10-Year CHD Risk Prediction Using CAC and Traditional Risk Factors: the Multi-Ethnic Study of Atherosclerosis

current prelim work: Robyn McClelland, Neal Jorgensen

University of Washington

October 2, 2013

Background/Goals

Primary Modeling Approach

Preliminary Results

Discussion Points

- Have median of 10 years of follow-up in MESA
- Existing risk scores poorly calibrated to MESA and do not incorporate CAC
- Our goal is to develop a risk score using traditional risk factors and CAC
- Statistically want to use techniques to avoid over-fitting and over-optimism in performance measures.

Risk Factors and Interactions

- Risk Factors: age, gender, race/ethnicity, HDL, total cholesterol, lipid lowering medication use (yes/no), SBP, anti-hypertensive medication use (yes/no), BMI, current smoking, family history of heart attack, diabetes, and CAC.
- interactions (pre-specified): age, gender, race and CAC with all other predictors; anti-hypertensive medications-by-SBP; and lipid lowering medications-by- total cholesterol.
- non-linear terms: $age^2, age^3, sbp^2, sbp^3$

Endpoint

- hard CHD: n=259 events
- hard CHD + revascularization (with angina): n=393 events

Primary Modeling Approach: General Strategy

- avoid any p-value based selection;
- find model that will produce good predictions on new data;
- summarize properties accurately;
- used penalized Cox regression models to avoid over-fitting ('shrinkage'models)
- exploit bias-variance tradeoff to improve predictions on new data

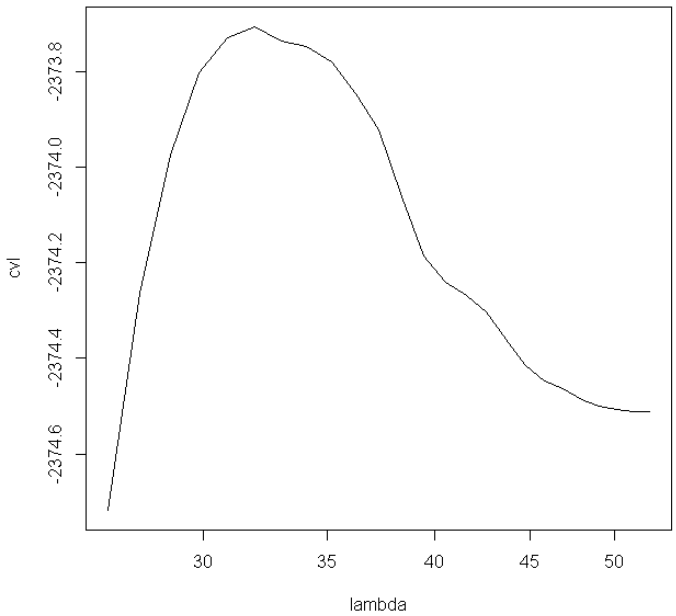
Primary Modeling Approach: Shrinkage Model

- These models 'shrink'coefficients towards zero relative to the MLE
- maximize (partial) likelihood subject to the constraints $\lambda_1 \sum |\beta_j| + \lambda_2 \sum \beta_j^2$
- $\lambda_1 = 0$ yields ridge regression (Hoerl and Kennard, 1970); λ_2 =0 gives the Lasso (Tibshirani, 1997); use of both is the elastic net (Zou and Hastie, 2005)
- use cross-validated log partial likelihood to select tuning parameter
- we used the `penalized` package in R

Primary Modeling Approach: Performance Metrics

- Discrimination: Do people who have the outcome have higher risk predictions than those who do not? *C-statistic; Discrimination Slope*
- Calibration: Do close to x out of 100 of people with a risk prediction of x% experience the event? *Brier Score, Calibration Slope, Hosmer-Lemeshow Test*

Results: First Stage Penalized Model



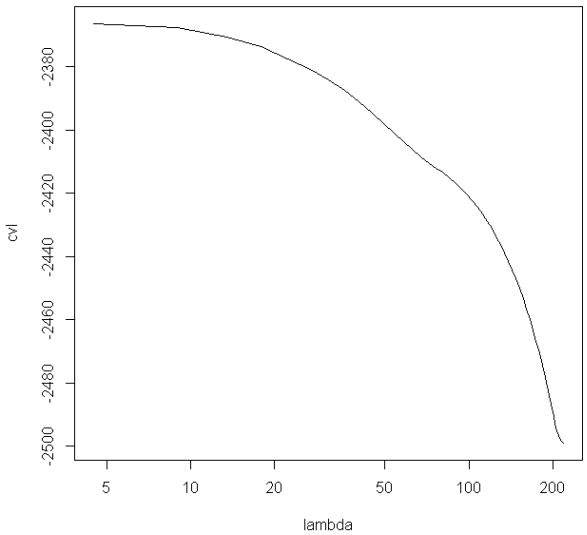
Main effects + Penalized interactions

Lasso (λ_1 = 32.02)

