NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Prediction of Incident Heart Failure in General Practice: The ARIC Study

Sunil K. Agarwal, MD, MPH, PhD[1], **Lloyd E. Chambless, PhD**[1], **Christie M. Ballantyne, MD**[2], **Brad Astor, PhD**[3], **Alain G. Bertoni, MD, MPH**[4], **Patricia P. Chang, MD, MHS**[1], **Aaron R. Folsom, MD, MPH**[5], **Max He, PhD**[1], **Ron C. Hoogeveen, PhD**[2], **Hanyu Ni, MPH, PhD**[6], **Miguel Quibrera, PhD**[1], **Wayne D. Rosamond, PhD**[1], **Stuart D. Russell, MD**[3], **Eyal Shahar, MD, MPH**[7], and **Gerardo Heiss, MD, MSc, PhD**[1]

[1]Departments of Epidemiology and of Biostatistics, University of North Carolina, Chapel Hill, NC

[2]Baylor College of Medicine and Methodist DeBakey Heart and Vascular Center, Houston, TX

[3]Departments of Epidemiology and Medicine, Johns Hopkins University, Baltimore, MD

[4]Department of Epidemiology and Prevention, Wake Forest University School of Medicine, Winston-Salem, NC

[5]Division of Epidemiology and Community Health, University of Minnesota, Minneapolis, MN

[6]National Institute of Health, Bethesda, MD

[7]Division of Epidemiology and Biostatistics University of Arizona, Tucson, AZ

Abstract

Background—A simple and effective Heart Failure (HF) risk score would facilitate the primary prevention and early diagnosis of HF in general practice. We examined the external validity of existing HF risk scores, optimized a 10-year HF risk function, and examined the incremental value of several biomarkers, including NT-proBNP.

Methods and Results—During 15.5 years (210,102 person-years of follow-up), 1487 HF events were recorded among 13,555 members of the bi-ethnic Atherosclerosis Risk in Communities (ARIC) Study cohort. The area under curve (AUC) from the Framingham-published, Framingham-recalibrated, Health ABC HF recalibrated, and ARIC risk scores were 0.610, 0.762, 0.783, and 0.797, respectively. Upon addition of NT-pro-BNP, the optimism corrected AUC of the ARIC HF risk score increased from 0.773 (95% CI: 0.753 – 0.787) to 0.805 (95% CI: 0.792 – 0.820). Inclusion of NT-proBNP improved the overall classification of re-calibrated Framingham, re-calibrated Health ABC, and ARIC risk scores by 18%, 12%, and 13%, respectively. In contrast, Cystatin C or hs-CRP did not add towards incremental risk prediction.

Disclosures
None.

Conclusions—The ARIC HF risk score is more parsimonious yet performs slightly better than the extant risk scores in predicting 10-year risk of incident HF. The inclusion of NT-proBNP markedly improves HF risk prediction. A simplified risk score restricted to a patient's age, race, gender, and NT-proBNP performs comparably to the full score (AUC = 0.745), and is suitable for automated reporting from laboratory panels and electronic medical records.

Keywords

heart failure; risk prediction; external validation; NT-proBNP; Cystatin C; hs-CRP; biomarkers

Each year about 500,000 individuals are diagnosed with heart failure (HF) for the first time in the U.S.[1] In the setting of advances in clinical management, the incidence as well as the prevalence of HF have increased[2]. While progress in the therapy of HF appears to be associated with improved survival, greater efforts are needed towards early detection of ventricular dysfunction and the prevention of symptomatic heart failure[3].

Most HF patients present for the first time and are managed by general practitioners (GPs)[4]. However, both a poor appreciation of the importance of early diagnosis and treatment of LV dysfunction[5], and poor confidence in establishing HF diagnosis[6] may be important barriers to HF management. Appropriate risk stratification tools[7] can alert the clinician to patients at high risk of development of HF long-term, allowing for their risk factors to be aggressively managed while still primarily under primary care. A risk prediction score that is parsimonious, based on information easily available to the GP, and effective is required to implement the ACCF/AHA Guidelines for the Diagnosis and Management of Heart Failure in Adults (2009 update)[8].

We examined the external validity of the extant HF risk scores, i.e., the Framingham Heart Study [9] and Health ABC[10, 11] scores in the large, bi-racial cohort of middle aged participants sampled from four US communities by the Atherosclerosis Risk in Communities (ARIC) Study. We also derived a parsimonious HF risk function focused on primary care settings, called here the 'ARIC HF risk score,' and gauged its performance relative to the Framingham and Health ABC functions in predicting the 10-year risk of HF. Given the increasing use of biomarkers in clinical settings, we also examined the incremental value of few biomarkers, including N-terminal pro-B-type natriuretic peptide (NT-proBNP), for the long term risk prediction of HF.

Methods

Study population

The ARIC Study enrolled 15,792 men and women ages 45–64 years sampled from four U.S. communities[12]. Baseline examinations of the cohort were conducted from 1987 to 1989 to collect standardized information on socioeconomic indicators, medical history, family history, cardiovascular risk factors, serum chemistries, electrocardiograms (ECGs), and medication use. Three re-examinations, annual telephone interviews and active surveillance of hospitalizations and death followed the baseline visit. The last complete cohort visit was done in 1996–98.

Analysis involved use of covariates at two different index visits i.e., at the baseline examination (1987–89) and at visit 4 (1996–98). For analyses using baseline data, those with prevalent HF (n=775) or missing (n=325) data on HF[13]; missing information on any of the predictors shown in Table 1 (n=1502); or race other than black or white (n= 48) were excluded, thus leaving a cohort of 13,555 observations for analysis. After derivation of the ARIC HF model, we compared the beta estimates and discrimination statistics with those

from a dataset that excluded observations with missing information only for variables included in the final model (n = 279); all estimates were found to be quite similar.

