ARIC Manuscript Proposal #2485

PC Reviewed: 1/13/15   **Status: <u>A</u>**   **Priority: <u>2</u>**
SC Reviewed: _____   **Status: _____**   **Priority: ____**

1.a. Full Title: Risk prediction of microvascular and macrovascular complications and all-cause mortality in persons with diabetes

  b. Abbreviated Title (Length 26 characters): Predicted risk in diabetes

2.   Writing Group:
     Writing group members: Christina M. Parrinello; Kunihiro Matsushita; Mark Woodward; Michael Steffes; Josef Coresh; Elizabeth Selvin; others welcome

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

First author: **Christina Parrinello**
Address: 2024 E. Monument Street
            Suite 2-600
            Baltimore, MD 21287

            Phone: (443) 287-4679   Fax:
            E-mail: cparrin1@jhu.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: **Elizabeth Selvin**
    Address: 2024 E. Monument Street
            Suite 2-600
            Baltimore, MD 21287

            Phone: 410-614-3752   Fax: 410-955-0476
            E-mail: eselvin@jhu.edu

3.   Timeline: Upon approval of this manuscript proposal, we expect to be able to submit this completed manuscript for review within one year.

4. Rationale:

Diabetes is an established risk factor for micro- and macrovascular complications, including chronic kidney disease (CKD) and end-stage renal disease (ESRD), coronary heart disease (CHD), stroke, heart failure, and mortality.[1–5] Diabetes is currently considered a "CHD risk equivalent",[6] which implies that all persons with diabetes are in a high-risk category similar to those persons who have a history of CHD. However, recent evidence suggests that persons with diabetes may have varying degrees of risk depending on the presence and severity of other risk factors and co-morbidities, and that HbA1c and other biomarkers not included in traditional cardiovascular risk equations may potentially inform this risk.[7] Aggressive treatment of hyperglycemia and other cardiovascular risk factors such as hypertension may not benefit all individuals equally and is associated with adverse events (e.g., hypoglycemia and hypotension). It is crucial to distinguish between persons who will and will not benefit from such aggressive treatment. Risk prediction equations can identify those individuals at high risk for certain outcomes and may have utility to inform certain treatment interventions.

Risk prediction models for microvascular and macrovascular complications are of clinical interest. Several models have been developed to predict ESRD in persons with diabetes complicated by CKD,[8–10] as well as in persons with diabetes who do not have kidney disease.[11,12] Many more risk prediction models have been developed for macrovascular complications. Risk scores for CVD developed in the general population have tended to underestimate risk when applied to persons with diabetes.[13,14] To improve predictive accuracy, several risk scores for CVD have been developed in persons with diabetes.[15–22] A risk score for CHD was previously developed in ARIC in persons with diabetes, but this model was never externally validated, nor did it include HbA1c or nontraditional markers of hyperglycemia, cardiac damage, kidney function, inflammation or liver function.[18]

Whereas there is some evidence that scores developed in persons with diabetes may better discriminate risk compared to those developed in the general population,[23] many of these have still been problematic when applied to external populations of persons with diabetes. First, many have not been able to accurately predict risk in external populations.[24] For instance, the well-known United Kingdom Prospective Diabetes Study (UKPDS) risk engine, which is a risk prediction tool for CHD and stroke in persons with diabetes,[15,20] has been shown to greatly overestimate risk (by up to 5 fold) in external populations, which could lead to unnecessary treatment in persons with diabetes.[25] Second, most of these risk scores have been developed in white European populations,[13] which may limit their generalizability. Third, having to use multiple risk scores to predict risk of diabetes complications is burdensome for practitioners,[26,27] whereas a risk prediction tool that comprehensively predicts risk of multiple diabetes complications may be convenient for clinical use. A recent paper developed a risk prediction model for multiple endpoints that included micro- and macrovascular complications in a Japanese population, and found that combining these outcomes improved classification of persons into low- and high-risk groups.[28] Fourth, most of these risk scores have not incorporated the use of nontraditional markers of hyperglycemia, cardiac damage, kidney function,

inflammation or markers of liver function,[13] which have been associated with increased risk of complications in persons with diabetes.[8–12,29,30]

5.   Main Hypothesis/Study Questions:

Aim: To develop and internally validate a comprehensive risk prediction equation for 10-year risk of key micro- and macro-vascular complications and death in persons with diabetes using demographic and clinical information and a panel of traditional and nontraditional markers of hyperglycemia, cardiac damage, kidney function, inflammation, and liver function.

