1. Full Title: Cardiovascular Disease Global Risk Score Development for Diabetes Mellitus from Five Multiethnic US Cohorts

Note: This is a pooled analysis proposal involving multiple cohorts. Also see attached MESA P&P approval letter.

b. Abbreviated Title (Length 26 characters): CVD Risk Score in Diabetes

2. Writing Group:

Writing group members: Yanglu Zhao, MD, MS, Ralph D’Agostino, PhD, Alain Bertoni, MD, Matthew Budoff, MD, Aaron Folsom, MD, David Jacobs, PhD, Elizabeth Selvin, PhD, Herman Taylor, MD, and Nathan D. Wong, PhD (senior/corresponding author)

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

First author: Yanglu Zhao, MD, MS (Senior author: Nathan D. Wong, PhD)

Address: Heart Disease Prevention Program, Division of Cardiology, C240 Medical Sciences, University of CA, Irvine, CA 92697-4079
Phone: 949-824-5561 Fax: 949-824-2200
E-mail: yangluz@uci.edu / ndwong@uci.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: Aaron Folsom
Address: University of Minnesota
Division of Epidemiology and Community Health
1300 South Second Street
Suite 300
Minneapolis, MN 55454
Site Phone: 612-626-8862
Office Phone: 612-626-8862
Fax: 612-624-0315
folso001@umn.edu

3. Timeline: July 1, 2017 – December 31, 2018

4. Rationale:

Patients with diabetes mellitus (DM) are at 2-4 times increased risk of developing cardiovascular disease (CVD) including coronary heart disease (CHD). But there is great heterogeneity in risk among patients with DM and they should not be assumed as CHD risk equivalents. A meta-analysis of over
45,000 patients showed those with DM but without a prior MI had a 43% lower risk of future CHD events compared to those with a prior MI but no DM [1]. Risk assessment is important in preventive cardiology and used to appropriately target intensity of therapy [2]. Accurate risk assessment is crucial for guidelines and treatment algorithms to help determine appropriateness of further diagnostic tests and preventive therapy.

Current risk assessment for DM patients is mainly based on risk scores derived from the general population, such as the Framingham Risk Score (FRS) for CVD or the 2013 AHA/ACC Pooled Cohort Equation (PCE) for hard atherosclerotic cardiovascular disease [3,4], or are from other countries, such as the U.K. Prospective Diabetes Study (UKPDS) risk engine [5]. These risk scores have inadequate calibration or discrimination from external validation, with a tendency to overestimate the risk in modern populations. Based on this need for improved tailored risk assessment for persons with DM, several risk engines for patients with DM have been developed [5-8], but only one has been specifically developed in the US population, which used the ARIC cohort for CVD risk and was relatively limited in sample size [9].

5. Main Hypothesis/Study Questions:

**Primary Objective:**

1) To develop a new risk score calculator of 10-year CVD event prediction for the US diabetes population using five US cohorts (MESA, ARIC, JHS, CARDIA and FHS Generation 3) and externally validate the risk scores using the control arm of the Action to Control Cardiovascular Risk in Diabetes (ACCORD) study cohort.

**Secondary Objectives:**

1) To develop and validate risk score calculators for 10-year incident CHD events, incident stroke and incident heart failure (HF) in the US diabetes population.

2) To compare performance regarding discrimination and calibration between the above DM specific risk score and existing risk scores for CVD (AHA/ACC pooled cohort equation for ASCVD, Framingham Risk Score for total CVD), CHD (Framingham risk score for CHD, UKPDS for CHD), stroke (Framingham stroke risk score, UKPDS for stroke) and heart failure (Framingham HF risk score).

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).

