ARIC Manuscript Proposal # 1646

PC Reviewed: 4/13/10  Status: A  Priority: 2
SC Reviewed: _________  Status: _____  Priority: _____

1.a. **Full Title**: Towards a novel cardiovascular risk score to reduce socioeconomic disparities in cardiovascular risk: social and biological approaches to improve calibration of Framingham scores

b. **Abbreviated Title (Length 26 characters)**: Risk prediction & disparities

2. **Writing Group**:
Writing group members: Adam Richards, MD MPH ; Kathryn Rose, MA MSPH PhD ; Jose Escarce, MD PhD ; Melonie Heron, PhD ; Kevin Fiscella, MD MPH ; Martin Shapiro, MD PhD ; Chloe Bird, PhD ; Teresa Seeman, PhD

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

**First author**: Adam Richards
Address:
Robert Wood Johnson Clinical Scholars Program
911 Broxton Ave, Los Angeles CA 90024

Phone: 310-794-2904  Fax: 310-794-3288
E-mail: adamrichards@mednet.ucla.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: Kathryn Rose
Address:
Department of Epidemiology
University of North Carolina at Chapel Hill
Bank of America Center
137 E Franklin St
Chapel Hill, North Carolina 27514
Phone: 919-966-4596  Fax: 919-966-9800
E-mail: kathryn_rose@unc.edu

3. **Timeline (2010)**:
**April, May:** Obtain Data for ARIC, CHS and NHANES III - NDI
June, July, August: Data Analysis – model development & validation
September, October, November: manuscript preparation & submission for peer review

4. Rationale:
Coronary heart disease (CHD) accounts for approximately 27% of all deaths in the United States, and clinicians recommend the initiation or intensification of preventive interventions based on estimated 10-year risk of cardiovascular heart disease (CVD) or death. Cardiovascular risk is negatively associated with socioeconomic status (SES) independent of traditional CHD risk factors but clinicians lack the tools to incorporate this SES gradient into their calculation of CHD risk. Current CHD risk calculators in the United States such as the ATP III and Framingham algorithms appear well calibrated to predict first coronary events among the overall United States population, but they under-estimate CHD risk among low SES groups. This systematic bias may lead clinicians to under-treat low SES groups and may inadvertently exacerbate SES disparities in serum cholesterol and CHD outcomes. A risk score that improves CHD risk prediction among low SES groups would provide clinicians with a simple tool to reduce CHD disparities by harnessing the power of their every-day treatment decisions.

The U.S. Preventive Services Task Force (USPSTF) recently outlined five major criteria to evaluate emerging risk factors for possible inclusion in Framingham risk scores to improve CHD risk prediction. The USPSTF applied these criteria in a comprehensive review of nine emerging CHD risk factors, and rejected all of them. However, their criteria did not evaluate potential bias among population subgroups, nor did they consider socioeconomic status as a possible risk factor. Fiscella et al added a composite SES outcome based on individual-level income and education to traditional Framingham risk scores (FRS), and their bi-variate model attenuated systematic SES bias among a representative middle-age population in the United States. Risk classification was upgraded for 15.1% of low-SES participants (95% CI 13.9-29.4%), although the authors did not report a risk score for clinical use. Investigators in England and Scotland have developed and validated entirely new risk scores (QRISK, QRISK2 and ASSIGN) that incorporate SES measures and improve model calibration among low SES groups, but a similar clinical risk score has not been developed for use in the United States.

Two approaches exist to improve risk score model calibration among low SES groups. In addition to directly incorporating SES measures into the risk model, similar improvements may also be achieved by the introduction of biological variables that are associated both with SES and with CHD outcomes. This project takes both the direct and biological approaches by evaluating how the inclusion of individual or neighborhood SES or CRP, separately and in concert, may minimize SES disparities in CHD risk prediction models.