Hypothesis: The addition of traditional and nontraditional biomarkers improves the prognostic ability of the risk prediction equations compared to non-laboratory markers alone.

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).

Study population

We will include ARIC participants with diabetes at visit 2 (defined by self-reported diagnosis or medication use), excluding those with prevalent CVD or reduced kidney function (N~800). Prevalent CVD will be defined using self-reported history of CHD, stroke or heart failure at visits 1 or 2, or hospitalization for any of these events prior to visit 2. Prevalent reduced kidney function will be defined using creatinine-based estimated glomerular filtration rate <60 mL/min/1.73 m^2 at visit 2. We will further exclude non-black and non-white participants, as well as black participants from Minneapolis or Washington County, due to small numbers, as well as persons who are missing key covariates.

Demographic and clinical measurements

Models will include age, sex and race-center. We may additionally include the following variables during model development (described later): duration of diabetes, family history of CVD, education level, alcohol consumption, smoking status, physical activity (Baecke sport activity index), systolic blood pressure, diastolic blood pressure, antihypertensive medication use, cholesterol-lowering medication use, low-density lipoprotein cholesterol, high-density lipoprotein cholesterol, total cholesterol, body mass index and waist circumference.

Laboratory measurements

We will consider inclusion of both traditional and nontraditional laboratory markers of hyperglycemia, kidney function, cardiac damage, inflammation and liver function during model development. These will include:

	Traditional	**Nontraditional**
Hyperglycemia	Fasting glucose, HbA1c	Fructosamine, Glycated albumin, 1,5-anhydroglucitol
Kidney function	Creatinine	ß-2 microglobulin, Cystatin C
Cardiac damage	--	hs-cTnT, NT-proBNP
Liver function	ALT, AST	GGT
Inflammation	--	hs-CRP

Outcomes

Microvascular and macrovascular events and all-cause mortality are ascertained via continuous surveillance of hospitalizations and death certificates, annual telephone follow-up with the participant or a proxy, and linkage with the National Death Index. CHD will be defined as a first occurrence of either adjudicated hospitalization for definite/probable myocardial infarction or CHD death.[31] Incident stroke will be defined as a first occurrence of adjudicated hospitalization or definite/probable ischemic stroke death.[32] Incident heart failure will be defined as a first occurrence of either hospitalization with a discharge code of 428 in any position for diagnosis using the International Classification of Diseases, 9th Revision (ICD-9) or heart failure death, based on a 428 ICD-9 code or an ICD, 10th Revision (ICD-10) code of I50.[33] Incident CKD will be defined as eGFR<60 mL/min/1.73 m^2 and $\geq$25% decline in eGFR since visit 2, or hospitalization due to kidney disease, kidney transplant or dialysis or death due to kidney disease.

Statistical analysis

We will use a 3-level approach for model development: 1) including non-laboratory demographic and clinical characteristics; and adding 2) traditional markers and 3) nontraditional markers of hyperglycemia, cardiac damage, kidney function, inflammation and liver function alone and in combination with one another.

We will conduct analyses using standard Cox proportional hazards models with a combined endpoint (so modeling time to first of any event) or using cause-specific proportional hazards models for comparison. The cause-specific hazards model censors persons with the competing event(s), just as in standard survival analysis. However, the absolute risk (cumulative incidence) is calculated differently.[34] We may also compare results using a competing risk framework by fitting a proportional hazards model for the subdistribution (or Fine and Gray model). Standard survival analysis, including the Kaplan-Meier method and Cox proportional hazards regression, overestimates

cumulative incidence when competing risks are present, especially when the competing risks are strong, such as in the setting of diabetes. These approaches may therefore affect the calibration of the risk prediction model, more so than the discrimination.[34,35] Therefore, the Fine and Gray method may more accurately assess both discrimination and calibration, since accurate determination of absolute risk is vital for clinical prognosis and treatment decisions.