Main Effects	
age1c	1.0357
gender	1.4650
race	
White	--
Chinese	0.5731
black	1.0389
hisp	1.2082
lncac	1.2748
diabetes	1.5172
lipids	0.9500
htnmed1c	1.0895
smoking	2.0623
family hx	1.2877
hdl	0.9888
chol1	1.0030
sbp1c	1.0107

Interactions	
dm*age	0.9996
smoke*age	0.9828
hisp*dm	1.0055
sbp1c*lncac	0.9995

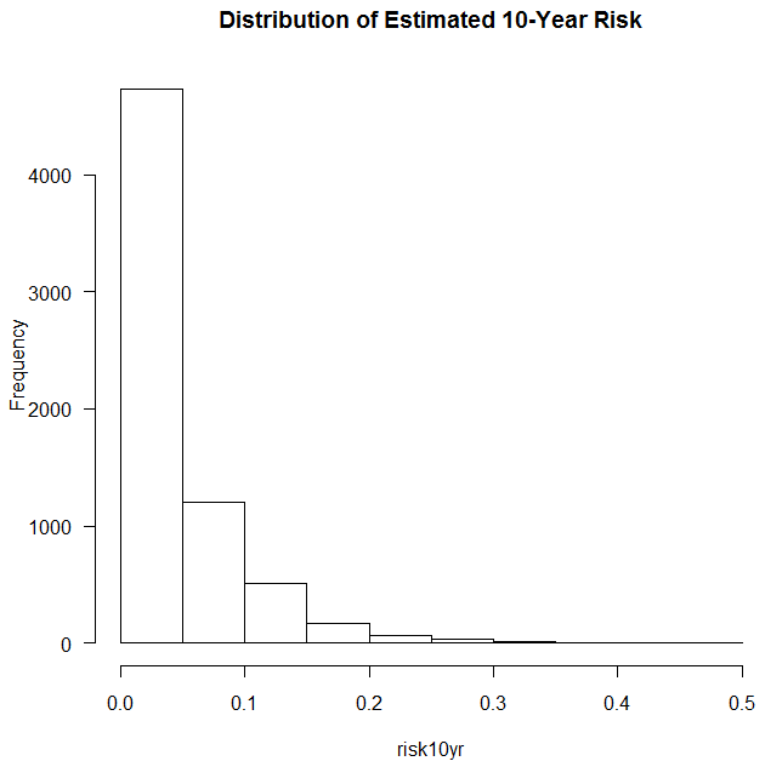
Results: Second Stage Penalized Model



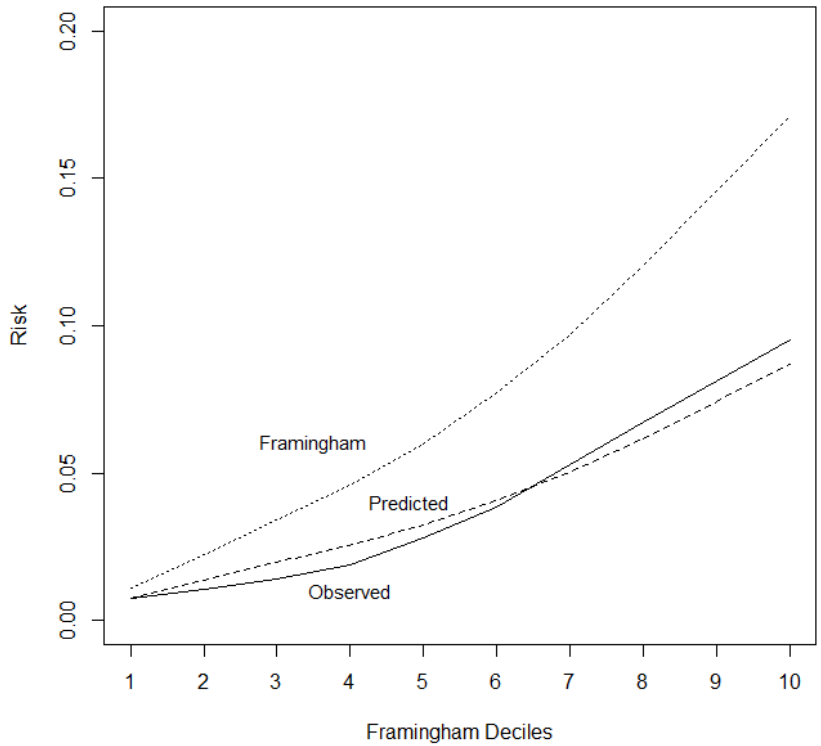
Penalized Main effects & interactions
Lasso (λ1 = 5.13)

