ARIC Manuscript Proposal # 1194

1.a. Full Title: Effect of outcome misclassification on risk prediction

b. Abbreviated Title (Length 26 characters): Outcome PPV effect on AUC

2. Writing Group:
   Writing group members: Nina Paynter, Josef Coresh, Thomas Lumley

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

First author: Nina Paynter
Address: 2024 E. Monument St.
Suite 2-600
Baltimore, MD 21205
Phone: 410-502-6705 Fax: 410-955-0476
E-mail: npaynter@jhsph.edu

Corresponding/senior author (if different from first author correspondence will be sent to both the first author & the corresponding author): Josef Coresh
Address: 2024 E. Monument St.
Suite 2-600
Baltimore, MD 21205
Phone: 410-955-0495 Fax: 410-955-0476
E-mail: coresh@jhu.edu

3. Timeline: begin analysis upon approval, circulate draft in Nov 2006

4. Rationale:

The area under the receiver-operator curve (AUC) has been used to measure the ability to predict disease outcomes but the effect of outcome misclassification on the AUC has not been quantified. Outcome misclassification is frequent and can contribute to the imperfect prediction often observed. We have derived a formula for the effect of random misclassification in the outcome.

To complement the theoretical main component of the paper, we would like to present an example using actual collected data. We propose to use AIRC adjudicated CHD events...
(INC_BY02) as the gold standard and all possible CHD events identified for ARIC adjudication as the misclassified outcome to illustrate the effect of misclassification in reducing the AUC of a prediction model. We will compare the AUCs for the Cox-proportional hazards model of events using the traditional CHD risk factors for each outcome definition separately for men and women.

5. Main Hypothesis/Study Questions:

Our main study hypothesis is that the AUC will be larger when the gold-standard outcome is used as compared to the “misclassified” outcome of all identified potential CHD events.

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).

Individuals without information on smoking, blood pressure, hypertensive medications, diabetes and lipids for Visit 1 will be excluded from this analysis. Visit 1 will be used for all covariate and risk factor measurements and as the beginning of outcome follow-up time.

Exposures: smoking, blood pressure, hypertensive medication use, diabetes and lipids (total and HDL cholesterol)
Outcomes: Adjudicated incident CHD event, Potential incident CHD event
Covariates: age, race, gender, center
Analysis: 10 year AUC from Cox-proportional hazards models with bootstrapped confidence intervals.

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

8.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.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)?

None Found

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/

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.