To visually assess the discrimination of the model, we will create histograms of the predicted risk in events and non-events, separately. We will use the following measures of discrimination to assess incremental improvements in prediction: 1) the Harrell's c-statistic, which accounts for censoring in survival analysis; 2) the overall net reclassification improvement (NRI) to quantify upward and downward reclassification, as well as the event and nonevent NRI separately, in order to determine the amount and direction of reclassification separately in people who do and do not experience an event; and 3) the integrated discrimination improvement (IDI) and relative IDI to assess the improvement in average sensitivity.[36–42]

We will assess the calibration of the model using: 1) extensions of the Hosmer-Lemeshow goodness of fit test for survival data; and 2) calibration curves for competing risks models.[43] Additionally, we will create risk distribution plots, which display the distribution of risks calculated from each model in the overall study population, as well as in events and non-events, separately. These plots allow visualization of the proportion of individuals classified as low versus high risk.[44] We will plot risk predictiveness curves (risk percentile across the X-axis and the predicted risk along the Y-axis) and overlay the observed proportions of the outcome of interest at the midpoint of each risk decile, which will graphically complement the Hosmer-Lemeshow goodness of fit test.[41]

We will internally validate the risk prediction equation using bootstrapping methods to obtain an estimate of the optimism-corrected c-statistic, which is a better estimate of the expected predictive accuracy from external validation.[45,46] We will have already calculated the c-statistic of the model developed in the original sample (c_{app}). We will then sample with replacement to create 200 bootstrapped samples of the same size as the original data set and develop a model in each of the bootstrapped samples. To calculate the optimism in each bootstrapped sample, we can find the difference in the c-statistic of the model applied to the bootstrapped sample (c_{boot}) and applied to the original sample (c_{orig}). The average optimism can then be calculated as the sum of the optimisms from each bootstrapped sample divided by the number of bootstrapped samples. Lastly, the optimism-corrected c-statistic is the difference of the c-statistic from the original model applied to the original sample and the average optimism. Compared to the more common method of splitting the data into training and validation data sets, or even cross-validation (repeated splitting of the data into training and validation data sets), bootstrapping should result in less biased estimates and smaller variance. Additionally, it allows for use of the entire data set to develop the model.

Sensitivity analyses may additionally consider stratification by race and/or duration of diabetes.

Limitations

We are limited to using a baseline visit at which all proposed analytes were measured. Therefore, we have used visit 2 as the baseline. Whereas it would be informative to use visit 4 as the baseline, since data are more recent, not all biomarkers of interest were assayed using store specimens from that visit (HbA1c, in particular). Nonetheless, our results will be informative for developing a useful risk prediction equation.

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)?
Published articles:

Selvin E, Rawlings AM, Grams M, Klein R, Steffes M, Coresh J. Association of 1,5-anhydroglucitol with diabetes and microvascular conditions. Clin Chem. 2014 Nov;60(11):1409-18.

Selvin E, Rawlings AM, Grams M, Klein R, Sharrett AR, Steffes M, Coresh J. Fructosamine and glycated albumin for risk stratification and prediction of incident diabetes and microvascular complications: a prospective cohort analysis of the Atherosclerosis Risk in Communities (ARIC) study. Lancet Diabetes Endocrinol. 2014 Apr;2(4):279-88.

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* 2009.16)
  ___   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.

REFERENCES

1. Standards of medical care in diabetes--2014. *Diabetes Care* 2014;37 Suppl 1:S14-80.

2. National Kidney Foundation. KDOQI Clinical Practice Guideline for Diabetes and CKD: 2012 Update. *Am. J. Kidney Dis.* 2012;60(5):850-86.

3.   Foster MC, Rawlings AM, Marrett E, et al. Cardiovascular risk factor burden, treatment, and control among adults with chronic kidney disease in the United States. *Am. Heart J.* 2013;166(1):150-6.

4.   Selvin E, Juraschek SP, Coresh J. Kidney disease in people with diabetes: the expanding epidemic. *Am. J. Kidney Dis.* 2012;59(3):340-2.

5.   Fowler MJ. Microvascular and Macrovascular Complications of Diabetes. *Clin. Diabetes* 2008;26(2):77-82.

6.   Third Report of the National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III) final report. *Circulation* 2002;106(25):3143-421.

7.   Paynter NP, Mazer N a, Pradhan AD, Gaziano JM, Ridker PM, Cook NR. Cardiovascular risk prediction in diabetic men and women using hemoglobin A1c vs diabetes as a high-risk equivalent. *Arch. Intern. Med.* 2011;171(19):1712-8.

8.   Keane WF, Zhang Z, Lyle PA, et al. Risk scores for predicting outcomes in patients with type 2 diabetes and nephropathy: the RENAAL study. *Clin. J. Am. Soc. Nephrol.* 2006;1(4):761-7.

