Published in final edited form as:

Stat Med. 2014 July 10; 33(15): 2577–2584. doi:10.1002/sim.6165.

Assessing the incremental predictive performance of novel biomarkers over standard predictors

Vanessa Xanthakis, PhD^{1,2,3}, Lisa M. Sullivan, PhD², Ramachandran S. Vasan, MD^{1,3}, Emelia J. Benjamin, MD, ScM^{1,3}, Joseph M. Massaro, PhD^{1,2,4}, Ralph B. D'Agostino Sr., PhD^{1,4}, and Michael J. Pencina, PhD^{1,2,4}

¹Framingham Heart Study, Framingham, MA

²Department of Biostatistics, Boston University School of Public Health, Boston, MA

³Section of Preventive Medicine and Epidemiology, Boston University School of Medicine, Boston, MA

⁴Department of Mathematics and Statistics, Boston University

Abstract

It is unclear to what extent the incremental predictive performance of a novel biomarker is impacted by the method used to control for standard predictors. We investigated whether adding a biomarker to a model with a published risk score overestimates its incremental performance as compared to adding it to a multivariable model with individual predictors (or a composite risk score estimated from the sample of interest), and to a null model. We used 1000 simulated datasets (with a range of risk factor distributions and event rates) to compare these methods, using the continuous Net Reclassification Index (NRI), the Integrated Discrimination Index (IDI), and change in the C-statistic as discrimination metrics. The new biomarker was added to a: **null** model; model including a **published risk score**; model including a **composite risk score** estimated from the sample of interest; and multivariable model with **individual predictors**. We observed a gradient in the incremental performance of the biomarker, with the null model resulting in the highest predictive performance of the biomarker and the model using individual predictors resulting in the lowest (mean increases in C-statistic between models without and with the biomarker: 0.261, 0.085, 0.030, and 0.031; NRI: 0.767, 0.621, 0.513, and 0.530; IDI: 0.153, 0.093, 0.053 and 0.057, respectively). These findings were supported by Framingham Study data predicting atrial fibrillation using novel biomarkers. We recommend that authors report the effect of a new biomarker after controlling for standard predictors modeled as individual variables.

Keywords

biomarkers; model discrimination; risk model; risk prediction

INTRODUCTION

With advances in technology and the availability of new prognostic markers of disease, cardiovascular disease (CVD) risk prediction models are constantly evaluated for improvements in prediction with the inclusion of new biomarkers. The critical underlying methodological and clinical question is whether a new biomarker provides a more accurate estimate of the absolute risk of CVD events compared to a set of standard predictors. It has been argued that simply relying on the statistical significance of the association between the new biomarker and CVD risk is insufficient to gauge its discrimination ability [1]. Statistical significance may or may not provide adequate evidence to support the inclusion of the biomarker in a prediction model with standard predictors as it does not necessarily imply that the inclusion of the biomarker improves the model's predictive accuracy [2, 3], nor may it indicate "clinical significance" of the new biomarker. Therefore, identification of new biomarkers that improve CVD risk prediction presents both challenges and opportunities for clinicians and statisticians interested in providing the best possible estimate of the absolute risk of developing a CVD event.

The area under the receiver operating characteristic curve (AUC,) is quantified by the C-statistic. Initially investigators assessed the incremental predictive performance of a new biomarker of CVD risk by comparing the C-statistic between models with and without the new biomarker [4]. The results of these initial investigations demonstrated that the increments in the C-statistic with the addition of new biomarkers were generally very modest unless the effect size for the biomarker was substantial [5]. The awareness of this limitation of the C-statistic prompted the development of additional complementary indices of discrimination performance, such as the Net Reclassification Improvement (NRI) and the Integrated Discrimination Improvement (IDI) [3], to assess improvement in model discrimination when evaluating the addition of a new biomarker to a set of existing predictors in a model.

