, subscapular skinfold, waist, hip, and WHR exceeded 0.90. The only self-reported metric is weight reported at age 25. Recalled weight, even from the distant past, is highly correlated with measured weight. Moreover, we used the self-reported measures in obesity GWAS and EWAS. Of note, in ARIC we previously flagged those who lost weight due to pathology and those who did not (preliminary analysis). For example, our previous work considered the following...
criteria when assessing morbidity/chronic conditions associated with weight gain or loss: cancer, heart failure and kidney failure, as well other conditions following the literature.

**Event definitions.** Follow-up for all outcomes (CHD, HF, and stroke; baseline visit through 12/31/13 or the latest date for which outcome data are available) was accomplished through a combination of active surveillance of local hospital discharge lists, annual participant interviews querying hospitalizations, examination of vital records, and interviews with decedent’s next of kin. Incident HF was defined by hospitalization or death certificate codes listing 428 or I-50 in any position, shown previously to have high levels of accuracy.\(^{31}\) Incident CHD was defined as a validated definite or probable hospitalized myocardial infarction (MI), a definite CHD death, or an unrecognized MI identified by electrocardiograph. The criteria for definite or probable hospitalized MI were based on combinations of chest pain symptoms, electrocardiogram changes, and cardiac enzyme levels.\(^{32}\) Physician-adjudicated stroke was identified based on the presence of ICD-9 codes 430 to 438 and neurological signs and symptoms. Differences in associations by stroke subtype will be evaluated in sensitivity analysis to the degree possible given modest numbers of non-ischemic stroke. As of 12/31/2014, ARIC investigators identified 2,431, 2,867, and 1,332 cases of incident CHD, HF, and stroke, respectively over a mean of 20 years follow-up.

**Statistical methods.** For all continuous outcomes, normality will be assessed and appropriate transformations will be applied, as required. Confounders will be assessed through literature review and construction of directed acyclic graphs (DAGs).\(^{33}\) Although we present our approach for the entire population, we also will examine evidence for effect modification by sex and race/ethnicity as secondary exploratory extensions to each aim. We also will develop software in flexible and freely accessible programming packages (e.g. R, via the CRAN repository, https://cran.r-project.org/web/packages/) to ensure that our software is widely and freely available.

**C.5.a. Extension of statistical framework to evaluate obesity subtypes.** We will leverage novel statistical methods, longitudinal obesity distribution, duration, and severity measures, as well as validated CVD events, to interrogate heterogeneity in both obesity (distribution, duration and severity) and heterogeneity in risk of CVD events and mortality, operationalized as subtypes of obesity relative to CVD/mortality risk. Our statistical development and extension is as follows: suppose that there are \(n\) study subjects who belong to \(G\) unobserved obesity latent classes (i.e. obesity and CVD/mortality risk subtypes for example high overall obesity, low abdominal obesity and moderate risk of a CV event) defined by both obesity distribution, duration, and severity (see Table 1 for obesity metrics used to define subtypes) and as well as risk of CVD event or mortality.

\[\text{Let } C_i \text{ denote the latent class membership for the } i\text{th subject, with } C_i = g \text{ if the } i\text{th subject belongs to latent class } g \quad (g = 1, \ldots, G). \text{ We relate } C_i \text{ to a set of time-independent covariates } W_i \text{ (e.g., sex, center, etc.) through a multinomial logistic regression model:}\]

\[
\Pr(C_i = g \mid W_i) = \frac{e^{\alpha_{0g} + \alpha_{1g}^TW_i}}{\sum_{l=1}^{G} e^{\alpha_{0l} + \alpha_{1l}^TW_i}},
\]

<table>
<thead>
<tr>
<th>Metric</th>
<th>Mean number of longitudinal measures(^{3})</th>
<th>N. participants with (\geq 1) measurement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Obesity</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BMI (kg/m(^2))</td>
<td>3.9</td>
<td>15,732</td>
</tr>
<tr>
<td>Calf girth, cm</td>
<td>1.0</td>
<td>15,739</td>
</tr>
<tr>
<td>Hip, cm</td>
<td>3.9</td>
<td>15,743</td>
</tr>
<tr>
<td>Subscapular skinfold</td>
<td>1.9</td>
<td>15,726</td>
</tr>
<tr>
<td>Tricep skinfold</td>
<td>1.9</td>
<td>15,733</td>
</tr>
<tr>
<td>Waist</td>
<td>3.9</td>
<td>15,742</td>
</tr>
<tr>
<td>Waist hip ratio</td>
<td>3.8</td>
<td>15,735</td>
</tr>
<tr>
<td>Weight age 25(^{5})</td>
<td>1</td>
<td>15,546</td>
</tr>
</tbody>
</table>

\(^{1}\)Per participant. \(^{2}\)Self-reported at baseline. \(^{3}\)Omics measures available at select time points
where $\alpha_{0g}$ is the intercept for class $g$, and $\alpha_{1g}$ is the vector of class-specific regression parameters associated with $W_i$. For identifiability, we set $\alpha_{0G} = 0$ and $\alpha_{1G} = 0$.

