ARIC Manuscript Proposal # 1661

PC Reviewed: 6/8/10	**Status: A**	**Priority: 2**
SC Reviewed: _____	**Status: _____**	**Priority: ____**

1.a. Full Title:
Novel risk factors and the prediction of type 2 diabetes in the Atherosclerosis Risk in Community Study (ARIC)

b. Abbreviated Title (Length 26 characters):
Risk factors and diabetes

2. Writing Group:
Writing group members: Laura A. Raynor, James S. Pankow, Bruce Duncan, Maria Ines Schmidt, Mark Pereira, Ron Hoogeveen, Christie Ballantyne, J. Hunter Young

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

F**irst author**: **Laura A. Raynor**
Address: Division of Epidemiology and Community Health
University of Minnesota
1300 S. Second Street, Suite 300
Minneapolis, MN 55454-1015
Phone: 612-626-7921 Fax: 612- 624-0315
E-mail: rayno007@umn.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: **James S. Pankow**
Address: Division of Epidemiology and Community Health
University of Minnesota
1300 S. Second Street, Suite 300
Minneapolis, MN 55454-1015
Phone: 612-624 2883 Fax: 612-624-0315
E-mail: panko001@umn.edu

3. Timeline:
June 2010: Proposal submitted
June-September 2010: Analysis

October-November 2010: Writing of the paper
December 2010: Revisions by co-authors
January 2011: Submission to journals

4. Rationale:

Type 2 diabetes is an epidemic in the United States, affecting around 7% of the population and leading to significant increases in morbidity, mortality, and long-term healthcare costs (Rahman et al. 2008; Kolberg et al. 2009). Clinical trials have shown that interventions in individuals at high risk can delay the development of diabetes; however, a screening tool is necessary to identify individuals at high-risk so that these pharmacological and lifestyle interventions can be initiated (Tuomilehto et al. 2001; Knowler et al. 2002). Fasting plasma glucose and the oral glucose tolerance tests have been used most often for risk assessment but, by the time that glucose regulation is identified as abnormal, the underlying disease have been progressing for years and complications have already occurred (Kolberg et al. 2009).

Thus, solely utilizing glucose measures to identify individuals at risk prior to disease progression is inadequate, resulting in the development of several risk scores that assess risk using multiple variables (Kolberg et al. 2009). These diabetes prediction models often include combinations of measurements of obesity, glycemia, age, sex, ethnicity, family history, systolic blood pressure, triglycerides, and high density lipoprotein (HDL) cholesterol. The area under the curve (AUCs) for these prediction models ranges from 0.66 to 0.84, indicating that there is room for improvement (Sattar et al. 2008).

A number of novel risk factors have been found to be associated with type 2 diabetes in recent studies including inflammatory markers, markers of endothelial dysfunction, adipose-derived markers, hepatic-derived markers, and genetic markers. To date, there have been no risk prediction studies that have added the majority of these novel risk factors to a model and tested the contribution of these predictors in the context of the aforementioned existing predictors.

References

1. Knowler WC et al. Reduction in the incidence of type 2 diabetes with lifestyle intervention or meformin. New England Journal of Medicine. 2002;346(6):393-403.
2. Kolberg JA et al. Development of a type 2 diabetes risk model from a panel of serum biomarkers from the Inter99 cohort. Diabetes Care. 2009;32(7):1207-12.
3. Rahman M et al. A simple risk score identifies individuals at high risk of developing Type 2 diabetes: a prospective cohort study. Family Practice. 2008;25(3):191-6.
4. Sattar et al. Can metabolic syndrome usefully predict cardiovascular disease and diabetes? Outcome data from two prospective studies. Lancet. 2008;371(9628):1927-35.
5. Tuomilehto J et al. Prevention of type 2 diabetes mellitus by changes in lifestyle among subjects with impaired glucose tolerance. New England Journal of Medicine. 2001;3(344):1343-50.

5. Main Hypothesis/Study Questions:

We will investigate the predictive ability of clinical and genetic risk factors for incident type 2 diabetes by creating prediction models for two ARIC populations. The first model will utilize the ARIC case-cohort ancillary study, which had an extensive collection of novel biomarkers measured from baseline blood samples in approximately 1200 subjects. The second analysis will include members of the full cohort that had a more limited set of risk factors measured at baseline.

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).

An individual will be considered to have diabetes if they report a history of physician-diagnosed diabetes, current use of anti-diabetes mediation, have a fasting serum glucose ≥126 mg/dL, or have a non-fasting glucose ≥200 mg/dL. For both analyses we will start with a simple prediction model, previously validated in ARIC, that includes parental history of diabetes, hypertension, race/ethnicity, age, fasting glucose, triglycerides, systolic blood pressure, use of anti-hypertensive medications, height, high density lipoprotein cholesterol, and waist circumference (Schmidt et al. 2003).