Cardiovascular disease risk scores (e.g., the Framingham risk score or equivalent for coronary heart disease risk [6-9]) are often applied to populations other than the one from which they were derived. An important methodological question that often arises is whether to re-estimate the regression coefficients for the standard CVD predictors included in the published risk score using the current study data (derived from the population of interest) or to use the published regression coefficients. It is generally accepted that in this context it is more appropriate to estimate the regression coefficients for the standard predictors included in the risk score using the current study data [10]. Often, investigators evaluate the performance of a new biomarker by adding it to a model using a published risk score treated as a single variable. It is unclear to what extent the apparent discrimination ability of the new biomarker is influenced by the method used to model the standard predictors (i.e., applying published regression coefficients to the current sample vs. using regression coefficients estimated from the current sample). In the present investigation, we addressed this issue by comparing four methods for including a new biomarker in a prediction model to evaluate its incremental performance over standard predictors. The new biomarker was added to a: **null** model; model including a **published risk score**; model including a **composite risk score** estimated from the sample of interest; and multivariable model with

individual predictors. We focused on measures of improvement in discrimination with the addition of new biomarkers, including the continuous Net Reclassification Index [¹¹], the Integrated Discrimination Index, and changes in the C-statistic.

METHODS

We address the research questions in theoretical simulations and practical application to Framingham Heart Study (FHS) data.

Simulation study

Logistic regression analysis was used to model the association between standard predictors and a dichotomous outcome (e.g., presence/absence of CVD).

Our hypotheses were as follows: the incremental discriminatory ability of a novel biomarker will be gradually higher as the biomarker is added to the standard predictors using the different models below, with model 1) resulting in the highest increase and model 4) resulting in the lowest increase:

1. null model (unadjusted, method 4)
2. partially-adjusted model using a “published” risk score (method 3)
3. partially-adjusted model using a composite risk score created from data taken from the sample of interest (method 2)
4. fully-adjusted, refitting the model with current data (method 1)

We employed numerical simulations to investigate our hypotheses and to assess the true values of discrimination metrics used for testing the aforementioned hypotheses. Information from novel biomarkers is likely to be used clinically in a range of heterogeneous populations that may differ from the one in which the markers were initially measured. Therefore, we chose a 2-stage simulation design, shown in Table 1. Specifically, this simulation design captures the predictive performance of a novel biomarker W over a range of possible distributions of the standard predictors and of possible event rates. A detailed description of the simulation scheme is presented in Online Supplement A.

The parameter values were chosen to mimic those seen within the FHS in our previous work. More specifically, they resemble Odds Ratios (per SD of risk factor) similar to common predictors used for analyses within the FHS. It should be noted that the use of different parameter values could result in bigger differences in the discrimination metrics used in this study, depending on the sample size used and also on the number of parameters estimated. Online Supplement B shows a detailed example of a replicated dataset. In brief, for a simulated event rate of e.g., 15% using the first simulation scenario, one of the 1,000 generated vectors of $\mu_1, \mu_2, \mu_3, \mu_4$ could be 0.25, 0.45, 0.75, and 0.85, respectively; this generated values for X_1, X_2, X_3, X_4 , and W for those with events based on the distributions $N(0.25, 1^2), N(0.45, 1^2), N(0.75, 1^2), N(0.85, 1^2)$, and $N(1, 1^2)$, respectively, and based on $N(0, 1^2)$ for those without events. We also generated 1,000 “published” datasets to match the number of the current study datasets, each of sample size of $n=5000$. Finally, to establish the “typical reference” values of the discrimination metrics used, we generated a dataset of

n=1,000,000, with a 20% event rate intended to represent a hypothetical population (lower part of Table 1). We compared mean NRI, mean IDI, and differences in mean C-statistic, between sets of models with and without the biomarker, and also across methods used to model the biomarker with the standard predictors. We also calculated the difference in estimates (i.e., the difference between the discrimination values from each method and the values resulting from the hypothetical population) associated with the use of each of the three methods of incorporating a new biomarker (excluding the null model).