We propose to develop a novel CHD risk score that improves calibration of traditional Framingham score and appropriately re-classifies low SES individuals as moderate- or high-risk of CHD events in the next ten years. We plan to evaluate three risk factors for possible inclusion in the risk score: C-reactive protein (CRP), individual SES and neighborhood SES. Although the USPSTF recently rejected CRP as a risk factor, CRP is negatively associated with SES and may attenuate the systematic bias of traditional FRS among low SES groups. Given that individuals
may under-estimate or refuse to report their individual income\textsuperscript{15} it is important to consider alternative means of acquiring SES-related information for the purposes of clinical decision-making. British clinicians use a postal-code-level “deprivation score” as a proxy for individual SES to improve CHD risk prediction\textsuperscript{11–13}, and a similar measure of neighborhood SES could easily be derived in United States populations by geocoding patient addresses available in electronic medical records and linking with census-tract SES data (David Bates, personal communication). We plan to evaluate an existing index of neighborhood SES, comprised of six publicly available variables from the United States Census, that is associated with allostatic load\textsuperscript{16}.

We propose to compare the observed 10-year CHD mortality risk to the risk predicted by the new score, as well as by the Framingham risk score. Sensitivity analyses will compare risk predicted by the Reynold’s and Karlamangla risk scores (for women and elderly, respectively).

This study is not designed to elucidate precise mechanisms underlying the etiology of CHD risk but to develop a practical score clinicians can use to modify treatment decisions. Specifically, we anticipate that the risk score could be used to inform recommendations for intensive lifestyle modification, and treatment with aspirin or a statin. Our analysis is premised on the appropriateness of current guidelines that recommend for or against treatment based on an individual’s predicted absolute 10-year risk of CHD. A recent study has strengthened the evidence in favor of multivariate-risk-model–based guidelines for statin treatment by demonstrating that a tailored strategy would prevent more CAD events while treating fewer persons with high-dose statins than current treat-to-target approaches (such as ATP-III).\textsuperscript{17} Our primary assumption, supported by multiple studies, is that pharmacologic intervention with aspirin or a statin provides a relative risk reduction that remains fairly constant across the etiology or the magnitude of absolute CHD risk\textsuperscript{18–21}.

We are unaware of any ongoing studies using ARIC, CHS or NHANES datasets to develop a novel risk score to improve calibration among low SES groups. As noted above, Fiscella et al have evaluated the addition of individual SES measures to traditional Framingham risk scores, and this group is also evaluating a simple measure of neighborhood SES based on census tract median household income (Kevin Fiscella, personal communication). Our proposed study builds on these analyses in three ways: first, we will refine and extend their findings among a cohort over 65 years of age; second, we will develop a novel risk score for clinical use; third, we will estimate the number of CHD events that potentially could be averted with implementation of the new risk score.

Current risk prediction models may exacerbate SES disparities by under-estimating CHD risk among low SES groups. A novel risk score clinicians can use to base treatment decisions has the potential to ameliorate SES disparities in CHD outcomes. As part of the analysis, we plan to quantify this possible benefit by estimating the number of CHD events that potentially could be averted with implementation of the new risk score. We plan to identify individuals participating in NHANES-III who would have been re-classified as eligible for statin treatment based on their 10-year risk of CHD, and to multiply the number of CHD deaths in this eligible group by the
relative risk reduction of statin therapy from previous primary prevention trials (25 or 35% for moderate- and high-intensity interventions, respectively). Our findings may support initiation of the first randomized trial of alternative risk scoring systems to reduce SES disparities in cardiovascular outcomes.

5. **Specific Aims and Hypotheses:**

**Aim 1:** To determine whether currently available risk scores, as well as prediction models based on their component predictor variables, under-estimate CHD risk among low-SES persons in the ARIC, CHS and NHANES cohorts.

**Hypothesis 1:** the Framingham risk score, as well as a model comprised of the six component predictor variables, will under-estimate CHD risk among low SES persons in all three cohorts.

**Aim 2:** To determine whether C-reactive protein, individual income and education, or neighborhood socioeconomic status, alone and in concert, improve calibration of CHD risk prediction models among middle-age and elderly populations in the United States?

**Hypothesis 2:** A clinically practical model that includes individual education and neighborhood SES will improve prediction of CHD events compared to a model that includes only the six Framingham predictor variables.

**Aim 3:** Develop (in ARIC and CHS) and validate (in NHANES) a novel coronary heart disease risk score that improves calibration and minimizes under-estimation of CHD risk among low socioeconomic populations.