**Data / Sample Selection:**

1) We plan to pool five NIH sponsored population-based cohorts to increase the study sample size, expand the age range and widen ethnic diversity: ARIC, MESA, CARDIA, JHS and FHS Gen 3. The summary of each study were shown in Table 1.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline Age</td>
<td>45-64</td>
<td>45-84</td>
<td>18-30</td>
<td>21-84</td>
<td>19-72</td>
</tr>
<tr>
<td>Sample Size</td>
<td>15,792</td>
<td>6,814</td>
<td>5,115</td>
<td>5,302</td>
<td>4,095</td>
</tr>
<tr>
<td>App.DM sample</td>
<td>*892</td>
<td>*740</td>
<td>332</td>
<td>*451</td>
<td>150</td>
</tr>
<tr>
<td>Race/Ethnicity</td>
<td>Caucasian</td>
<td>Caucasian</td>
<td>Caucasian</td>
<td>African-</td>
<td>Caucasian</td>
</tr>
</tbody>
</table>
2) We will include subjects aged 40-79 years with DM and free of known CVD at baseline. DM is defined as (1) physician diagnosed DM; (2) use of insulin or oral medication; (3) fasting blood glucose ≥ 7.0 mmol/l (126 mg/dl); (4) 2h oral glucose tolerance test ≥ 11.1 mmol/l (200 mg/dl); and/or (5) a HbA1c ≥ 6.5% (48 mmol/mol) at the time of (or earlier than) the identified baseline visit where HbA1c and other risk factor information were available. Cohort participants will be excluded if they had a prevalent CVD event at the designated baseline as noted above (MI, stroke, heart failure, percutaneous coronary intervention, or bypass surgery).

3) The primary outcomes of interest will be ASCVD events, and secondary endpoint are CHD, stroke and HF. According to the designated baseline exam we will use for our study, maximum follow-up time in years will be approximately 25 years for ARIC, 15 years for Jackson, 12 years for MESA, 10 years for CARDIA and 13 years for FHS Gen3.

4) Our new cohort is anticipated to have approximately 2600 DM patients with a minimum of 10 years of follow-up and CVD events with over 800 CVD events expected. We examined the latest follow-up data in MESA to get an estimate of individual endpoints. Among 6809 MESA participants (including both DM and non-DM) with a maximal follow-up of 14 years and mean follow-up of 11.7 years, 283 MI events, 267 stroke events and 308 HF events occurred. From our pooled cohort sample of 2700 DM subjects, we estimate the projected MI, stroke and HF events will be 226, 214 and 246 assuming DM has at least twice of CVD risk as the non-DM patients, which will allow for approximate 20 risk factors to be included in the model.

5) The following risk factors will be examined in the risk prediction model: Age, sex, race/ethnicity, body mass index, diabetes duration, premature family history of CVD, hemoglobin A1c (HbA1c), atrial fibrillation, urinary albumin/creatinine ratio, use of lipid-lowering medication, antihypertensive treatment, smoking status, albuminuria, estimated glomerular filtration rate (eGFR), systolic and diastolic blood pressure, high-density lipoprotein and low-density lipoprotein cholesterol, triglycerides, hs-CRP.

**Analytic Plan:**

1) We will calculate our risk scores separately in gender and race subgroups (blacks, whites, others as sample sizes permit). We will use a two-step method to determine the inclusion and exclusion of a certain factors: Random survival forest analysis will be applied to rank the risk factors according to their permutation importance, and predictors with a negative importance score were excluded from further analysis.[15] After the first round of selection, the selected factors, their higher power terms and interaction terms will be examined in the
Cox regression model with hazard ratio (HR), Harrell c-statistics, net reclassification index (NRI) and integrated discrimination improvement (IDI) index. Covariates with statistical significance for above measures will be remained in the final model.

2) Calculation of absolute risk score. The Cox proportional hazard regression model will be used to produce both relative risk (hazard ratio, or the exponential of beta coefficient) and an estimation of the absolute risk of an event occurring at 10 years. Mean Survival rate is calculated as \( S_{10} \) in each gender and race group. Given the individual value of other risk factors, an individual’s estimated absolute risk score is calculated as:

\[
1 - S_{10}^{\exp(\text{Individual}\times\beta - \text{Mean}\times\beta)}
\]

3) Evaluation of risk score performance. In this step, we will evaluate the performance of the new risk score by discrimination, calibration and internal validity. Discrimination will be evaluated using Harrell’s c-statistics, an equivalent to AUC specifically designed for censoring data, NRI and IDI. [16] Calibration for time-to-event analyses can be evaluated by the Nam-D’Agostino Chi-Square test [17].