Several biomarkers included in these analyses were assayed from stored specimens from this last visit (visit 4). From the 11,656 cohort members examined at the ARIC field centers during visit 4, those with prevalent HF at baseline visit (n= 469) and those with incident HF between baseline visit and visit 4 (n = 227) were excluded; those not self-identified as black or white (n=31), and those missing any important covariate (n = 187) or NT-proBNP (n=327) or NT-pro-BNP values   6025 pg/mL (n=6), were excluded, thus leaving a sample of 10,106 for the analyses using visit 4 as index visit.

Predictors of heart failure

Prevalent coronary heart disease (CHD) was ascertained from medical history as well as the adjudicated baseline ECG. Following 5 minutes of rest and while seated, blood pressure was measured three times using a random-zero sphygmomanometer, and an average of the last 2 readings was taken. Serum glucose was measured using the hexokinase method. Diabetes was defined by the presence of any serum glucose level >= 200 mg/dl, an 8-hour fasting glucose level >=126 mg/dl, a self-reported history of diabetes, or the current use of medications for diabetes. Cigarette smoking status was defined as self reported current smoking. The self reported average number of cigarettes/day and numbers of years of smoking were multiplied to derive cigarette-years of smoking, and this number was divided by 20 to get pack-years of smoking. COPD was defined as a self report of physician diagnosis of either emphysema or chronic bronchitis (or a chronic lung disease when using visit 4 as baseline). Body mass index was defined as the ratio of measured weight (in kilograms) and measured height^2 (in meters^2). Race and was identified at baseline as White, Black, American Indian or Alaskan Indian, or Asian or Pacific Islander.

An auscultatory finding of either a diastolic murmur or a systolic murmur of grade 4 or above by a trained physician assistant was considered as positive for presence of a valvular heart disease (VHD). A supine 12-lead ECG at rest was obtained using the MAC PC10 personal cardiogram (Marquette Electronics, Milwaukee, Wisconsin). Processing, monitoring, and quality control of the ECG data have been described elsewhere [14]. The presence of left ventricular hypertrophy was defined as a Cornell voltage >28 mm in men or >22 mm in women when using a 12 lead resting ECG. QRS duration (derived from 10 second resting ECG) of >120 ms as commonly used in clinical settings was used to define bundle branch block (BBB) was used as a binary variable.

In addition to N-terminal pro - B type natriuretic peptide (NT – pro BNP), two biomarkers putatively predictive of HF were considered. Using stored samples from visit 4 (1996–98), high-sensitive C-reactive protein (hs-CRP) using the immunoturbidimetric assay[15], cystatin C employing particle-enhanced immunonephelometry[16], and NT-proBNP using the Elecsys proBNP II immunoassay[17] were assayed.

Characterization of Heart Failure

Incident HF was defined as the first HF hospitalization or presence of HF code on death certificate since baseline visit through 2007. These events were identified from hospital discharge records and death certificates that showed a HF code in any position. International Classification of Diseases Code, Ninth Revision (ICD-9) code 428.x, and deaths with ICD-9/10 codes of either 428.x or I50 were considered as HF. The % agreement between HF events adjudicated by a standardized physician panel and ICD-9 code 428.x at any position was 73%, similar to that with the Framingham HF criteria (70%)[18].

Statistical Methods

The following analyses were done using SAS version 9.2 statistical software, SAS Institute, Cary, NC. Means (standard deviations) and proportions of characteristics at baseline were estimated by incident HF status. Cox proportional hazard models were used to estimate age, race, and gender independent hazard ratios for presence vs. absence (categorical variables), or per SD increment (for continuous variables). Those without incident HF were censored at death or at the end of study follow up. Model performance measures such as area under the ROC curve (AUC) was estimated for discrimination, and Gronnesby-Borgan (GB) statistics[19] for model fit. Also, Log likelihood ratio chi-square statistics were estimated by subtracting −2 log likelihood of model containing age, race, gender, and each variable from −2 log likelihood of a model with age, race, and gender only. Since net reclassification improvement (NRI) may be more meaningful for clinical decision making than the AUC[20], it was estimated for cut-offs for 10-year risk (<5%, 5 to <10%, 10 to <20%, and 20% or more)[20, 21]. All the variables presented in Table 1 with the exception of gender and race were retained when using likelihood ratio chi-square test with backward elimination exit criteria p value of 0.1. AUC, NRI, and IDI were estimated by methods which allow these to vary by time and which account for censoring [22]. The variables considered in optimizing an HF risk score in ARIC are shown in Appendix Table 1. To test the external validity of the extant risk functions in the ARIC study population, we estimated each participant's risk score by multiplying the published regression coefficients from the extant risk functions [9, 11] by the respective measurements for the ARIC cohort members. A Cox regression model was then fit with the risk function as the sole independent variable and performance statistics were estimated. Model fits for 5 and 10 years of follow-up yielded estimates that were quite similar. For the estimation of performance statistics of extant scores, we used both published regression coefficients[9–11], and regression coefficients derived within the ARIC cohort using the variables in the respective risk scores.

To derive the ARIC risk function we computed AUC for multiple models starting with variables that contributed most to AUC independently, while considering their easy availability to the primary care practitioners and measurement quality characteristics in practice settings.

To test the incremental value of NT-proBNP to the performance of the ARIC study risk score we considered NT-proBNP and its log transformation after exclusion of 6 observations with NT-pro-BNP values   6025 pg/mL, which upon examination were considered influential outliers. We evaluated models linear in NT-proBNP, polynomial in NT-proBNP, piecewise linear in NT-proBNP, categorical with 6 categories (percentiles 20, 40, 60, 80, 90) and categorical with 7 categories (20, 40, 60, 80, 90, and 95), linear in log (NT-proBNP), polynomial in log (NT-proBNP), and piecewise linear in log (NT-proBNP). The comparative fit of these models is summarized in Appendix Figures 1a and 1b, and the comparative index of discrimination in Appendix Table 2. Given its fair discrimination performance, model fit, and simplicity of use, log (NT-proBNP) was chosen for these analyses, though it is not the best fitting transformation at the lower tail of the distribution.