Clinical application - FHS example on atrial fibrillation

The design and selection criteria of the original FHS [12] and the Framingham Offspring Study [13] have been previously described. We considered a “current sample” that was comprised of second-generation FHS participants attending the sixth examination cycle (1995–1998), when circulating CRP and BNP were measured. Examination cycle 6 was different than the one on which the FHS AF risk score was developed. At the sixth examination cycle, there were 3120 attendees, of whom 203 (6.5%) developed AF within approximately 10 years of follow-up. In this sample, we assessed the predictive performance of separately adding CRP or BNP to the same predictors (i.e., age, sex, body mass index, systolic blood pressure, hypertension treatment, PR interval, presence of a heart murmur, and presence of heart failure) that are included in the AF risk score created by Schnabel et al. using FHS participants attending different examinations [14]. We performed the following comparisons of differences in C-statistic, and NRI and IDI generated from a Cox proportional hazards regression analysis to evaluate the impact of the addition of CRP or BNP to: individual standard predictors versus to a risk score from current data; a published standard risk score versus to a risk score from current data; and a published standard risk score versus to individual standard predictors. The SAS version 9.2. software (SAS, Cary, NC) was used for all analyses.

The study protocols for Offspring examinations were approved by the Institutional Review Board at the Boston University Medical Center, and all attendees at the examinations provided written informed consent. The authors had full access to and take full responsibility for the integrity of the data. All authors have read and agree to the manuscript as written.

Role of the Funding Source

The funding source had no role in the design, conduct or reporting of study results.

RESULTS

Simulations

Results from simulation scenario 2 (Table 1) are shown in Table 2, which displays the mean, median, and standard deviation of the C-statistic for models with and without the new biomarker W, the difference in the C-statistic, the NRI and IDI for each of the four methods. Method 1 resulted in a minimally larger (practically identical) mean increase in the C-statistic as compared to method 2 (0.031 versus 0.030, respectively). Of note, a larger mean increase in the C-statistic (0.085) was observed with method 3 as compared to methods 1

and 2 (Table 2). Finally, adding W to a null model resulted in the highest increase in the C-statistic among all methods (0.26, Table 2).

Comparing the mean NRI and IDI values, we observed the same trend as noted for the difference in the C-statistic (Table 2). Additionally, Table 3 compares the three methods (not including the null model) with respect to the difference in estimates they introduce when contrasting it to the discrimination values resulting from the hypothetical population. More specifically, the first column of data shows the increase in C-statistics, NRI, and IDI for the hypothetical population. The next 3 columns show the mean increase in C-statistic, as well as the mean NRI and IDI values for the four methods. The final 3 columns show the difference in estimates resulting from using each method when comparing it to the hypothetical population. We observed that Method 3 introduces a larger difference in estimates as compared to methods 1 and 2. Moreover, the difference in estimates in NRI and IDI values follow the same pattern. The Online Supplement C shows that the difference in estimates can be even larger when different simulation parameters are used (i.e., instead of using 0.5 as the standard deviation for the distribution used to generate the independent means to be used for the generation of the values of the predictors, we have used separately 0.3 and 0.7).

FHS atrial fibrillation example

Table 4 shows the descriptive characteristics of the current and published study data, as well as the regression coefficients for the standard predictors resulting from Cox proportional hazards regression analysis. Both CRP and BNP were natural-logarithmically transformed to normalize their distributions.

Adding CRP or BNP to a model that includes the single AF risk score estimated from the current study (method 2) resulted in a slightly smaller (and practically identical) increase in the C-statistic (0.0035 and 0.0231 for CRP and BNP, respectively) as compared to adding them to a multivariable model with the individual standard predictors for AF (0.0043 and 0.0243 for CRP and BNP, respectively), with the latter model showing a slightly greater discrimination ability (Table 5). The IDI showed a similar trend. The NRI results also followed a similar pattern when adding BNP; a somewhat larger improvement in the NRI was observed when adding CRP using method 2 versus method 1 (0.2148 versus 0.1690, respectively; Table 5), perhaps because the association between CRP and AF is not as strong as the association between BNP and AF.