9.   Desai AS, Toto R, Jarolim P, et al. Association between cardiac biomarkers and the development of ESRD in patients with type 2 diabetes mellitus, anemia, and CKD. *Am. J. Kidney Dis.* 2011;58(5):717-28.

10.   Krolewski AS, Warram JH, Forsblom C, et al. Serum concentration of cystatin C and risk of end-stage renal disease in diabetes. *Diabetes Care* 2012;35(11):2311-6.

11.   Elley CR, Robinson T, Moyes SA, et al. Derivation and validation of a renal risk score for people with type 2 diabetes. *Diabetes Care* 2013;36(10):3113-20.

12.   Jardine MJ, Hata J, Woodward M, et al. Prediction of kidney-related outcomes in patients with type 2 diabetes. *Am. J. Kidney Dis.* 2012;60(5):770-8.

13.   Chamnan P, Simmons RK, Sharp SJ, Griffin SJ, Wareham NJ. Cardiovascular risk assessment scores for people with diabetes: a systematic review. *Diabetologia* 2009;52(10):2001-14.

14.   Brindle P, Beswick A, Fahey T, Ebrahim S. Accuracy and impact of risk assessment in the primary prevention of cardiovascular disease: a systematic review. *Heart* 2006;92(12):1752-9.

15.   Stevens RJ, Kothari V, Adler a I, Stratton IM. The UKPDS risk engine: a model for the risk of coronary heart disease in Type II diabetes (UKPDS 56). *Clin. Sci. (Lond).* 2001;101(6):671-9.

16.   Donnan PT, Donnelly L, New JP, Morris AD. Derivation and validation of a prediction score for major coronary heart disease events in a U.K. type 2 diabetic population. *Diabetes Care* 2006;29(6):1231-6.

17.   Cederholm J, Eeg-Olofsson K, Eliasson B, Zethelius B, Nilsson PM, Gudbjörnsdottir S. Risk prediction of cardiovascular disease in type 2 diabetes: a risk equation from the Swedish National Diabetes Register. *Diabetes Care* 2008;31(10):2038-43.

18.   Folsom AR, Chambless LE, Duncan BB, Gilbert AC, Pankow JS. Prediction of coronary heart disease in middle-aged adults with diabetes. *Diabetes Care* 2003;26(10):2777-84.

19. Yang X, So W-Y, Kong APS, et al. Development and validation of stroke risk equation for Hong Kong Chinese patients with type 2 diabetes: the Hong Kong Diabetes Registry. *Diabetes Care* 2007;30(1):65-70.

20. Kothari V, Stevens RJ, Adler AI, et al. UKPDS 60: Risk of Stroke in Type 2 Diabetes Estimated by the UK Prospective Diabetes Study Risk Engine. *Stroke* 2002;33(7):1776-1781.

21. Elley CR, Robinson E, Kenealy T, Bramley D, Drury PL. Derivation and validation of a new cardiovascular risk score for people with type 2 diabetes: the new zealand diabetes cohort study. *Diabetes Care* 2010;33(6):1347-52.

22. Kengne AP, Patel A, Marre M, et al. Contemporary model for cardiovascular risk prediction in people with type 2 diabetes. *Eur. J. Cardiovasc. Prev. Rehabil.* 2011;18(3):393-8.

23. Echouffo-Tcheugui J-B, Kengne AP. Comparative performance of diabetes-specific and general population-based cardiovascular risk assessment models in people with diabetes mellitus. *Diabetes Metab.* 2013;39(5):389-96.

24. Simmons RK, Coleman RL, Price HC, et al. Performance of the UK Prospective Diabetes Study Risk Engine and the Framingham Risk Equations in Estimating Cardiovascular Disease in the EPIC- Norfolk Cohort. *Diabetes Care* 2009;32(4):708-13.

25. Bannister CA, Poole CD, Jenkins-Jones S, et al. External Validation of the UKPDS Risk Engine in Incident Type 2 Diabetes: A Need for New Type 2 Diabetes-Specific Risk Equations. *Diabetes Care* 2014;37(2):537-45.

26. Muller-Riemenschneider F, Holmberg C, Rieckmann N, et al. Barriers to Routine Risk-Score Use for Healthy Primary Care Patients. *Arch. Intern. Med.* 2010;170(8):719-724.

27. Dhippayom T, Chaiyakunapruk N, Krass I. How diabetes risk assessment tools are implemented in practice: A systematic review. *Diabetes Res. Clin. Pract.* 2014.