To obtain stable estimates for AUC corrected for optimism, due to fitting the risk score in the sample from which it was derived, 1000 bootstrap samples were processed to achieve stable estimates with lower bias than split-sample and cross-validation[23]. The average optimism i.e., (measure$_{\text{bootstrap samples}}$ − measure$_{\text{original dataset}}$) was subtracted from the original performance measure.

Results

Derivations of the ARIC HF risk score

Based on an average of 15.5 years of follow-up since baseline, 1487 (11%) incident HF events were observed in the 13,555 cohort members. The characteristics of study participants at baseline by incident HF status, as well as age, race, and gender adjusted hazard ratios, AUC, and GB statistics of models for each additional variable are presented in Table 1. Briefly, NT-proBNP, diabetes, blood pressure-lowering medication use and BMI each contributed more than 0.03 to the AUC of a model that included age, race, and gender only (AUC = 0.673).

Next we compared the AUC of several multivariable models to predict 10-year risk of HF, detailed in Appendix Table 1. The main criteria in the selection of candidate variables and model building were a) easy availability of the measurements to primary care physicians, b) measurement reliability, and c) parsimony relative to the degree of improvement in the AUC. Based on the above, the optimal model in the prediction of 10-year HF risk in the ARIC population included age, race, gender, prevalent CHD, systolic blood pressure, use of blood pressure-lowering medication, diabetes, smoking status, heart rate, and body mass index. After correction for optimism, the AUC achieved by this model was 0.7937 (95% CI = 0.7932, 0.7942). Appendix Figure 2 displays the predicted and observed events in the ARIC cohort from the ARIC risk function, by deciles of risk prediction score. While less than 5% of the cohort members below the 50th percentile had an HF event in 10 years, more than 30% of cohort members in the highest decile had an HF event. The ratio of predicted probability of those in the highest decile as compared to those in the lowest decile was 24.8, and the absolute 10 years risk difference was 31.2%.

The variables included in the HF risk score to predict 10 years of HF risk and their regression coefficients (log of hazard ratio of heart failure) are shown in Table 2 for all participants and by race and gender. Similarly, we have included yearly estimates from the baseline survival function in Supplemental Table 3. Modest variability in the magnitude and direction of associations can be seen by race and gender but there is limited statistical power to differentiate these associations in subsets of the cohort. The plots of 10-year risk of HF vs. percentile of risk shown in Figure 1 depict the overall goodness of fit achieved. We explored several fits for NT-proBNP on HF as an outcome. Although the fit was improved in piecewise models, the discrimination achieved was similar. Analysis repeating the piecewise linear fit (in log (proBNP)) after changing the lowest and highest knots to 10/90 percentiles of cases instead of 5/95 showed negligible differences in performance.

External validation of extant risk scores and comparison with the ARIC risk score model

Table 3 presents the AUC and model fit statistics for the ARIC, the Framingham and the Health ABC risk functions based on models fit to predict the 10-year risk of HF. The AUC for the Framingham and the Health ABC risk scores were estimated using the published beta estimates from the respective cohorts, as well as with beta coefficients estimated in the ARIC cohort using the variables of the respective risk scores, as described in the methods section. The AUC from the ARIC HF risk function was highest at 0.7966; the AUCs estimated in the ARIC cohort using variables from the Framingham HF risk score (0.7618), and using the variables in the Health ABC risk score (0.7835) were lower. The AUCs estimated using published coefficients from the Framingham and the Health ABC HF risk scores were 0.6139, and 0.7848, respectively. Reclassification using the ARIC HF risk score improved the overall classification of the individuals classified by the Framingham HF risk score: the net improvement score subtracting off the percentage of worsening classification from the percentage of improving classification was 13.5%. The gains in reclassification

using the ARIC risk score relative to that of the Health ABC risk score (which includes additional variables) were modest (NRI=3%). The overall goodness of fit by decile of risk score was good for the three risk functions, although the large number of events led to a sensitivity of the GB test.

Incremental value of biomarkers

With visit 4 as baseline, the average follow-up was 9.6 years, yielding 870 incident HF events in the 10,109 followed through 2007 (8.6%).

High sensitivity C-reactive protein (hs-CRP) did not contribute appreciably to the prediction of HF beyond that of a model with age, race, and gender only (AUC = 0.668 vs. 0.663), nor relative to the ARIC HF risk model (AUC = 0.772 without vs. 0.775 with hs-CRP) which yielded an NRI of 0.4%. Similarly, cystatin C did not improve the ability to predict the 10 year risk of HF compared to model with age, race, and gender only (AUC = 0.668 vs. 0.672), nor to the ARIC HF risk model (0.772 without vs. 0.779 cystatin C; NRI of 3.2%). These two biomarkers were not considered further.

Of all the variables shown in Table 1, NT-proBNP had the largest contribution to the AUC once added to a model based on the 1996–98 exam data containing age, gender, and race (data not shown). The AUC of the model that included age, race, gender, and NT-proBNP was 0.745. NT-proBNP was modeled using several distribution-based approaches to the treatment of NT-proBNP, as shown in Appendix Figures 1a and 1b; the resulting indices of discrimination and of fit of these models can be seen in Appendix Table 2. For ease of replication and use in practice, the log transformation of NT-proBNP was selected for further analyses.