**Hypothesis 3:** The final model developed in ARIC and CHS cohorts will also improve calibration of CHD among low SES populations in the NHANES cohort.

**Aim 4:** Estimate the number (proportions) of low SES individuals in the United States eligible for primary prevention therapy based on the new score, compared to the number eligible under current treatment guidelines (ATP-III or Framingham/NHLBI risk-based strategy)

**Hypothesis 4:** The proportion of low SES individuals eligible for statin therapy using the new risk score will be larger than under current guidelines. Conversely, the proportion of high SES individuals eligible for treatment may be lower.

**Aim 5:** Evaluate the impact of re-classification on the socioeconomic gradient (disparity) in CHD outcomes by estimating the number of CHD events that potentially could be averted with implementation of the new risk score Coronary heart disease

**Hypothesis 5:** The number deaths that potentially could be averted by implementation of the new score will be substantial.
6. Design and analysis

6A. Data Sources
We are in the process of applying for access to three datasets described below. The NHANES data request has been submitted to the NCHS. CHS and ARIC requests are drafted and are awaiting official sponsorship by the respective study investigator. The RAND IRB has approved the study protocol. (ARIC and CHS proposals require documentation of IRB approval prior to submission)

Atherosclerosis Risk in Communities (ARIC)
The ARIC Study is a prospective study of atherosclerosis and cardiovascular disease incidence in a cohort of 15,792 persons, initially aged 45 to 64 years and sampled from 4 US communities in 1987–1989. A baseline examination and 3 subsequent triennial examinations were conducted. Follow-up is ongoing and complete for incident CHD (ie, myocardial infarction, fatal CHD, or coronary revascularization). The Coordinating Center and Publications & Presentations (P&P) Committees for ARIC and CHS review proposals and prepare datasets for each project using CHS and ARIC data. The National Heart Lung Blood Institute also review all proposals to ensure proper adherence to confidentiality procedures. Because data on CRP levels at the first visit is not available for all ARIC participants, we will defer the CRP-related analyses in ARIC.

The Cardiovascular Health Study (CHS)
The CHS is a population based longitudinal study of coronary heart disease and stroke in adults 65 years of age and older. A total of 5201 men and women (242 black participants and 4959 white participants) were recruited from four areas: Forsyth Co, NC; Washington Co, MD; Sacramento Co, CA; and Pittsburgh, PA. Participants were sampled from Medicare eligibility lists in each area. Eligible participants were 65 years or older at the time of the examination (June 1989–May 1990), not institutionalised, and did not require a proxy respondent at baseline. Overall, 57% of those eligible were enrolled in the study. In 1992–93, three of the four centers (Forsyth Co, Sacramento, and Pittsburgh) recruited an additional 687 black participants into the CHS cohort using the same sampling frame and data collection procedures used for the original cohort. The baseline interview and examination took place in 1989–1990 for the initial cohort and in 1992–93 for the supplemental cohort. Both cohorts have been followed up since then with semi-annual contacts. We plan to combine the two cohorts for these analyses.

NHANES III-NDI-Census linked data
We plan to access data through the Remote Data Center of the The National Center for Health Statistics (NCHS). SAS program requests are submitted to the NCHS, which returns tabulated results to the remote analyst’s secure email server.
This data will include (1) individual-level data from the geocoded Third National Health and Nutrition Examination Survey, 1988-1994 (NHANES III); (2) Census data characterizing neighborhood socioeconomic status (NSES) for each individual NHANES respondent, and (3) individual-level data from the National Death Index linkage file. NHANES data include interview, clinical examination, and laboratory data for a representative sample of U.S. residents. Blacks
and Mexican-Americans were over-sampled. Respondents’ residential addresses were geocoded to 1990 census tracts, using either address or closest street intersection. Fourteen percent of the sample (predominantly rural residents) could not be geocoded to a census tract and will be excluded. We plan to use census tracts to represent neighborhoods as described by Bird et al.\textsuperscript{16} Neighborhood characteristics were derived from the 1990 decennial Censuses and linked to NHANES III respondents through census-tract identifiers based on 1990 census-tract boundary definitions. At no time will we have access to the census-tract identifiers of individuals; rather, we will provide the NCHS with a neighborhood SES index for all census tracts in the United States in 1990, which the NCHS will then link with each participant in the publicly available dataset. NDI-linked cause of death (cardiovascular; cardiovascular + stroke; or other) and duration of follow-up (years / months) will be requested through the NCHS.