28. Tanaka S, Tanaka S, Iimuro S, et al. Predicting macro- and microvascular complications in type 2 diabetes: the Japan Diabetes Complications Study/the Japanese Elderly Diabetes Intervention Trial risk engine. *Diabetes Care* 2013;36(5):1193-9.

29. Kengne AP, Batty GD, Hamer M, Stamatakis E, Czernichow S. Association of C-reactive protein with cardiovascular disease mortality according to diabetes status: pooled analyses of 25,979 participants from four U.K. prospective cohort studies. *Diabetes Care* 2012;35(2):396-403.

30. Sluik D, Beulens JWJ, Weikert C, et al. Gamma-glutamyltransferase, cardiovascular disease and mortality in individuals with diabetes mellitus. *Diabetes. Metab. Res. Rev.* 2012;28(3):284-8.

31. White AD, Folsom AR, Chambless LE, et al. Community surveillance of coronary heart disease in the Atherosclerosis Risk in Communities (ARIC) Study: methods and initial two years' experience. *J. Clin. Epidemiol.* 1996;49(2):223-33.

32. Rosamond WD, Folsom AR, Chambless LE, et al. Stroke Incidence and Survival Among Middle-Aged Adults : 9-Year Follow-Up of the Atherosclerosis Risk in Communities (ARIC) Cohort. *Stroke* 1999;30(4):736-743.

33. Loehr LR, Rosamond WD, Chang PP, Folsom AR, Chambless LE. Heart failure incidence and survival (from the Atherosclerosis Risk in Communities study). *Am. J. Cardiol.* 2008;101(7):1016-22.

34. Wolbers M, Koller MT, Witteman JCM, Steyerberg EW. Prognostic models with competing risks: methods and application to coronary risk prediction. *Epidemiology* 2009;20(4):555-61.

35. Pencina MJ, D'Agostino RB, Larson MG, Massaro JM, Vasan RS. Predicting the 30-year risk of cardiovascular disease: the framingham heart study. *Circulation* 2009;119(24):3078-84.

36. Pencina MJ, D'Agostino RB. Overall C as a measure of discrimination in survival analysis: model specific population value and confidence interval estimation. *Stat. Med.* 2004;23(13):2109-23.

37. Pencina MJ, D'Agostino RB, Steyerberg EW. Extensions of net reclassification improvement calculations to measure usefulness of new biomarkers. *Stat. Med.* 2011;30(1):11-21.

38. Pencina MJ, D'Agostino RB, Vasan RS. Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. *Stat. Med.* 2008;27(2):157-72; discussion 207-12.

39. Zheng Y, Parast L, Cai T, Brown M. Evaluating incremental values from new predictors with net reclassification improvement in survival analysis. *Lifetime Data Anal.* 2013;19(3):350-70.

40. Pepe MS, Feng Z, Gu JW. Comments on "Evaluating the added predictive ability of a new marker: From area under the ROC curve to reclassification and beyond" by M. J. Pencina et al., Statistics in Medicine (DOI: 10.1002/sim.2929). *Stat. Med.* 2008;27(2):173-81.

41. Pepe MS, Feng Z, Huang Y, et al. Integrating the predictiveness of a marker with its performance as a classifier. *Am J Epidemiol* 2008;167(3):362-368.

42. Uno H, Cai T, Pencina MJ, D'Agostino RB, Wei LJ. On the C-statistics for evaluating overall adequacy of risk prediction procedures with censored survival data. *Stat. Med.* 2011;30(10):1105-17.

43. Gerds TA, Andersen PK, Kattan MW. Calibration plots for risk prediction models in the presence of competing risks. *Stat. Med.* 2014.

44. Pepe MS, Gu JW, Morris DE. The potential of genes and other markers to inform about risk. *Cancer Epidemiol Biomarkers Prev* 2010;19(3):655-665.

45. Moons KGM, Kengne AP, Woodward M, et al. Risk prediction models: I. Development, internal validation, and assessing the incremental value of a new (bio)marker. *Heart* 2012;98(9):683-90.

46. Kengne AP, Masconi K, Mbanya VN, Lekoubou A, Echouffo-Tcheugui JB, Matsha TE. Risk predictive modelling for diabetes and cardiovascular disease. *Crit. Rev. Clin. Lab. Sci.* 2014;51(1):1-12.