The AUC of the ARIC HF risk score model (basic) compared to this model with the addition of NT-proBNP is shown in Appendix Figure 3. The AUC of the ARIC HF risk score model at visit 4 after addition of log(NT-proBNP) was 0.805 (95% CI: 0.792 – 0.820) as compared to 0.773 (95% CI: 0.753 – 0.787) of the risk model without this biomarker. The increment in the AUC is 0.032 is statistically significantly different from 0 (P<0.05). . The NRI following the addition of NT-proBNP was 13% (95% CI: 10.2, 19.9%) as shown in Table 4, with an integrated discrimination index (IDI) of 0.057 (95% CI: 0.043, 0.076). Inclusion of NT-proBNP in the Framingham HF risk score derived within the ARIC cohort achieved an AUC value of 0.792 (base model = 0.760), a NRI of 18% and IDI of 0.044. The corresponding AUC for the Health ABC risk score after addition of NT-proBNP was 0.799 (base model = 0.777), a NRI of 12% and an IDI of 0.036.

Considerable overlap in the AUC of models with NT-proBNP was observed when males (AUC 0.812; 0.795 – 0.833) were compared to females (0.799; 0.782 – 0.821). However, the NRI based on NT-proBNP was slightly higher in women = 17.9% (8.9, 23.3) compared to men 14.4% (8.2, 22.0).

Discussion

We optimized a parsimonious linear combination of variables to estimate the 10-year risk of incident HF in a population sample of men and women aged 45–64 years, based on information readily available in primary care settings. Compared to the extant HF risk scores, this ARIC HF risk function includes fewer variables and performs better than the Framingham abbreviated model and comparably to the Health ABC HF prediction model in terms of discrimination, using both traditional (AUC) and newer measures (NRI). Overall, the goodness of fit across deciles of risk was satisfactory. Addition of NT-pro BNP increased the AUC of the ARIC HF model by 0.03. Considering that the AUC of the ARIC

HF model was already high at 0.77 addition of this biomarker achieves a notable improvement relative to what is commonly seen in the literature after addition of multiple predictor variables.

Heart failure is a heterogeneous syndrome and often the culmination of prolonged and complex pathological processes [24]. In this respect, the robust discrimination ability for HF is remarkable, and comparable if not higher than the AUC statistics generally observed in CHD risk prediction, especially among women [25].

The Framingham HF risk score did not perform well in this sample, even when restricted to characteristics similar to the original derivation cohort (data not shown); in contrast, the discrimination achieved by the Health ABC risk score was good. The derivation cohorts for these risk scores are different, i.e., unselected middle aged participants in ARIC, a selected cohort of healthy elderly in the Health ABC, and a cohort with selective comorbidities in the Framingham HF risk score. The Health ABC risk score was validated in the CHS cohort[26]; although the Health ABC cohort includes African Americans, race was not included as a variable in the reported risk score.

The addition of NT-proBNP to risk models markedly improved 10-year HF risk prediction. Remarkably, the AUC of a model restricted to age, gender, race, and NT-proBNP is comparable to the AUC statistics achieved for most multivariable risk prediction equations. These results open the prospect of simple, automated estimates of the 10-year risk of HF to accompany the NT-proBNP values reported by clinical laboratories, similar to the current practice of automated reporting of an estimated glomerular filtration rate with measurements of serum creatinine.

While age and sex are almost invariably recorded and thus available, race may be missing in EMR and administrative claims data. We therefore estimated the impact of race on the performance of the full and the "simplified" ARIC models. In the 'simple' model (based only on age, sex, log NT-proBNP) the AUC was 0.7369, compared to an AUC of 0.7412 on addition of race (an increment of 0.0043). The risk in the highest decile - lowest decile for the former model was (28.5% – 1.5% = 27%), compared to (29.1% – 1.5% = 27.6%). Similarly, for a model with age, gender, diabetes, hypertension (clinically diagnosed), and current smoking the AUC was 0.7428, compared to the AUC of the same model following the addition of race 0.7437 (an increment of 0.0009).

Lastly, for the optimized ARIC risk score model with age, gender, diabetes, hypertension, systolic blood pressure, heart rate, body mass index, current and former smoking the AUC was 0.7734, compared to the AUC of 0.7733 for same model with addition of race (essentially unchanged).

Other measurements commonly available in clinical settings, such as ECG-defined left ventricular hypertrophy and the QRS interval, and serum markers such as creatinine, albumin, and HDL cholesterol, did not contribute importantly to the AUC for HF prediction, possibly reflecting saturation of the models, intermediate variable effects, modest collinearity with variables already in the model, or a low prevalence in this population-based cohort. Other proxy measures of fat mass such as waist circumference or waist to hip ratio may provide similar predictive value as BMI. These were not considered due to the widespread use of BMI in clinical settings and the higher measurement error associated with the other measures.

Convenient access to the predictor variables in the risk score, ease of use, parsimony in the number of predictor variables and the ability to modify the factors that adversely influence a patient's risk also seem influential for acceptability in clinical practice. A risk score with

these characteristics that can penetrate clinical practice can serve to identify patients at intermediate or high levels of predicted risk of HF who may benefit from education, aggressive risk factor management, regularly scheduled visits, and timely use of diagnostic testing or referral to a cardiologist, if felt necessary by the general practitioner. Current trends toward an increasing use of electronic medical records and availability of simple computational tools to practitioners may increase the penetrance and impact of risk estimation tools and research in this area will be needed. Prevention of the progression of Stage A HF to stages B or C, as highlighted in the guidelines recently released by several professional organizations[8, 27], will require an ability for simple and effective HF risk stratification in general practice. HF prevention trials that induct and monitor high risk individual also require such tools for initial screening and stratification of those eligible.