Comparing the effect of standard predictors alone (before adding CRP or BNP), method 3 produced a lower C-statistic (0.7535) as compared to method 2 (0.7789), which shows that method 2 had higher discrimination ability as compared to method 3 before even considering the new biomarker (Table 5). Method 3 led to a larger increase in the C-statistic when adding CRP, as compared to method 2. A similar pattern was observed with the addition of BNP. The IDI and NRI values showed a similar trend (Table 5),

Method 3 resulted in a larger increase in the C-statistic compared to method 1 (0.0122 versus 0.0043, respectively for CRP and 0.0442 and 0.0243, respectively for BNP), with the NRI and IDI values showing a similar pattern.

DISCUSSION

Principal findings

The current investigation compared four methods of incorporating new biomarkers into existing CVD risk prediction models and investigated the effect of these methods to best assess the incremental predictive value of the new biomarkers on model performance. Simulation studies and evaluation of empirical FHS data (using the AF risk score as an example) yielded consistent results that suggested there is a gradient effect of adding a biomarker to standard predictors. More specifically, we observed the highest increase in the mean C-statistic, the NRI, and the IDI when adding the biomarker to a null model, and the lowest increase when adding it to a model with the predictors as individual variables. We also re-calibrated the published risk score for external validation, which produced similar results (data not shown).

Explanation for Findings

A potential explanation for the higher difference in discrimination ability of the model using a published risk score may be that it combines coefficients for the new biomarker estimated from current study data with published coefficients for the standard predictors; the published coefficients have been often validated and therefore the effect sizes are not inflated, yielding a smaller C-statistic before adding the new biomarker. It should be noted, however, that if the effect sizes were to be almost identical between the cohorts used to develop the published risk score versus the current model, the discrimination ability could come very close. Yet, this would not suggest that the use of the published score is the optimal method.

In general, the effect of better apparent performance of the biomarker when added to a model containing only the published risk score can be attributed to a poorer performance of this published score in the new sample under investigation. This can be due to a number of reasons, including:

1. Over-fitting – the published model was optimized for the sample on which it was developed and hence it does not perform as well on the new sample.
2. Difference in populations. Even though we assume that the sample used to develop the published score and the new sample on which we test the biomarker come from the same population, this assumption is likely true only approximately. Because the true regression coefficients of the predictors are likely not identical, the published risk score performs more poorly.

Another related explanation might focus on the fact that the new biomarker is optimized to the new sample. At the same time, the risk score as a whole and not the individual components are being fitted on the new sample, limiting the degree of optimization; If the regression coefficients were really close in the new sample as compared to the sample on which the published score was developed, one would expect that the incremental values of the new biomarker would also be very close. Finally, it is important to stress that re-fitting the entire score provides a more likely scenario from a practical standpoint: if a new biomarker was considered useful, it would be incorporated into the new risk model by re-fitting the model with the marker added to the list of predictors.

One point of attention could also be the sample size used for the study. If an adequate sample size is available, refitting would be the best option. However, with smaller sample sizes, this may be a challenge.¹⁵ Our observations suggest that careful attention should be given to the method used to model new CVD biomarkers to avoid potential overestimation of their incremental performance. In the present investigation we chose to focus on the best way to model a set of appropriate covariates for a specific outcome of interest. A broader question is the variation in the choice of covariates researchers choose to include in their model, as highlighted by Tzoulaki et al [16]; this important question is, however, beyond the scope of the present investigation. Hypothesis testing was not the focus of our investigation. However, if testing is desired, we recommend performing only one test: the standard likelihood ratio test (or its approximation, the Wald test).¹⁷