The expanded model for the case cohort sub-sample will consider the following measures obtained from visit 1 samples:

- Total adiponectin
- High molecular weight adiponectin
- Leptin
- Gamma-glutamyl transpeptidase (GGT)
- Alanine aminotransferase (ALT)
- Fetuin A
- Ferritin
- C-reactive protein (CRP)
- Intercellular adhesion molecule 1 (ICAM-1)
- Orosomucoid
- Sialic acid
- Interleukin 6
- Interleukin 18
- Complement C3
- Asymmetric dimethylarginine (ADMA)
- Retinol binding protein 4 (RBP-4)
- Free fatty acids
- Oxidized LDL
- Lactate

.
The expanded model for the full cohort will consider the following measures obtained at visit 1 and reported to be associated with incident diabetes in previous ARIC publications:

- White blood cell count (WBC)
- Fibrinogen
- Albumin
- von Willebrand factor antigen (vWF)
- Activated partial thromboplastin time (aPTT)
- Factor VIII coagulant activity
- Serum magnesium
- Forced vital capacity (FVC)
- Forced expiratory volume (FEV)
- Total blood viscosity
- Hematocrit level
- Leg length
- Hip circumference
- Heart rate
- Low frequency power heart rate variability
- Genetic risk score

Continuous variables will be tested for normality and log transformed if found to be non-normal or may be categorized if there is strong clinical evidence in support of categorization. A risk allele score will be constructed that is the sum of the number of risk alleles for each participant (0, 1, or 2) that has complete genotype information. This risk allele assumes all genetic variants have the same effect size. The genetic risk score will include variants from the following thirty-two genes: NOTCH2 (rs10923931), THADA (rs7578597), PPARG (rs1801282), ADAMTS9 (rs6795735), IGF2BP2 (rs1470579), WFS1 (rs10010131), CDKAL1 (rs10440833), JAZF1 (rs849134), SLC30A8 (rs13266634), CDKN2A/B (rs10811661) , CDC123/CAMKID (rs12779790), HHEX/IDE (rs5015480), TCF7L2 (rs7903146), KCNQ1 (rs163184), KCNJ11 (rs5215) , TSPAN8/LGR5 (rs7961581), FTO (rs9939609), HNF1B (rs75210), MTNRB1 (rs1387153), IRS1 (rs7578326), ZBED3 (rs4457053), KLF14 (rs972283), TP53INP1 (rs896854), CHCHD9 (rs13292136), KCNQ1 (rs231362), CENTD2 (rs1552224), HMGA2 (rs1531343), HNF1A (rs7957197), ZFANDG (rs11634397), PRC1 (rs8042680), and DUSP9 (rs5945326).

The case cohort sample and total cohort will be randomly divided into two sets of equal size, a training dataset and a testing dataset for analysis. Associations of individual risk factors for the prediction of incident type 2 diabetes in the training dataset will be investigated using multivariate Cox proportional hazards regression with forward stepwise regression. Akaike's information criterion (AIC) and the -2 log likelihood model fit statistics will be used to select between models and Harrell's C statistic will be used to determine the discriminative accuracy of the models. We will calculate the net-reclassification-improvement statistic and the integrated-discrimination-improvement statistic to evaluate the overall improvement in risk classification with the addition of

novel risk factors. The testing dataset will be used to validate the predictive models obtained from the training set, comparing the areas under the receiver operating curves, specificity, sensitivity, and predictive values. Finally, bootstrapping will be used to generate confidence estimates for the area under the curve using repetitive sampling and replacement of individuals in the testing dataset.

7.a. Will the data be used for non-CVD analysis in this manuscript?
X Yes ____ 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?
_X_ Yes ____ No
(This file ICTDER03 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)?
1. ARIC proposal #1349 Association of blood lactate with insulin resistance and type 2 diabetes: The Atherosclerosis Risk in Communities Carotid MRI Study
2. ARIC proposal #1558 Retinol Binding Protein 4 (RBP-4) in relation to the risk of type 2 diabetes in the ARIC study
3. ARIC proposal #1543 Relationship between circulating levels of Fetuin-A and risk of type 2 diabetes in the ARIC study
4. ARIC proposal #1273 Genetic risk score for type 2 diabetes
5. ARIC proposal #977 Liver Enzyme Activity and Risk of Diabetes
6. ARIC proposal #976 Adiponectin, complement component 3, leptin and incident diabetes mellitus (Ancillary study)

7. ARIC proposal #808 Prediction of diabetes mellitus and impaired glucose tolerance in middle-aged adults: The Atherosclerosis Risk in Communities Study
8. ARIC proposal #862 Fasting plasma non-esterified fatty acids (NEFA) and risk of type 2 diabetes
9. ARIC proposal #1492: High-molecular-weight adiponectin and the risk of type 2 diabetes in the ARIC study
10. ARIC proposal #946 Association of plasma ferritin and incident diabetes
11. ARIC proposal #406 Serum magnesium concentration and the risk of incident non-insulin dependent diabetes mellitus
12. ARIC proposal #912 Viscosity and incidence of type 2 diabetes mellitus
13. ARIC proposal #825 Pulmonary function and type 2 diabetes mellitus
14. ARIC proposal #900 Is height related to incident diabetes mellitus?
15. ARIC proposal #883 Cardiac autonomic function and the development of incident diabetes
16. ARIC proposal #853 IL-6, acute phase proteins and incident diabetes mellitus (Ancillary study)
17. ARIC proposal #564 Acute phase response lipids/lipoproteins as predictors of incident diabetes

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

X **A. primarily the result of an ancillary study (list number***
1995.09 Inflammatory Precursors of Type 2 Diabetes
2006.03 GWA for loci influencing incident CHD and other HLB phenotypes (NHLBI RFA for large scale genotyping) (STAMPEED) (GEI)
2007.02 The National Heart Lung and Blood Institute's Candidate Gene Association Resource (CARE): Phase I (CARE)

___ 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/

12. 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.