Among the strengths of this study is the derivation of a risk score in a large community-based cohort of middle aged white and black men and women, with long term follow-up and high retention rates. There are several limitations also worth noting. Although similar to that of the Health ABC study and other cardiovascular disease cohorts, the classification of HF as the outcome was not validated by review of all possible events. A recent validation study based on a large sample of hospitalizations discharged with ICD codes with high suspicion of HF indicates that the predictive value positive (PPV) of an ICD code 428.x for HF classified by a physician review panel was 77%, and the sensitivity was 95% [18, 28]. The corresponding PPV and sensitivity of Framingham Heart Failure criteria were 0.78 and 0.83, respectively[18, 28]. Thus, the HF outcome based on hospital discharge ICD codes used in our analyses may be considered at least as good as the Framingham HF diagnostic criteria. However, it should be noted that our study would miss HF cases managed successfully in outpatient settings if they were not hospitalized throughout the 15.5 year follow up period, although this would probably apply to rather small numbers. We must also note that little echocardiography data was available over the extended course of this cohort's follow-up, as a result of which we were unable to consider the potentially different predictors for systolic vs. diastolic HF.

We conclude that the ARIC HF risk score performs well in predicting 10-year risk of hospitalized HF in community settings. The inclusion of NT-proBNP in the ARIC HF risk score – and even in a model restricted to age, gender and race – markedly improves the prediction of 10-year risk of HF in middle-aged adults. These findings need replication, and possibly calibration in other cohorts. The risk calculators presented are available on www.ARICNEWS.net. Practice-based tools for early identification of susceptibility to HF and its efficient monitoring in community settings may contribute to the reduction of the growing burden of HF by enabling proactive risk management.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

1. Rosamond W, Flegal K, Furie K, Go A, Greenlund K, Haase N, Hailpern SM, Ho M, Howard V, Kissela B, Kittner S, Lloyd-Jones D, McDermott M, Meigs J, Moy C, Nichol G, O'Donnell C, Roger V, Sorlie P, Steinberger J, Thom T, Wilson M, Hong Y. Heart disease and stroke statistics--2008 update: A report from the american heart association statistics committee and stroke statistics subcommittee. Circulation. 2008; 117:e25–e146. [PubMed: 18086926]

2. Roger VL, Weston SA, Redfield MM, Hellermann-Homan JP, Killian J, Yawn BP, Jacobsen SJ. Trends in heart failure incidence and survival in a community-based population. Jama. 2004; 292:344–350. [PubMed: 15265849]

3. Fonseca C. Diagnosis of heart failure in primary care. Heart Fail Rev. 2006; 11:95–107. [PubMed: 16937029]

4. Fowler PB. Evidence-based diagnosis. J Eval Clin Pract. 1997; 3:153–159. [PubMed: 9276592]

5. Mair FS, Bundred PE. The diagnosis and management of heart failure: Gp opinions. 1996; 3:121–125.

6. Fuat A, Hungin AP, Murphy JJ. Barriers to accurate diagnosis and effective management of heart failure in primary care: Qualitative study. Bmj. 2003; 326:196. [PubMed: 12543836]

7. Kannel WB, D'Agostino RB, Sullivan L, Wilson PW. Concept and usefulness of cardiovascular risk profiles. Am Heart J. 2004; 148:16–26. [PubMed: 15215787]

8. Jessup M, Abraham WT, Casey DE, Feldman AM, Francis GS, Ganiats TG, Konstam MA, Mancini DM, Rahko PS, Silver MA, Stevenson LW, Yancy CW. 2009 focused update: Accf/aha guidelines for the diagnosis and management of heart failure in adults: A report of the american college of cardiology foundation/american heart association task force on practice guidelines: Developed in collaboration with the international society for heart and lung transplantation. Circulation. 2009; 119:1977–2016. [PubMed: 19324967]

9. Kannel WB, D'Agostino RB, Silbershatz H, Belanger AJ, Wilson PW, Levy D. Profile for estimating risk of heart failure. Arch Intern Med. 1999; 159:1197–1204. [PubMed: 10371227]

10. Butler J, Kalogeropoulos A, Georgiopoulou V, Belue R, Rodondi N, Garcia M, Bauer DC, Satterfield S, Smith AL, Vaccarino V, Newman AB, Harris TB, Wilson PW, Kritchevsky SB. Incident heart failure prediction in the elderly: The health abc heart failure score. Circ Heart Fail. 2008; 1:125–133. [PubMed: 19777072]

11. Butler J, Kalogeropoulos A, Georgiopoulou V, Belue R, Rodondi N, Garcia M, Bauer D, Satterfield S, Smith A, Vaccarino V. Incident heart failure prediction in the elderly: The health abc heart failure score. 2008

12. The ARIC Investigators. The atherosclerosis risk in communities (aric) study: Design and objectives. The aric investigators. American journal of epidemiology. 1989; 129:687–702. [PubMed: 2646917]

13. Loehr LR, Rosamond WD, Chang PP, Folsom AR, Chambless LE. Heart failure incidence and survival (from the atherosclerosis risk in communities study). Am J Cardiol. 2008; 101:1016–1022. [PubMed: 18359324]

14. Vitelli LL, Crow RS, Shahar E, Hutchinson RG, Rautaharju PM, Folsom AR. Electrocardiographic findings in a healthy biracial population. Atherosclerosis risk in communities (aric) study investigators. Am J Cardiol. 1998; 81:453–459. [PubMed: 9485136]

15. Folsom AR, Lutsey PL, Astor BC, Cushman M. C-reactive protein and venous thromboembolism. A prospective investigation in the aric cohort. Thromb Haemost. 2009; 102:615–619. [PubMed: 19806245]

16. Erlandsen EJ, Randers E, Kristensen JH. Evaluation of the dade behring n latex cystatin c assay on the dade behring nephelometer ii system. Scand J Clin Lab Invest. 1999; 59:1–8. [PubMed: 10206092]

17. Agarwal SK, Avery CL, Ballantyne CM, Catellier D, Nambi V, Saunders J, Sharrett AR, Coresh J, Heiss G, Hoogeveen RC. Sources of variability in measurements of cardiac troponin t in a community-based sample: The atherosclerosis risk in communities study. Clin Chem. 2011; 57:891–897. [PubMed: 21519038]

