1.a. Full Title:
Empirical validation of the metabolic syndrome components and cutpoints through the prediction of CHD and diabetes using recursive partitioning methods

b. Abbreviated Title (Length 26 characters):
Validation of Metabolic Syndrome Components

2. Writing Group (list individual with lead responsibility first):
Lead Annie McNeill, PhD

Writing group members: Sherita Hill Golden, Mercedes Carnethon, Latha Palaniappan, Cynthia Girman, Diane Catellier, Bruce Duncan and other interested ARIC Investigators

3. Timeline: Analyses will begin immediately upon approval by the Publications and Steering committees.

4. Rationale:

Recently, the National Cholesterol Education Program (NCEP), Third Adult Treatment Panel Report (ATP III)\(^1\) proposed a definition of the metabolic syndrome to facilitate the use of uniform criteria in epidemiologic studies and as a tool to aid in the clinical evaluation of individuals at risk for coronary heart disease.

Recently published studies using this definition have shown that African Amerian and Caucasian subjects with the metabolic syndrome are at increased risk of developing diabetes and cardiovascular disease\(^2,3\) although no association between this condition and incident CHD was observed in the Strong Heart Study of Native Americans.\(^6\) While most prospective studies have shown an increased incidence of CHD and diabetes with metabolic syndrome, few have investigated which of these highly correlated components of the syndrome confer the greatest risk.

One limitation of these prospective studies is that they have attempted to assess the association between the individual components of the metabolic syndrome and CHD or diabetes by modeling these associations using logistic or proportional hazards regression among a set of variables that are known to be highly correlated.\(^7\) Such collinearity may represent a violation of the assumptions of these models and result in biased estimates of the true relationship between each component and the outcome of interest. Further, previous studies that relied on the pre-specified dichotomous cutpoints to define each of the metabolic syndrome components did not explore whether there are alternative thresholds that, if applied, would result in greater discrimination between individuals with and without the clinical outcome of interest. Such information would have prognostic value for medical practitioners.

In contrast, recursive partitioning methods such as signal detection analysis (SDA) and Classification and Regression Tree (CART) do not impose such limitations. Both approaches are quantitative, exploratory
analytic tools to identify mutually exclusive subgroups of high and low risk subjects for a given outcome and employ a decision tree-based structure to classify individuals with and without the outcome of interest using a mathematical algorithm that maximizes sensitivity and specificity at each decision branch or node. (note to the writing group: Latha says that under SDA, variables can only enter the tree structure once while in CART, variables can enter multiple times. Both CART and SDA can weight the importance of sensitivity v. specificity at the users discretion)

SDA and CART overcome the problem of multiple collinearity among predictor variables and allow exploratory testing of higher order interactions. Furthermore, recursive partitioning methods have been shown to identify more homogenous groups of high and low risk subjects in a clinical setting than logistic regression methods. Both SDA and CART analysis will be employed separately on the ARIC data to ensure validity and robustness of the findings.

The Atherosclerosis Risk in Communities (ARIC) Study provides an opportunity to explore recursive partitioning techniques to provide clinically meaningful cutpoints that discriminate subpopulations at risk for CHD and diabetes based on criteria related to the metabolic syndrome.

5. Main Hypothesis/Study Questions:

The purpose of this manuscript is to use recursive partitioning techniques to evaluate discriminatory power of individual components of the metabolic syndrome and their cutpoints to predict incident CHD and diabetes in the ARIC cohort. Gender-stratified analyses, signal detection and CART techniques will be used to determine the optimal cut-points across all variables for maximizing both sensitivity and specificity for classifying the outcome of interest.

Advantages of these techniques over more traditional simultaneous regression approaches include the following:

- non-parametric, requires no assumptions of normal distributions of predictor variables
- identifies multiple interactions among predictors
- not adversely affected by highly correlated data, multicollinearity is not an issue as in logistic or proportional hazards regression.
- tree-based decision structure provides clinically relevant cutpoints on predictor variables that are selected by minimizing misclassification (maximize sensitivity and specificity)

Study Questions

- Which combination of the metabolic syndrome components offers the best discriminatory power in predicting who will develop CHD and diabetes?
- Are the binary cutpoints for each component defined by the ATP III and WHO definitions consistent with the cutpoints derived from the recursive partitioning methods?
- Does the tree-based structure (in terms of entry of specific components or specific cutpoint values) differ for men and women?
- Does the tree-based structure (in terms of entry of specific components or specific cutpoint values) differ for African-Americans and whites?
- Is the recently proposed change in the definition of impaired fasting glucose from 110 to 100 supported by the results of these analyses
6. Data (variables, time window, source, inclusions/exclusions):
   The study population will be taken from baseline data from the ARIC cohort. Individuals with the following conditions will be excluded:
   - Bloodwork obtained after < 8 hours fasting at the baseline visit.
   - Race other than African American or White
   - African-American participants not residing in Forsyth or Jackson centers. Race other than African American or White
   - Missing data on any component of the metabolic syndrome
   - Prevalent diabetes at the baseline visit (for analyses with incident diabetes as the outcome)
   - Prevalent CHD at the baseline visit (for analyses with incident CHD as the outcome)

   Variables will include those corresponding to individual components of the metabolic syndrome (e.g., systolic blood pressure, diastolic blood pressure, fasting insulin, fasting glucose, high-density lipoprotein (HDL), triglycerides, waist-to-hip ratio, waist circumference, and BMI.

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 ICTDER02 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 ICTDER02 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
   b. If yes, is the author aware that either DNA data distributed by the Coordinating Center must be used, or the file ICTDER02 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.escc.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)?

    The following manuscript proposals are related to the current proposal and, in each instance, the lead author has been contacted for comments and collaboration.

    #831A: Associations of new definitions of the metabolic syndrome with cardiovascular disease and atherosclerosis (cross-sectional analysis) (McNeill)

    #831B: Associations of new definitions of the metabolic syndrome with cardiovascular disease and atherosclerosis (prospective analysis). (McNeill)
11. 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.

The proposed writing group agrees to this timeline.

References