Each latent class is characterized by class-specific trajectories of multivariate longitudinal outcomes, specifically longitudinal obesity distribution, duration, and severity measures, and a class-specific risk of a CVD event or mortality. The longitudinal outcomes and the event time are assumed to be conditionally independent given the latent classes.

Suppose that there are $K$ measures of obesity and the $k$th measure is collected on the $i$th subject at $n_{ik}$ time points. For $i = 1, \ldots, n$, $k = 1, \ldots, K$, and $j = 1, \ldots, n_{ik}$, let $Y_{ijk}$ denote the $k$th longitudinal obesity measurement of the $i$th subject at time $t_{ijk}$, and let $X_{ijk}$ denote the corresponding covariates. We relate $Y_{ijk}$ to $X_{ijk}$ through the multivariate linear mixed model:

$$Y_{ijk} | c_i = g = \beta_{kg}^T X_{ijk} + b_i^T \tilde{X}_{ijk} + \epsilon_{ijk},$$

where $\beta_{kg}$ is a vector of class-specific regression parameters for the $k$th outcome, $\tilde{X}_{ijk}$ is a subset of $X_{ijk}$, $b_i$ is a vector of random effects assumed to be $d$-variate normal with mean 0 and covariance matrix $\Sigma_i(\xi_g)$ indexed by parameters $\xi_g$, and $\epsilon_{ijk}$ is zero-mean normal with variance $\sigma_{kg}^2$. By using a multivariate linear mixed framework we can capture both the complex longitudinal nature of the obesity measurements, taking into account the correlation over time, but we are also simultaneously modeling multiple obesity measures thereby accounting for the correlation across obesity measures. Let $T_i$ denote the time to incident CVD event or death for the $i$th subject. We relate $T_i$ to a set of potentially time-dependent covariates $Z_i$ through the proportional hazards model:

$$\lambda(t | Z_i, C_i = g) = \lambda_{0g}(t) e^{\gamma_g^T Z_i(t)},$$

where $\lambda_{0g}(\cdot)$ is an arbitrary class-specific baseline hazard function, and $\gamma_g$ is a set of class-specific regression parameters. In the presence of censoring, we observe $\tilde{T}_i = \min(T_i, U_i)$ and $\Delta_i = I(T_i \leq U_i)$, where $U_i$ is the censoring time on the $i$th subject, and $I(\cdot)$ is the indicator function. The joint likelihood function takes the form

$$\prod_{i=1}^n \prod_{g=1}^G \sum_{W_i} e^{\alpha_{0g} + \alpha_{1g}^T W_i} \int_{b_i}^{\infty} \prod_{k=1}^K \prod_{j=1}^{n_{ik}} f_{kg}(Y_{ijk} | X_{ijk}; b_i) f(b_i; \Sigma_i(\xi_g)) db_i \left\{ \lambda_{0g}(t) e^{\gamma_g^T Z_i(t)} \right\}^{\Delta_i} \exp \left\{ - \int_0^{\tilde{T}_i} e^{\gamma_g^T Z_i(t)} d\Lambda_{0g}(t) \right\},$$

where $f_{kg}(y | X_{ijk}; b_i)$ is the normal density function with mean $\beta_{kg}^T X_{ijk} + b_i^T \tilde{X}_{ijk}$ and variance $\sigma_{kg}^2$, $f(b_i; \Sigma_i(\xi_g))$ is the $d$-variate normal density function with mean 0 and covariance matrix $\Sigma_i(\xi_g)$, and $\Lambda_{0g}(t) = \int_0^t \lambda_{0g}(u) du$. We adopt the nonparametric maximum likelihood estimation approach, under which $\Lambda_{0g}(\cdot)$ is treated as a step function that jumps at the observed event times and $\lambda_{0g}(t)$ is replaced by the jump size of $\Lambda_{0g}(\cdot)$ at $t$. We will derive an EM algorithm to carry out the maximization by treating $C_i$ and $b_i$ as missing values. The resulting estimators will be shown to be consistent, asymptotically normal with a covariance matrix that can be estimated by the profile likelihood method. Finally, the number of latent classes $G$ is usually
unknown. We maximize the likelihood for a fixed number of latent classes, and the optimal number of latent classes is determined by the Bayesian Information Criterion (BIC).

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? ____ Yes  ___X___ 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”? ____ 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)?

    MS 2738 - Metabolically healthy obesity, cumulative exposure to obesity, and progression to incident metabolic syndrome: The Atherosclerosis Risk in Communities Study
    MS 1456 - Measures of obesity in predicting different CVD outcomes by race and sex in the ARIC study

The first proposal is distinct from ours in that it evaluated metabolic syndrome and the second was approved nearly 10 years ago.

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)* __________ __________ __________)

*ancillary studies are listed by number at http://www.cscc.unc.edu/aric/forms/
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.

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.
Literature Cited