18. Rosamond WD, Chang PP, Baggett C, Johnson A, Bertoni AG, Shahar E, Deswal A, Heiss G, Chambless LE. Classification of heart failure in the atherosclerosis risk in communities (aric) study: A comparison of diagnostic criteria. Circ Heart Fail. 2012; 5:152–159. [PubMed: 22271752]

19. Gronnesby JK, Borgan O. A method for checking regression models in survival analysis based on the risk score. Lifetime Data Anal. 1996; 2:315–328. [PubMed: 9384628]

20. Pencina MJ, D'Agostino RBS, D' Agostino RBJ, Vasan RS. Evaluating the added predictive ability of a new marker: From area under the roc curve to reclassification and beyond. Stat Med. 2008; 27:157–172. [PubMed: 17569110]

21. Pepe MS, Feng Z, Huang Y, Longton G, Prentice R, Thompson IM, Zheng Y. Integrating the predictiveness of a marker with its performance as a classifier. Am J Epidemiol. 2008; 167:362–368. [PubMed: 17982157]

22. Chambless LE, Cummiskey CP, Cui G. Several methods to assess improvement in risk prediction models: Extension to survival analysis. Stat Med. 2011; 30:22–38. [PubMed: 20827726]

23. Steyerberg EW, Harrell FE Jr, Borsboom GJ, Eijkemans MJ, Vergouwe Y, Habbema JD. Internal validation of predictive models: Efficiency of some procedures for logistic regression analysis. J Clin Epidemiol. 2001; 54:774–781. [PubMed: 11470385]

24. Braunwald, E.; Zipes, DP.; Libby, P. Heart disease : A textbook of cardiovascular medicine. Philadelphia: Saunders; 2001.

25. D'Agostino RB Sr, Grundy S, Sullivan LM, Wilson P. Validation of the framingham coronary heart disease prediction scores: Results of a multiple ethnic groups investigation. Jama. 2001; 286:180–187. [PubMed: 11448281]

26. Kalogeropoulos A, Psaty BM, Vasan RS, Georgiopoulou V, Smith AL, Smith NL, Kritchevsky SB, Wilson PW, Newman AB, Harris TB, Butler J. Validation of the health abc heart failure model for incident heart failure risk prediction: The cardiovascular health study. Circ Heart Fail. 2010; 3:495–502. [PubMed: 20427700]

27. Lindenfeld J, Albert NM, Boehmer JP, Collins SP, Ezekowitz JA, Givertz MM, Katz SD, Klapholz M, Moser DK, Rogers JG, Starling RC, Stevenson WG, Tang WH, Teerlink JR, Walsh MN. Hfsa 2010 comprehensive heart failure practice guideline. J Card Fail. 2010; 16:e1–e194. [PubMed: 20610207]

28. Rosamond WD, Chang P, Baggett C, Bertoni A, Shahar E, Deswal A, Heiss G, Chambless L. Abstract 1454: Classification of heart failure in the atherosclerosis risk in communities (aric) study: A comparison with other diagnostic criteria. Circulation. 2009; 120:S506.

A Heart Failure (HF) risk score using few simple clinical variables was optimized to predict 10 years risk of new-onset HF was developed. It is based on 15.5 years of follow-up of the bi-racial ARIC cohort from four U.S. communities (1,487 HF events, 210,102 person-years). In African Americans and whites, a parsimonious HF risk score containing only age, gender, race, and NT-proBNP performed comparably to the extant HF risk scores, and also to those that incorporate a large number of risk factors. These findings carry potential impact for the prevention and early diagnosis of HF. Awareness of a patient's risk of HF may lead to opportunities for patient education, proactive risk factor management, timely use of diagnostic testing, or referral to a cardiologist, if felt necessary by the general practitioner. Using only information readily available to practitioners and their patients, the ARIC HF risk score improves risk stratification and adds to the practitioner's ability to prevent HF or to influence its natural history.