Neither the NCHS nor the NHLBI (which coordinates CHS and ARIC data) releases data containing personal identifiers. If the possibility exists to identify individuals based on small number of participants in a given "cell" of data, that data is omitted from the output.

6B. Study variables

Outcomes
The primary outcome to be studied will be total incident CHD, which will include first myocardial infarction (including silent), and CHD death. Since the primary objective of this project is the development of a risk score for use with the lipid management guidelines from the NCEP and aspirin guidelines from the US Preventive Services Task Force, and since both these guidelines use incident CHD risk to guide selection of prevention strategies, our primary analyses will focus on CHD events. However, given documented disparities in access to revascularization procedures in the ARIC cohort we plan to exclude CABG and PCI from the primary outcome. Since aspirin and statins have a role in primary prevention of all cardiovascular disease, we will also examine a more comprehensive outcome, global incident cardiovascular disease, which will also include incident congestive heart failure, first stroke (non-fatal and fatal), first transient ischemic attack, new claudication, and peripheral revascularization.

Outcome variables (ARIC and CHS):
- Cardiovascular death (fatal MI; sudden or nonsudden coronary death)

Outcomes for sensitivity analyses (ARIC and CHS):
- incident “total cardiovascular disease” (hard CHD + fatal and nonfatal stroke + transient ischemic attack (TIA) + peripheral vascular disease + congestive heart failure)

Outcomes in the NHANES III dataset are limited to deaths and are derived from the National Death Index: The primary outcome is coronary heart disease (CHD) death, defined by International Classification of Diseases, 10\textsuperscript{th} Revision using codes I20 to I25.

For sensitivity analyses in the NHANES III cohort we will also report results for secondary outcomes:
1) a composite outcome of cardiovascular death based on codes I11-I78 (excluding I33-I40 [carditis]). This definition includes stroke (I60-69) but does not include rheumatic heart disease (ICD 10 codes I00-I09), diseases of veins, lymphatics and lymph nodes not classified elsewhere (I80-I89); or other and unspecified disorders of the circulatory system (I95-I99).  
2) any “cardiovascular disease” (ICD 10 codes I00-I99)

Predictor Variables
Framingham risk factor variables (age, sex, Total cholesterol, HDL cholesterol, current smoker, (treated) HTN

Additional risk score variables: waist-hip ratio, LDL cholesterol, former smoker, black race (Karlamangla score); and CRP, family history of premature CHD (Reynold’s score in CHS and NHANES only)

Individual SES variables: education, income

Neighborhood SES variable:
Census tract level neighborhood SES (NSES) index, developed by Bird et al from six United States census variables: (1) percent of adults older than 25 with less than a high school education; (2) percent male unemployment; (3) percent of households with income below the poverty line; (4) percent of households receiving public assistance; (5) percent of female-headed households with children; and (6) median household income. The NSES index has a zero mean and standard deviation of one (Cronbach’s alpha = 0.93), and the index is positively associated with allostatic load independent of individual-level characteristics, including individual SES.  

We plan to provide the data coordination center of each respective study with a dataset that contains two variables: 1) 1990 census tract, and 2) 1990 census tract neighborhood SES index, a primary predictor in our analyses. The individual identifiers from the geocoded ARIC file will be used only to merge the dataset with the neighborhood SES index variable. The final data file would include 1) publicly available variables, 2) the NSES index, 3) dummy variable for census tract. A census tract dummy variable is necessary to adjust variance estimates for clustering of observations at the census-tract level.

Sensitivity analysis variables (ARIC & CHS)
See below for sensitivity analysis rationale. Incident statin treatment (eg CHS visit 2, visit 3, visit 4); incident indication for statin therapy (traditional FRS>10%; requires Total/HDL cholesterol, smoking status, treated hypertension at visit 2, visit 3, visit 4); prevalent or incident aspirin therapy (for CHD); insurance status; usual source of care.