4) Internal validity will be examined using bootstrap method. Split sample method and cross validation method are also available to use but past experience from the Framingham study showed that these methods produced very similar results.

5) External validity will be tested in ACCORD cohort [18]. In ACCORD, 1370 subjects received standard HbA1c + standard lipid treatment and 1178 received standard HbA1c + standard blood pressure treatment, resulting in 2548 eligible subjects to be external validation cohort. We will recalculate the basic risks at 5 years \( S^5 \) in ACCORD population to recalibrate the risk score. We will use c-statistics and Nam-D’Agostino chi-square test to examine performance.

6) Comparison of old and new risk scores. We will also compare performance regarding discrimination and calibration between above DM specific risk score and existing 10-year risk scores for CVD (AHA/ACC pooled cohort equation for ASCVD [2], Framingham Risk Score for total CVD [1]), CHD (Framingham risk score for CHD [19], UKPDS for CHD [3]), stroke (Framingham stroke risk score [20], UKPDS for stroke [21]) and heart failure (Framingham HF risk score [22]). Again we will use c-statistics, NRI and Nam-D’Agostino chi-square test to examine the performance of risk scores.

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 __X__ 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)? Parrinello CM, Matsushita K, Woodward M, Wagenknecht LE, Coresh J, Selvin E. Risk prediction of major complications in individuals with diabetes: the Atherosclerosis Risk in Communities Study. Diabetes Obes Metab. 2016 Sep;18(9):899-906.

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.

References Cited
1. Bulugahapitiya U, Siyambalapitiya S, Sithole J, Idris I. Is diabetes a
Date: April 12, 2017

To: Yanglu Zhao

CC: Nathan Wong  
    Alain G. Bertoni  
    Ralph B. D’Agostino Sr.  
    Elizabeth Selvin  
    Matthew Budoff  
    David Jacobs Jr.  
    Aaron Folsom

From: Karen Hansen  
      P&P Committee Program Coordinator  
      CHS Coordinating Center

Re: Manuscript Proposal: Cardiovascular Disease Global Risk Score Development for Diabetes Mellitus from Five Multiethnic US Cohorts

The MESA P&P Committee has reviewed and approved your manuscript proposal shown above. The proposal has been numbered ME 352 in the database. In future correspondence with the P&P Committee, please include this assigned manuscript number.

Since this is a pooled/meta-analysis paper, no further review is required. An informational copy along with the P&P comments will be sent to the MESA Steering Committee. With this approval, you can now begin work on your pen draft.

For your easy reference, I’ve listed the streamlined process below for your meta-analysis paper (per the “How to Submit a Pen Draft” document located on the internal P&P Web page). Please note this when you submit the pen draft:

- The pen draft should be submitted per the usual process with a note in the submission e-mail that it is a pooled/meta-analysis paper. (See the requirements for pen drafts at the end of this memo.)

- The pen draft will receive electronic P&P review by a subset of committee members, with the main focus to ensure that MESA isn’t misrepresented and that there aren’t any major mistakes.

- Once P&P approval is received for the pen draft, the author can submit the manuscript to a journal. (Only an informational copy along with the reviewers’ comments will be submitted to the SC.)
There weren’t any P&P comments.

IMPORTANT:
Please check the acknowledgements information available under the “Acknowledgements and Methods Descriptions” subheading on the internal P&P Web page as you prepare your pen draft.

As chair of the writing group, it is your responsibility to see that this paper is completed on schedule and that all coauthors (including SC nominated authors) are included in the process. The P&P Committee, however, will assist you and your writing group in any way possible in the preparation of your manuscript. Thank you again for your efforts on behalf of MESA.