PLOT OF RISK VS PERCENTILE OF RISK
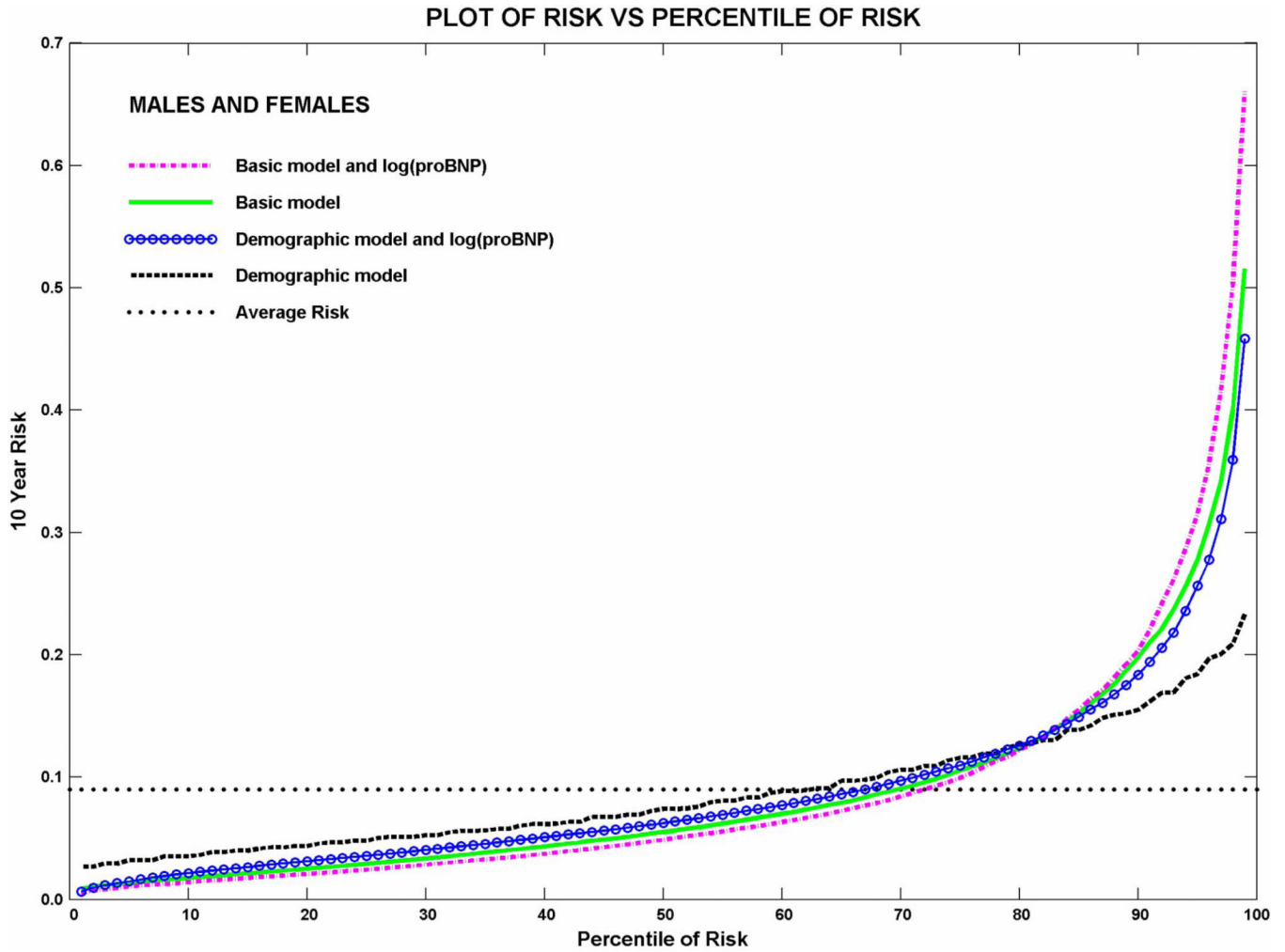
Figure 1.
Predicted 10-year risk of heart failure vs. percentile of risk for models parameterized as follows: **Demographic model** – Age, race and sex; **Demographic model + log(proBNP)** – Age, race, sex, log(NT-proBNP); **ARIC basic model** – age, race, sex, prevalent CHD, diabetes, systolic blood pressure, blood pressure medication use, heart rate, smoking status, and BMI; **ARIC basic model + log(proBNP)** – age, race, sex, prevalent CHD, diabetes, systolic blood pressure, blood pressure medication use, heart rate, smoking status, BMI and log(NT-proBNP).

Table 1

Baseline characteristics by incident heart failure status; age, race, and gender adjusted HR, AUC, and GB statistics: The ARIC cohort 1987–89 through 2005

Characteristics	HF during follow up* Mean (SD) or percentage		SD (all)	Hazard ratio‖			AUC	Incremental AUC	LR chi-square	Gronnesby Borgan§
	Yes (n =1487)	No (n=12068)		HR	95%CI					Wald Chi-Square
Age#(years)	56.6	53.8	5.8	1.69	1.60	1.78	0.638			6.73
Male**	0.54	0.44		1.43	1.29	1.58	0.653			2.12
Black**	0.34	0.25		1.78	1.60	1.98	0.660			18.33
Age, gender, race							0.673			24.18
Ln (NT-pro BNP) *** pg/mL	4.95	4.08		2.17	2.01	2.35	0.745	0.073	547.5	33.96
Diabetes	0.29	0.08		3.54	3.16	3.97	0.714	0.041	396.9	65.98
BP-lowering medication use	0.47	0.24		2.31	2.08	2.57	0.706	0.033	382.2	23.2
Body mass index (kg/(m*m))	29.6	27.3	5.2	1.45	1.38	1.52	0.704	0.031	190.4	16.55
Systolic blood pressure (mm of Hg)	130.0	119.8	18.7	1.42	1.36	1.48	0.701	0.028	255.4	19.14
Fasting glucose (mg/dL)	130.2	104.6	37.4	1.34	1.31	1.37	0.700	0.027	197.7	89.40
Pack-years of smoking	24.0	14.6	21.3	1.37	1.32	1.42	0.699	0.026	336.1	47.89
Prevalent CHD	0.13	0.03		4.12	3.53	4.82	0.698	0.025	202.8	26.16
HDL cholesterol (mg/dl)	47.2	52.7	17.0	0.67	0.62	0.71	0.698	0.025	227.3	14.79
Serum albumin (g/dL)	3.8	3.9	0.3	0.72	0.69	0.76	0.697	0.024	152.1	13.65
Current smoking status	0.37	0.24		2.05	1.84	2.28	0.696	0.023	140.9	24.54
Heart rate (beats/minute)	68.9	66.2	10.2	1.34	1.28	1.40	0.695	0.022	158.7	20.27
Log (serum creatinine in mg/dl)	0.11	0.10	0.24	1.18	1.12	1.24	0.682	0.009	139.7	6.31
Left ventricular hypertrophy	0.06	0.01		3.16	2.54	3.91	0.681	0.008	26.7	36.62
COPD	0.12	0.08		1.84	1.57	2.15	0.681	0.008	77.5	31.93
QRS duration >120 ms	0.07	0.03		2.20	1.81	2.68	0.679	0.006	47.0	21.79
Valvular heart disease	0.04	0.01		2.65	2.07	3.41	0.678	0.005	46.1	30.95
Cystatin-C*** mg/L	0.99	0.82		1.207	1.18	1.24	0.672	0.005	39.8	41.89
LDL cholesterol (mg/dl)	143.3	136.9	39.1	1.09	1.03	1.14	0.677	0.004	99.1	17.49