Sensitivity analysis variables (NHANES)
NHANES did not include clinical follow-up for participants and it is not possible to evaluate confounding by non-adherence to therapy after the baseline visit. Therefore sensitivity analyses
will be limited to baseline cholesterol lowering therapy, history of diabetes (or HbA1c > 6.5 or current insulin therapy)

6C. Inclusion / exclusion criteria
Eligible participants: men and women age 45 years or older without prevalent atherosclerotic vascular disease (history of CABG, PCI, or stroke; prevalent angina; prevalent peripheral arterial disease (ankle-brachial index <0.9 or claudication by questionnaire,) or a history of transient ischemic attack (TIA).

Carotid intima media thickness (IMT) will not be used as an inclusion / exclusion criteria because the goal of this study is to develop a risk score for general clinical use, and IMT is not currently measured in standard clinical practice in the United States.

Of note, our primary analyses also excludes individuals with prevalent diabetes (by history, HbA1c >= 6.5%, fasting blood glucose >126mg/dL or random blood glucose >200), but we plan to include participants with diabetes in the analytic file in order to perform sensitivity analyses.

6D. Descriptive statistics
We will provide descriptive statistics on participants in the analysis with and without low SES, defined as individuals with less than twelve years education or income less than 200% of federal poverty level. We will also present observed crude and age-standardised cardiovascular death rates per 1000 person years by age, sex, ethnicity, CRP, individual and neighborhood SES.

Next we will estimate predicted 10-year probability of cardiovascular events using the Framingham risk score (FRS) (and in sensitivity analyses, the Reynolds risk score among women only, and the novel risk score for elderly populations developed by Karlamangla among participants age >65 years).

6E. Multivariate Model Development
The first multivariate proportional hazards model will include the six FRS predictor variables (and in sensitivity analyses, composite variables in Reynold’s and Karlamangla risk scores).

In the proportional hazards model, the equation for the hazard for individual i at time t can be expressed as:

\[ h_i(t) = \lambda_0(t) \exp^{g(x_i)} \]  

(equation 1)

where \( \lambda_0(t) \) is the baseline hazard function that depends on time t and can be regarded as the hazard function for an individual whose covariates all have values of zero.

and \( g(x_i) = \{ \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_k x_{ik} \} \)  

(equation 2)
Variables \( \mathbf{X}_1, \ldots, \mathbf{X}_k \) represent predictor variables of the respective risk equation (e.g.: for Framingham these would include age, sex, hypertension, and HTN treatment, total cholesterol and HDL cholesterol, and current smoker).

The new risk score will evaluate three proportional hazards models that include, in addition to the original six FRS variables, the following three pairs of predictors: individual income and individual education; individual education and neighborhood SES; and (in CHS and NHANES) CRP and neighborhood SES. We will calculate robust variance estimates using the Huber Sandwich estimator to account for clustering of observations at the census tract level. We will then re-estimate predicted probabilities from each of these models to facilitate comparison with the distribution of probabilities predicted by FRS, and the distribution of observed events.

We plan to explore several means of incorporating individual income into the models: we will include poverty index (income poverty ratio) as a continuous variable; and we will examine alternative income thresholds from 100% to 400% of poverty to construct dichotomous and trichotomous poverty indicator variables. For participants lacking individual income data we will use an imputation algorithm previously developed by RAND specifically for the NHANES III database.

6F. Assessment of Model Performance
We will evaluate the effect of CRP and neighborhood SES on CHD risk prediction in four ways, as recommended by the USPSTF and the American Heart Association\(^2\):

1) we will report the magnitude and statistical significance of the multivariable survival analysis regression coefficients associated with CRP or SES

2) we will assess model discrimination by comparing the area under the receiver operator curve (AUROC) for each pair of models. We plan to perform additional evaluations of discrimination, following the strategy of Hippisley-Cox et al\(^1\): we will calculated the Brier score (a measure of goodness of fit where lower values indicate better accuracy) using the censoring adjusted version adapted for survival data, the D statistic (a measure of discrimination where higher values indicate better discrimination), and an \( R^2 \) statistic. The \( R^2 \) statistic is a measure of explained variation where higher values indicate more explained variation.