Conclusion

Overall, our observations indicate that the method used to control for standard predictors when assessing the impact of new biomarkers influences the apparent incremental performance. Specifically, adding a new biomarker to a model with a published risk score usually leads to greater NRI, IDI and increases in the C-statistic, and reliance on a published risk score might give an overly-optimistic view of the true predictive ability of the biomarker. This observation was likely due to the lower predictive ability, quantified by a lower C-statistic, of a model that includes a risk score (without the biomarker) based on published coefficients. Therefore, we suggest that the assessment of the incremental yield of a new CVD biomarker be performed by re-estimating the coefficients for the standard predictors using the current study data, as opposed to using a published risk score. Although we acknowledge that such refitting of models to individual study data may not always be possible in a research setting, our observations direct the attention of applied statisticians to the potential for overestimating the contribution of new biomarkers if the coefficients for the standard predictors are not re-estimated. It would be important for clinicians to work closely in consultation with statisticians to implement these best statistical practices.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by NIH contract N01-HC 25195, and NIH grants R01 HL 092577, R01 HL 102214 and RC1 HL 101056.

Reference List

1. Pencina MJ, D'Agostino RB, Vasan RS. Statistical methods for assessment of added usefulness of new biomarkers. *Clin Chem Lab Med*. 2010; 48(12):1703–1711. [PubMed: 20716010]
2. Hlatky MA, Greenland P, Arnett DK, Ballantyne CM, Criqui MH, Elkind MSV, Go AS, Harrell FE, Hong Y, Howard BV, Howard VJ, Hsue PY, Kramer CM, McConnell JP, Normand SL, O'Donnell CJ, Smith SC, Wilson PWF. on behalf of the American Heart Association Expert Panel on Subclinical Atherosclerotic Diseases and Emerging Risk Factors and the Stroke Council. Criteria for Evaluation of Novel Markers of Cardiovascular Risk. *Circulation*. 2009; 119(17):2408–2416. [PubMed: 19364974]

3. Pencina MJ, D'Agostino RB, D'Agostino RB, Vasan RS. Evaluating the added predictive ability of a new marker: From area under the ROC curve to reclassification and beyond. *Statistics in Medicine*. 2008; 27(2):157–172. [PubMed: 17569110]
4. 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):2109–2123. [PubMed: 15211606]
5. Pepe MS, Janes H, Longton G, Leisenring W, Newcomb P. Limitations of the Odds Ratio in Gauging the Performance of a Diagnostic, Prognostic, or Screening Marker. *American Journal of Epidemiology*. 2004; 159(9):882–890. [PubMed: 15105181]
6. Cook NR, Buring JE, Ridker PM. The effect of including C-reactive protein in cardiovascular risk prediction models for women. *Ann Intern Med*. 2006; 145(1):21–29. [PubMed: 16818925]
7. Ingelsson E, Schaefer EJ, Contois JH, McNamara JR, Sullivan L, Keyes MJ, Pencina MJ, Schoonmaker C, Wilson PWF, D'Agostino RB, Vasan RS. Clinical Utility of Different Lipid Measures for Prediction of Coronary Heart Disease in Men and Women. *JAMA: The Journal of the American Medical Association*. 2007; 298(7):776–785. [PubMed: 17699011]
8. Kim HC, Greenland P, Rossouw JE, Manson JE, Cochrane BB, Lasser NL, Limacher MC, Lloyd-Jones DM, Margolis KL, Robinson JG. Multimarker prediction of coronary heart disease risk: the Women's Health Initiative. *J Am Coll Cardiol*. 2010; 55(19):2080–2091. [PubMed: 20447530]
9. Ridker PM, Buring JE, Rifai N, Cook NR. Development and validation of improved algorithms for the assessment of global cardiovascular risk in women: the Reynolds Risk Score. *JAMA*. 2007; 297(6):611–619. [PubMed: 17299196]
10. Moons KGM, Kengne AP, Grobbee DE, Royston P, Vergouwe Y, Altman DG, Woodward M. Risk prediction models: II. External validation, model updating, and impact assessment. *Heart*. 2012
11. Pencina MJ, D'Agostino RB Sr, Steyerberg EW. Extensions of net reclassification improvement calculations to measure usefulness of new biomarkers. *Stat Med*. 2011; 30(1):11–21. [PubMed: 21204120]
12. Dawber TR, Meadors GF, Moore FE. Epidemiologic approaches to heart disease: the Framingham Study. *Am.J.Public Health*. 1951; 41:279–286.
13. Kannel WB, Feinleib M, McNamara PM, Garrison RJ, Castelli WP. An investigation of coronary heart disease in families. The Framingham offspring study. *Am.J.Epidemiol*. 1979; 110(3):281–290. [PubMed: 474565]
14. Schnabel RB, Sullivan LM, Levy D, Pencina MJ, Massaro JM, D'Agostino RB Sr, Newton-Cheh C, Yamamoto JF, Magnani JW, Tadros TM, Kannel WB, Wang TJ, Ellinor PT, Wolf PA, Vasan RS, Benjamin EJ. Development of a risk score for atrial fibrillation (Framingham Heart Study): a community-based cohort study. *The Lancet*. 2009; 373(9665):739–745.
15. Steyerberg EW, Borsboom GJ, van Houwelingen HC, Eijkemans MJ, Habbema JD. Validation and updating of predictive logistic regression models: a study on sample size and shrinkage. *Stat Med*. 2004; 23(16):2567–2586. [PubMed: 15287085]
16. Tzoulaki I, Liberopoulos G, Ioannidis JP. Assessment of claims of improved prediction beyond the Framingham risk score. *JAMA*. 2009; 302(21):2345–2352. [PubMed: 19952321]
17. Pepe MS, Kerr KF, Longton G, Wang Z. Testing for improvement in prediction model performance. *Stat Med*. 2013; 32(9):1467–1482. [PubMed: 23296397]