Main Effects	
age1c	1.0356
gender	1.4235
race	
White	--
Chinese	0.9979
black	0.6239
hisp	1.0021
lncac	1.2826
diabetes	1.2593
lipids	0.9910
htnmed1c	1.0405
smoking	1.9507
family hx	1.2327
hdl	0.9899
chol1	1.0022
sbp1c	1.0142
Interactions	
dm*age	0.9989
smoke*age	0.9640
hisp*dm	1.6529
sbp1c*lncac	0.9982
Baseline 10-Year Risk	0.9826

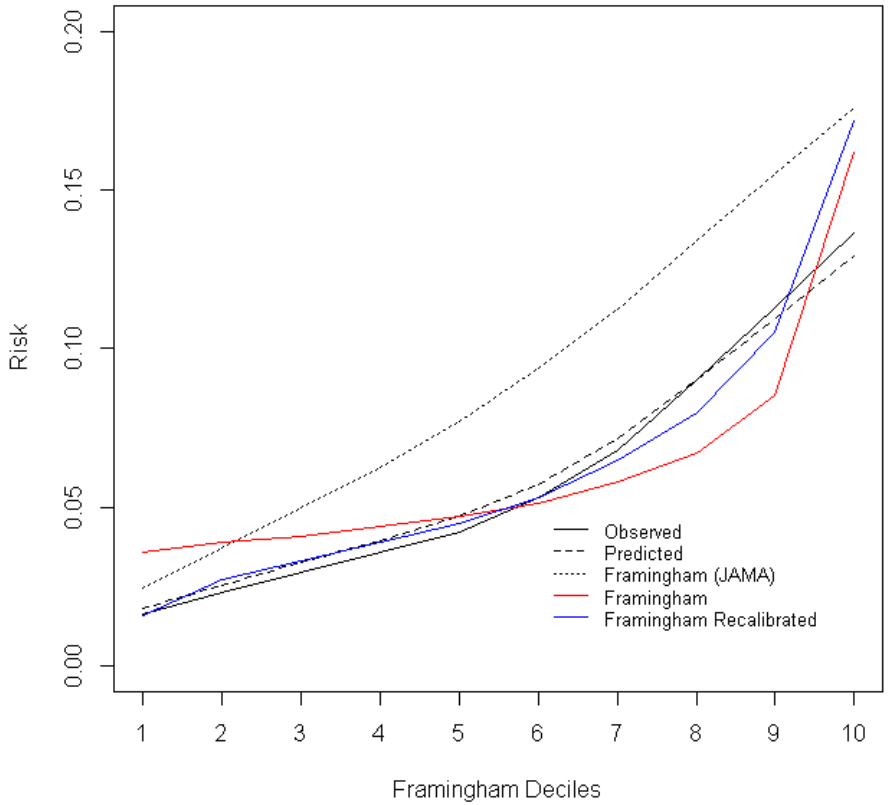
Results: Estimated 10-Year Risk



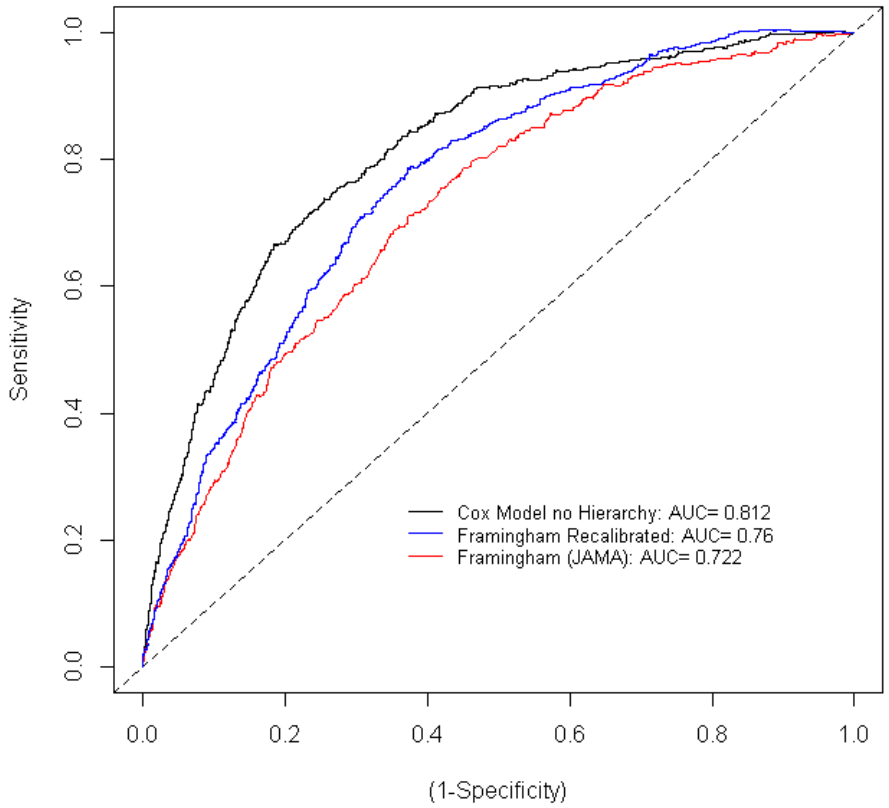
Results: Comparison with Framingham



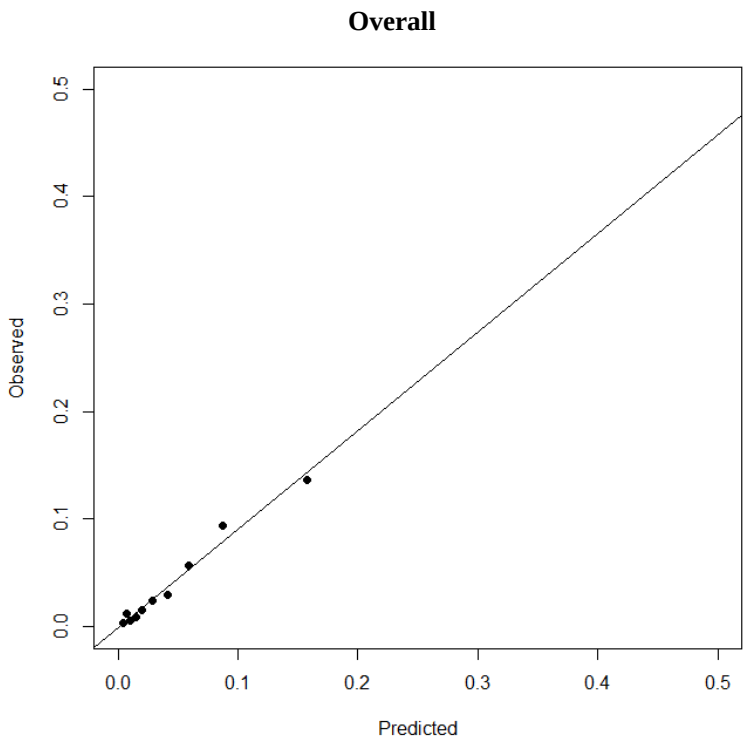
Results: Comparison with recalibrated FRS