3) we will assess model calibration by calculating the Hosmer-Lemeshow goodness-of-fit test; and we will calculate the mean predicted and observed cardiovascular disease risk at 10 years and compare the observed CHD event rate (O) vs. predicted probability of CHD (P) for each tenth of FRS, in the entire cohort and among a sub-population of low SES individuals. Because the degree of calibration may depend on the choice of risk categories, we also will use a smoothing function to display predictions over the range of values

4) reclassification: we plan to evaluate the proportion of low- and intermediate-risk individuals who would be re-classified above/below 10% and 20% risk thresholds by the new risk score. We plan to estimate observed risk at 10 years using the 10 year Kaplan-Meier estimate. Calibration will also be assessed by calculating the net reclassification improvement (NRI) and integrated discrimination improvement (IDI) as per the methods
of Pencina and associates. The improvement in reclassification can be quantified as a sum of differences in the proportions of individuals moving into a higher risk group minus the proportion moving below a risk threshold for people who develop events, and the proportion of individuals moving below a risk threshold minus the proportion moving above a risk threshold for people who do not develop events. The IDI test measures the extent to which the use of a new risk marker correctly revises upward the predicted risk of individuals who experience an event and correctly revises downward the predicted risk of individuals who do not experience an event. NRI and IDI will be calculated for low SES participants, as well as for the entire sample population.

6G. Model Validation
We plan to evaluate the performance of the risk score (developed using ARIC/CHS) in the validation dataset (NHANES III-NDI). Performance will be assessed by evaluating discrimination, calibration and re-classification as described above.

2) If we are not granted access to NHANES III data then we perforce will develop and validate the score within ARIC. The approach to internal validation will take advantage of bootstrapping methods to estimate confidence limits of parameter estimates. In each bootstrap sample, we will quantify the discrimination ability of the new risk score using the c statistic, which has the interpretation as the area under the receiver operating curve (ROC). The distribution of the c statistic over a large number (~1000) of bootstrapped samples will be used to construct confidence intervals for the ROC area. We can take a similar approach to estimating confidence intervals for other parameter estimates. Adjustment will be made for optimism (resulting from internal validation instead of external or cross validation) using methods developed by Efron and Tibshirani and Harrell et al.

Estimating model impact on SES disparities
We plan to estimate the number of CHD events that potentially could be averted with implementation of the new risk score by identifying those individuals who would have been reclassified as eligible for statin treatment and multiplying the number of deaths in that group by the relative risk reduction of statin therapy from previous primary prevention trials (10, 25 or 35% for low-, moderate-, and high-intensity interventions, respectively).

6H. Sensitivity Analyses
We plan to conduct the following sensitivity analyses:

1) Statin therapy: low SES individuals may be less likely to receive a statin, likely due to lack of a usual source of care. In order to explore the possibility of confounding by treatment with a potential intervention drug class (statins), we will re-analyze the data after excluding participants taking a statin at the baseline visit. We also will evaluate the change in beta coefficients and other parameter estimates including / excluding participants with 1) interval statin treatment during follow-up, 2) no insurance and 3) no usual source of care.

2) We will re-calculate expected CHD risk using logistic regression models
3) We will estimate the proportion of the cohort recommended to initiate statin therapy under the new risk score, compared to the proportion under the current ATP-III guidelines.

4) Ethnicity and SES interaction. Because ethnicity is not included in our risk prediction models (except in Karlamangla score among the elderly) we cannot test directly for interaction; however, we will construct separate models for Whites and African Americans to evaluate the possible impact of previously reported ethnic variation in SES/CRP gradients.14, 27

5) Diabetes is considered a “CHD risk equivalent” under current guidelines, which recommend that diabetic patients receive an aspirin and statin unless contraindicated. Future guidelines may change, however, particularly among the elderly with late onset diabetes, for whom diabetes may not represent an elevation of CHD risk equivalent to prevalent CHD. We therefore plan to perform sensitivity analyses of a model that includes an indicator variable for prevalent diabetes, in a sample that includes participants with prevalent diabetes.

