1.a. Full Title: Chronic kidney disease (CKD) incidence prediction

1.b. Abbreviated Title (Length 26 characters): CKD prediction

2. **Writing Group:**
   Writing group members: Robert Nelson, Josef Coresh, Shoshana Ballew, Yingying Sang, Kunihiro Matsushita, Varda Shalev, and others for the CKD Prognosis Consortium

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

**First author:** Josef Coresh, MD, PhD, MHS  
Address: 2024 E. Monument, Suite 2-630  
Baltimore, MD 21287  
Phone 410-955-0495, Fax 410-955-0476  
Email: coresh@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: Josef Coresh, MD, PhD, MHS  
Address: 2024 E. Monument, Suite 2-630  
Baltimore, MD 21287  
Phone 410-955-0495, Fax 410-955-0476  
Email: coresh@jhu.edu

3. **Timeline:**
Data to be used in this proposal are already available. Analyses and manuscript preparation will be performed over the next 6 months.

4. **Rationale:**
Chronic kidney disease (CKD) is a major global public health problem affecting 10 to 16% of the adult population worldwide.\(^1\)\(^,\)\(^2\) Previous CKD Prognosis Consortium (CKD-PC) work has shown that low eGFR and high albuminuria are strong independent predictors of risk for mortality and kidney outcomes.\(^3\)\(^-\)\(^7\) Recently, a few risk prediction models for end-stage renal disease (ESRD) have been published,\(^8\) and those models were
globally validated in a CKD-PC meta-analysis manuscript currently under review.\textsuperscript{9} While accurate prediction of ESRD is important to guide appropriate nephrology care, most people with CKD experience other complications such as cardiovascular disease or die before reaching ESRD.\textsuperscript{10} Unfortunately, appropriate measurement intervals to define progression to CKD are unclear in clinical guidelines and medical therapy directly targeting kidney function and damage is limited.\textsuperscript{11} Thus, it is important to identify individuals at high risk of developing CKD in order to try to prevent the progression to CKD through management of possible risk factors such as hypertension, diabetes, smoking, and obesity. Therefore, we will conduct meta-analysis with the CKD-PC as a platform, explore potent predictors of incident CKD, and try to develop global risk prediction tools of incident CKD. This analysis will inform physicians about which patients are at high risk and provide guidance on management of CKD risk factors.

5. **Main Hypothesis/Study Questions:**
To develop risk prediction models of incident CKD for individuals from all over the world.

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).**

**Population:** All ARIC participants with data on estimated glomerular filtration rate (eGFR) and albuminuria will be included.

**Exposure Variables from ARIC visit 4:**
- eGFR (serum creatinine). eGFR will be assessed by CKD-EPI epi equation.\textsuperscript{12}
- Albuminuria (urinary albumin-to-creatinine ratio). Albuminuria will be expressed as urinary albumin-to-creatinine ratio (ACR).
- Demographics: Age, sex, race, socioeconomic status, geography
- Medical history/comorbidities: history of cardiovascular disease (myocardial infarction, bypass grafting, percutaneous coronary intervention, heart failure, stroke, or peripheral artery disease), hypercholesterolemia, hypertension, diabetes mellitus
- Laboratory variables: cholesterol levels (total, HDL, LDL), triglycerides, glucose levels with fasting status, smoking (current, former, never), hemoglobin A1c
- Vital measurements: systolic blood pressure, diastolic blood pressure, heart rate, anthropometry (BMI [height, weight], waist circumference, waist-hip ratio)
- Interfering medication: antihypertensive medications including ACE inhibitors /ARB, cholesterol-lowering medication (Statins), as well as glucose lowering medication.

**Outcome Variables:**
- Incident CKD during follow-up visits (CKD will be defined as newly identified eGFR (using serum creatinine or cystatin C\textsuperscript{13}) < 60 mil/min/1.73 m\textsuperscript{2} or urinary albumin-to-creatinine ratio (ACR) $\geq$ 30 mg/g.)
- Prevalent CKD (CKD will be defined as the presence of eGFR (using serum creatinine or cystatin C\textsuperscript{13}) < 60 mil/min/1.73 m\textsuperscript{2} or urinary albumin-to-creatinine ratio (ACR) $\geq$ 30 mg/g.): The primary outcome would be incident CKD but the secondary analysis of prevalent CKD would have implications on who to be tested for CKD measures.
Brief analysis plan and methods:
Various cohorts from North America, Europe, Asia, and Australia will be pooled on individual participant level. Variables with more than 50% missing values in the cohort will not be included in the analysis. All other missing data will be imputed using the multiple imputation technique. The complete case analysis will be conducted as a sensitivity analysis. The time horizons for risk prediction will be 2, 5, and 10 years. We will develop a sequential series of models and compare those with more variables (i.e., greater complexity) to simpler ones. We will mainly use clinical guidance to determine variable selection. In univariate Cox proportional hazards regression models, variables not associated with outcomes (P>0.1) will be excluded from further analyses. Improvement in model performance through addition of new candidate variables in multivariate Cox proportional hazards regression models will be tested using metrics described below. For the secondary analysis with prevalent CKD as the outcome variable, logistic regression models will be used.

Statistical metrics to evaluate model performance
i. **Discrimination:** Discrimination refers to the ability of a model to correctly distinguish between those with and without outcomes. Concordance statistics (C statistics) and integrated discrimination improvement will be computed as measures of discrimination.

ii. **Calibration:** Calibration describes how closely the predicted probabilities agree numerically with the observed outcomes. We will compare the observed vs. predicted risk of outcomes of interest for each quintile of predicted risk and determined the magnitude of the deviation using the Gronnesby and Borgan test.

iii. **Goodness of Fit:** Overall model fit for sequential models will be compared using the Akaike Information Criterion (AIC).

iv. **Reclassification:** Reclassification improvement will be quantified using the net reclassification improvement (NRI) statistic. To evaluate the effect of definition of risk categories on reclassification, we will calculate NRI using an alternative method that does not require categories (continuous NRI).

v. **Sensitivity and specificity:** For the secondary analysis with prevalent CKD, these indices will be also assessed.

Summary/conclusion:
By pooling various cohorts, from all over the world, on individual participant level; we will be able to develop strong risk prediction models. These results will serve as key work for future guidelines and patient care.

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.csc.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)?

  1362 - Chronic kidney disease and risk of end-stage renal disease: The Atherosclerosis Risk in Communities Study; Bash, LD.

  1574 - Comparison of novel markers of kidney function and prediction of cardiovascular events, mortality, and kidney failure: the Atherosclerosis Risk in Communities (ARIC) Study; Astor, BC.

  1581 - Novel markers of kidney function and prediction of incident chronic kidney disease and end-stage renal disease: the Atherosclerosis Risk in Communities (ARIC) Study; Astor, BC.

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.csc.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 for the Use of Linked ARIC CMS Data, approved manuscripts using linked ARIC 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

5. Gansevoort RT, Matsushita K, van der Velde M, Astor BC, Woodward M, Levey AS, Jong PE, Coresh J. Lower estimated gfr and higher albuminuria are associated with adverse kidney outcomes in both general and high-risk populations. A collaborative meta-analysis of general and high-risk population cohorts. Kidney Int. 2011;80:93-104