All coauthors should let the first author know of a change in contact information. Failure to respond within a reasonable amount of time to a first author’s request for coauthor feedback, could result in removal from the writing group.

Please note the following P&P policies related to data use agreements and access to MESA data:

- Before an author can get access to the MESA data he/she must complete a data use agreement. Copies of the forms and instructions are available on the internal P&P Web page at: http://mesa-nhlbi.org/MesaInternal/Publications.aspx
- Enter the usual internal site login (user name followed by the password).
- Scroll down to the “Data Use Agreement Forms (DDA and DMDA)” subheading.
- David Vu at voodoo@uw.edu processes the data use agreements and is available to answer questions.
- Once a completed data use agreement form has been sent to the Coordinating Center, an author can contact David Vu at (voodoo@uw.edu) or 206-897-1913. (For security reasons the password cannot be emailed. It is only provided via phone or secure web page.)

Please note the following P&P policies related to submission of an abstract based on an approved proposal:

- Abstracts must be submitted via the MESA online abstract submission form (not as an e-mail attachment).
- Please ensure that abstracts are based exclusively on an approved MESA proposal. (This means that the scope of the abstract needs to be based entirely on the content of the originally approved proposal.)
- The P&P abstract submission deadline is two weeks prior to the conference submission deadline. (This deadline remains as stated regardless of any holidays that may fall on or before it.)
- It is strongly advised that authors submit abstracts well before the deadline in order to allow sufficient time for revisions.

Please note the following P&P policies related to submission of a pen draft based on an approved proposal:

- Send the pen draft to the P&P Coordinator as an e-mail attachment.

Include the following required information with each new pen draft:

1. MESA manuscript number (examples: MC 001, AC 025)
2. In January 2015 the committee set a limit of only 2 separate documents (excluding the lay summary) for pen draft submissions.

3. Confirmation that all coauthors have seen and approved the manuscript prior to submission

4. Specify one target journal that the author is thinking of submitting the manuscript to

5. Lay summary (see below for details)

As of October 2007 authors are also required to attach a separate lay summary (Word) document when submitting a new pen draft. In some cases, the lay summary may be featured on the participant website, or included in the participant newsletter, and should meet the following criteria:

- Please describe the main findings of this research using plain language that can be understood by our participants. For some summaries, this may be brief and involve just a few sentences (perhaps 100-200 words).
- If an author believes that a manuscript is not appropriate for a lay audience, please state so.

SUGGESTED QUESTIONS WHICH COULD BE ADDRESSED (Not all need to be answered in each lay summary):

1. Is there any interesting previous research or common thought about the topic? If so, what was the prevailing opinion? Does this research confirm or refute the previous prevailing opinion?

2. What MESA component (data) was used in the research (e.g. blood draw from exam 1, cardiac MRI procedure...)?

3. Was an ancillary study or special cohort involved? If so which one?

4. Was there an important finding for doctors? If so, what was it?

5. Was there an important finding for the general public? If so, what was it?

6. Would future research be helpful to add to or better explain the findings? If so, what would be involved in the future research?

7. Is there anything else that you think would be of general interest to MESA participants?

LAY SUMMARY EXAMPLES:

“This paper investigates whether healthy food availability in neighborhoods is related to the diet of residents. We surveyed food stores in the neighborhoods of Baltimore MESA participants. The healthy food availability at different food stores was characterized using a special survey. People living in areas with better food availability had better diets. Improving healthy food availability may be an important component of strategies to improve diet and prevent heart disease.”

“Diabetic retinopathy, a leading cause of visual impairment in working-aged adults affect one in three participants with diabetes. African and Hispanic Americans were more likely to have diabetic retinopathy than white and Chinese Americans. We have identified several modifiable risk factors of diabetic retinopathy, including higher blood sugar levels and abdominal obesity.”

If you would like further information, please contact Karen Hansen, P&P Committee Program Coordinator, at (206) 897-1939, email: hansenk3@u.washington.edu.