6) Evidence suggests that socioeconomic gradients in biological risk and mortality may be greater among young- and middle-aged populations than among elderly populations due to the premature deaths of high risk individuals at an earlier age.28 We therefore will examine possible interaction effects of age and SES; and consider stratified analyses by age (+/- 65 years old), as appropriate.

7) We will re-evaluate discrimination and calibration in models that include the alternative CVD outcomes as described above (eg: CHD + stroke and ‘global cardiovascular risk’).

6I. Limitations
There are several important limitations to this study. First, the cardiovascular outcomes in the validation dataset (NHANES III-NDI) are limited to cardiovascular mortality, and do not include other incident cardiovascular events (such as non-fatal myocardial infarction) that are included in the ARIC and CHS data. Furthermore, death certificate data (NHANES III-NDI) are notoriously inferior to adjudicated outcomes, and may lead to mis-classification of primary and secondary outcomes. This miss-classification would tend to diminish our ability to validate the risk score in the NHANES-III.

The second important limitation of this study is lack of inclusion of information on socioeconomic status across the life course. Evidence suggests that low socioeconomic status in childhood,29 and possibly in utero,30 may influence risk of cardiovascular disease later in life. The absence of maternal or childhood SES information in the CHS and NHANES III datasets precludes the assessment of life course SES in our analyses. However, the available current information at both the individual level (CRP and individual income) and neighborhood level (NSES) is likely to capture most of the variation in cardiovascular outcomes attributable to SES across the lifespan. Furthermore, the objective of this project is to develop a tool for clinical use, and we are unaware of validated retrospective measures of early life SES that might be routinely asked in clinical settings. Future studies may be able to further improve calibration of risk prediction among low SES individuals, possibly by including life-course SES.
The third important limitation is the lack of direct clinical trial evidence supporting one of our primary assumptions. As noted above, this risk score is based on the primary assumption that the relative risk reduction of statin therapy is independent of the magnitude and the etiology of cardiovascular risk. Specifically, we assume that statin therapy will reduce the risk of cardiovascular events due to factors related to CRP, and individual or neighborhood SES, despite incomplete understanding of the mechanism between the exposures and cardiovascular outcomes. It is not necessary to fully understand mechanisms or causal pathways in order to endorse therapeutic interventions. If we find evidence in favor of our hypotheses in three cohorts – that is, if current practice (that ignores SES in the prediction of cardiovascular risk) appears to exacerbate socioeconomic disparities in cardiovascular outcomes – a strong case can be made on ethical grounds to modify currently available tools. Clinicians may be faced with two alternatives to reduce the systematic under-treatment of low SES groups: either to improve risk prediction among low SES populations (as we would endorse); or, for those who demand clinical trial evidence, to conduct a randomized trial of statins prescribed differentially among low SES individuals.

Fourth, data on CRP is not available for all ARIC participants at the baseline visit and the relatively small number of outcome events occurring among participants with valid CRP measurements preclude calculation of stable parameter estimates in this cohort. Therefore the CRP-related analyses are limited to participants in NHANES, and elderly participants in CHS.

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 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?_____ 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.uc.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)?

As noted above, Fiscella et al have evaluated the addition of individual SES measures to traditional Framingham risk scores, and this group is also evaluating a simple measure of neighborhood SES based on census tract median household income. Kevin Fiscella agrees that the current study would extend the findings of their past work, and he has agreed to participate as part of this study team.

In addition, the ongoing study MS 1004 (ARIC CHD risk prediction from behavioral, psychosocial, and socioeconomic factors) covers similar territory. We contacted the lead investigator Dr. W. Chambless who does not perceive a conflict with our study. (personal communication).

11. a. Is this manuscript proposal associated with any ARIC ancillary studies or use any ancillary study data?  
___X___ Yes  ____ No

The geocoded information used to link the neighborhood SES index would come from ARIC AS 1998.02 "Life course SES, social context, and CVD (SESCVD)” (Heiss PI/Rose contact)

11.b. If yes, is the proposal  
___ A. primarily the result of an ancillary study (list number* __________)  
___X___ B. primarily based on ARIC data with ancillary data playing a minor role (usually control variables; list number(s)* ___ AS 1998.02 __________ __________)  

*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.

_________________________