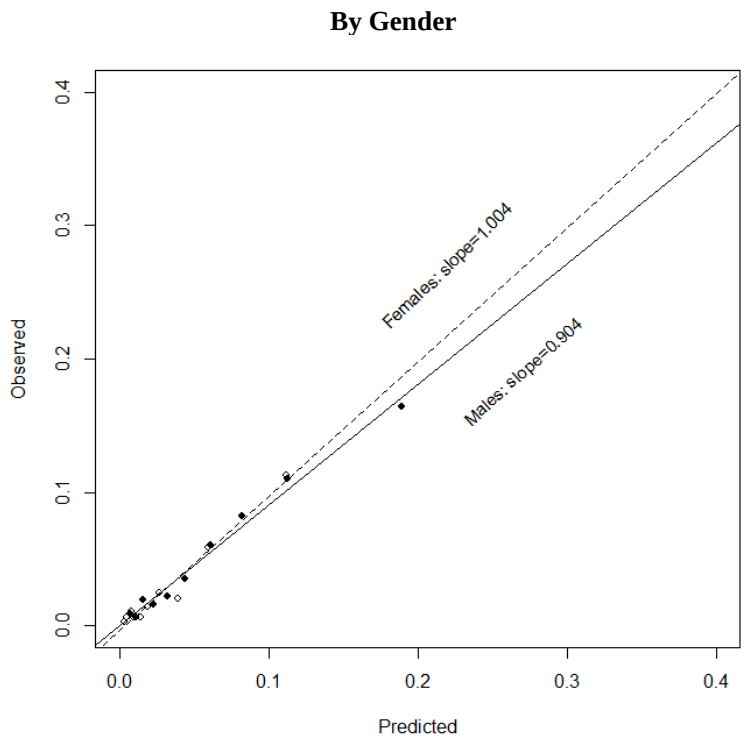


Results: ROC Curve

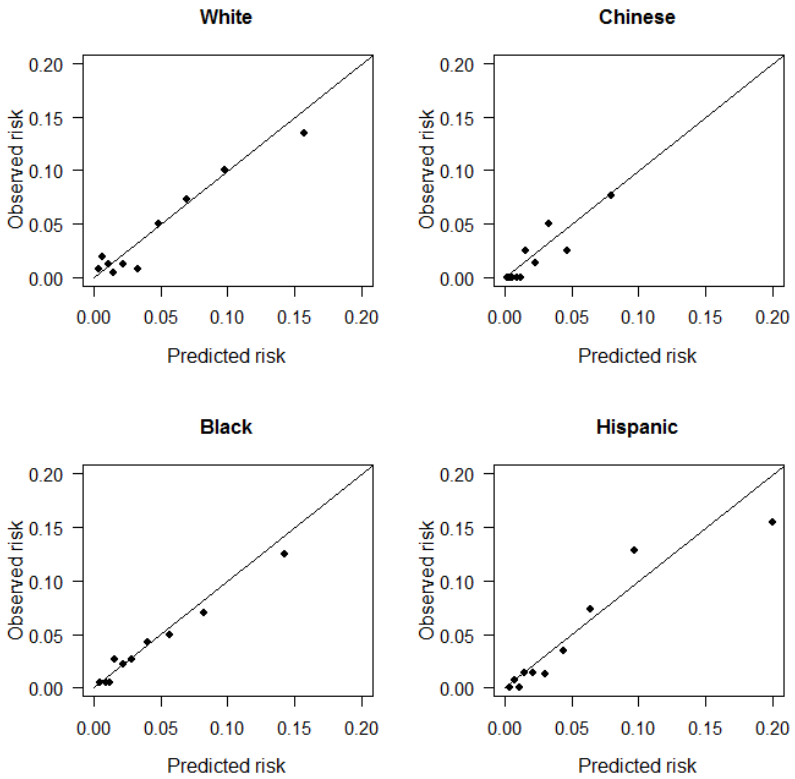


Results: Calibration





Results: Calibration by Race



Results: Performance Metrics

Metric	Apparent	Bootstrapped Optimism	Optimism Corrected
AUC	0.81	TBD	TBD
Brier Score	0.036	TBD	TBD
Discrimination Slope	0.055	TBD	TBD
Calibration Slope	0.91	TBD	TBD
Hosmer Lemeshow Test	p=0.184	TBD	TBD

- Take bootstrap sample, develop model on it.
- calculate performance metric on bootstrap (development) sample and on original
- difference in these average performances is "optimism".
- subtract optimism from Apparent to get optimism-corrected

Secondary/Sensitivity Modeling

- Bayesian model averaging
  - basic idea: averaging over all models provides better prediction than using a single model
  - uses a weighted average over (some sensible set of) models, weighted by the posterior probability for each model
- one-stage versus two-stage models (determines whether interactions are forced in or not)

Discussion Points

- Plan for dissemination includes a web tool (calculator).
- Model may be too complicated for simple points-based system–problem?
- Best model for *prediction* hard to interpret for *biology*.
- Problematic or good idea to call it "MESA CHD Risk Score"?

Thank you!