Table 1

Distributions Used for Data Generation

Generation of Means for X_1, X_2, X_3, X_4 ,* Among Events						
Scenario	Number of replications [†]	Distributions of $\mu_1, \mu_2, \mu_3, \mu_4$ for events				Ranges of event rates
		μ_1	μ_2	μ_3	μ_4	
1	1000	$N(0.3, 0.3^2)$	$N(0.5, 0.3^2)$	$N(0.7, 0.3^2)$	$N(0.9, 0.3^2)$	5–35%
2	1000	$N(0.3, 0.5^2)$	$N(0.5, 0.5^2)$	$N(0.7, 0.5^2)$	$N(0.9, 0.5^2)$	5–35%
3	1000	$N(0.3, 0.7^2)$	$N(0.5, 0.7^2)$	$N(0.7, 0.7^2)$	$N(0.9, 0.7^2)$	5–35%
4	1000	$N(0.3, 1^2)$	$N(0.5, 1^2)$	$N(0.7, 1^2)$	$N(0.9, 1^2)$	5–35%
Generation of X_1, X_2, X_3, X_4, W among events						
		X_1	X_2	X_3	X_4	W
		$N(\mu_1, 1^2)$	$N(\mu_2, 1^2)$	$N(\mu_3, 1^2)$	$N(\mu_4, 1^2)$	$N(1, 1^2)$
Generation of a hypothetical population ($n=1,000,000$)						
	Sample size	X_1	X_2	X_3	X_4	W
Events	200,000	$N(0.3, 1^2)$	$N(0.5, 1^2)$	$N(0.7, 1^2)$	$N(0.9, 1^2)$	$N(1, 1^2)$
Non-events	800,000	$N(0, 1^2)$	$N(0, 1^2)$	$N(0, 1^2)$	$N(0, 1^2)$	$N(0, 1^2)$

* Predictors X_1, X_2, X_3, X_4 and W for non-events followed a normal distribution with $N(0, 1^2)$.

[†] There were 4,000 replicated datasets (1,000 per scenario), each with sample size of $n=5,000$.

Table 2

Comparison Among the Four Methods – Simulation Results (Scenario 2)

	Mean	Standard deviation	Median
Method 1 (add W to X1, X2, X3, X4)			
C before adding W	0.841	0.069	0.851
C after adding W	0.873	0.044	0.874
Difference in C-statistic	0.031	0.028	0.022
NRI	0.530	0.129	0.531
IDI	0.057	0.034	0.051
Method 2 (add W to risk score from current study)			
C before adding W	0.841	0.069	0.851
C after adding W	0.871	0.046	0.873
Difference in C-statistic	0.030	0.026	0.021
NRI	0.513	0.120	0.515
IDI	0.053	0.031	0.047
Method 3 (add W to “published” risk score)			
C before adding W	0.730	0.108	0.738
C after adding W	0.815	0.047	0.806
Difference in C-statistic	0.085	0.113	0.075
NRI	0.621	0.120	0.628
IDI	0.093	0.043	0.089
Method 4 (add W to a model with only intercept, null model)			
C before adding W			
C after adding W	0.5	N/A	0.5
Difference in C-statistics	0.761	0.008	0.761
NRI	0.26	N/A	N/A
IDI	0.767	0.033	0.767
	0.153	0.022	0.155

Difference in estimates* resulting from the use of the three methods for estimating the predictive performance of a new biomarker

Table 3

	Hypothetical Population	Method 1	Method 2	Method 3	Difference in estimates from method 1	Difference in estimates from method 2	Difference in estimates from method 3
Difference in C-statistic	0.035	0.031	0.030	0.085	0.004	0.005	0.050
NRI	0.529	0.530	0.513	0.621	-0.001	0.016	0.091
IDI	0.060	0.057	0.053	0.093	0.003	0.007	0.032

* Difference in estimates is calculated by subtracting the discrimination metric for a given method from that estimated for the full population

Table 4

Descriptive Characteristics and Estimated Regression Coefficients – Current and Published* Study

Risk Factor	Descriptive characteristics		Estimated regression coefficients β (SE)	
	Current study n=3120	Published study n=4764	Current study n=3120	Published study n=4764
Age, years	58.4 (9.7)	60.9 (9.9)	0.0596 (0.0991)	0.1505 (0.0577)
Squared age			0.0003 (0.0008)	−0.0004(0.0004)
Male sex, %	46	45	0.4559 (1.1256)	1.9941(0.3933)
Body mass index, kg/m ²	27.9 (5.2)	26.3 (4.3)	0.0264 (0.0144)	0.0193 (0.0111)
Systolic blood pressure, mm Hg	128 (19)	136 (21)	0.0025 (0.0039)	0.0062 (0.0023)
Hypertension treatment, %	27	24	0.5109 (0.1506)	0.4241 (0.1010)
PR Interval, ms	163 (24)	164 (23)	0.0005 (0.0277)	0.0071 (0.0017)
Presence of heart murmur, %	3	3	5.1308 (2.4362)	3.7959 (1.3353)
Presence of congestive heart failure, %	0.5	1	0.2246 (5.0354)	9.4283 (2.2698)
Male sex * age			0.003 (0.0168)	−0.0003 (0.00008)
Heart murmur * age			−0.0697 (0.0362)	−0.0424 (0.019)
Congestive heart failure * age			0.0112 (0.0710)	−0.1231 (0.0335)

* Schnabel et al. *The Lancet* 2009;373:739-45.

Values are presented as mean (SD) or percentages

Table 5
Discrimination Measures for Three Methods of adding a novel biomarker for predicting AF risk

Model	CRP			BNP		
	Method 1 n=3120	Method 2 n=3120	Method 3 n=3120	Method 1 n=3120	Method 2 n=3120	Method 3 n=3120
C before adding biomarker	0.7789	0.7789	0.7535	0.7789	0.7789	0.7535
C after adding biomarker	0.7832	0.7826	0.7657	0.8032	0.8020	0.7977
Difference in C-statistic	0.0043	0.0035	0.0122	0.0243	0.0231	0.0442
P-value for difference	0.2184	0.2777	0.0225	0.0018	0.0015	0.00003
NRI	0.1690	0.2148	0.2572	0.4244	0.3581	0.4506
IDI	0.0025	0.0018	0.0035	0.0172	0.0161	0.0249