Characteristics	HF during follow up * Mean (SD) or percentage		SD (all)	Hazard ratio¶			AUC	Incremental AUC	LR chi-square	Gronnesby Borgan§
	Yes (n =1487)	**No (n=12068)**		**HR**	**95%CI**					**Wald Chi-Square**
Former smoking status	0.32	0.32		0.90	0.81	1.01	0.675	0.002	6.2	20.68
C–reactive protein *** mg/L	6.62	4.15		1.177	1.14	1.22	0.663	−0.005	0.6	36.43

SD = Standard Deviation, AUC = Area Under Curve of a receptor operating function, CHD = Coronary Heart Disease, COPD = Chronic Obstructive Pulmonary disease (self report of a physician diagnosis); LR = Likelihood ratio

*
Expressed as Mean and Standard deviation (SD) or Percentages

¶
Each row represents a model. For variables other than age, models includes age as a simultaneous independent variable; hazard ratios contrast the presence of a categorical characteristics versus its absence or 1 SD unit increase in the continuous variable, independent of age. For variables other than race, and gender, model contains age, race, and gender. Thus, the incremental AUC is as compared to a model with only age, race, and gender.

‖
AUC is calculated over 10 years of follow up (not corrected for optimism),

§
GB test has 9 degree of freedom corresponding to nine indicator variables to represent deciles.

#
Model has only age as an independent variable;

**
Model has age and the corresponding variable

***
Measured at visit 4 (1996–98) and the Incremental AUC is compared to a model with age, gender, and race, fitted with visit4 as baseline and estimates 10 years risk

Log likelihood ratio chi-square statistics calculated by subtracting −2 log likelihood of respective model from the −2 log likelihood of a model with age, race, and gender only (AIC = 27153.8)

Table 2

Regression coefficients (log of hazard ratio of heart failure) from multivariable models fit with the variables in the ARIC heart failure risk score, with and without NT-proBNP, for all participants and by race and gender

Variables in risk score	Full Cohort sans proBNP	Full Cohort +proBNP	Subgroup with variables included in the heart failure risk score				Full cohort with demographic + NT-proBNP
			White+proBNP		Black+proBNP		
			Female	*Male*	*Female*	*Male*	
Age (per year)	0.093	0.068	0.081	0.072	0.069	0.013	0.061
African American (yes vs. no)	0.019	0.286	-	-	-	-	0.702
Male (yes vs. no)	0.202	0.405	-	-	-	-	0.540
Heart rate (beats/minute)	0.026	0.026	0.031	0.023	0.021	0.031	-
Systolic blood pressure (per mm of Hg)	0.008	0.002	0.006	0.001	−0.003	0.005	-
BP-lowering medication use (yes vs. no)	0.338	0.229	0.270	0.194	0.081	0.326	-
Diabetes (yes vs. no)	0.672	0.763	0.795	0.620	1.042	0.830	-
Prevalent CHD (yes vs. no)	1.084	0.677	0.531	0.731	0.360	0.682	-
Current smoker vs. never smoker	0.976	0.840	1.055	0.971	0.241	0.882	-
Former smoker vs. never smoker	0.360	0.329	0.444	0.327	0.160	0.471	-
Body mass index (per kg/(m*m))	0.054	0.059	0.062	0.073	0.043	0.032	-
Log (NT-proBNP, pg/mL)	-	0.574	0.633	0.628	0.648	0.409	0.659

The ARIC study cohort's visit 4 (1996–98) was used to predict risk of HF through 2007

CHD = Coronary Heart Disease, NT-proBNP = Amino terminal pro-B type natriuretic peptide

The categorical variables such as African American, Male, BP-lowering medication use, diabetes, prevalent CHD, current smoker, former smoker were coded as 1 if condition was present and 0 if absent. For continuous variables the absolute value was used. All the variables were statistically significant (P<0.05) for prediction of HF.

Table 3

Discrimination of several models optimized to predict the 10-year risk of heart failure.

Sample		Optimism corrected Areas Under the ROC Curve						
		Framingham (V1)		Health-ABC (V1)		ARIC		
		From published beta	Derived beta	From published beta	Derived beta	Basic V1	Basic V4	Basic V4 with NT-proBNP
All		0.614	0.762	0.785	0.783	0.797	0.772	0.805
	Male	0.727	0.764	0.772	0.773	0.782	0.771	0.812
	Female	0.700	0.766	0.790	0.790	0.813	0.770	0.799

V1 and V4 refers to use of elements from the ARIC cohort's field center visits i.e., visit 1 (1987–89) and visit 4 (1996–98), respectively. V4 was used as baseline in estimations involving NT-proBNP. The number of participants when using V1 = 13555, and using V4 = 10103.

The variables included in the Framingham HF risk score[9] are age, gender, CHD, diabetes, ECG-based left ventricular hypertrophy, valve disease, heart rate, and systolic blood pressure. Health ABC score[10,11] includes the Framingham study variable with following differences: added serum albumin, serum creatinine, and smoking status; replaced glucose for diabetes; removed valve disease. The ARIC basic HF risk score includes age, race, gender, CHD, diabetes, systolic blood pressure, blood pressure medication use, heart rate, smoking status, and body mass index. AUCs are corrected for optimism.

The GB statistics for all the models presented above was <0.01.

Table 4

Net improvement on reclassification of individual risk of heart failure using the variables in the basic ARIC risk score and the addition of NT-proBNP

ARIC basic model	ARIC basic model with addition of NT-proBNP														
	10-Year Risk														
	<5%			5 to <10%			10 to <20%			20% or more			Overall		
10-Year Risk	n	row %	Risk	n	row %	Risk	n	row %	Risk	n	row %	Risk	n	%	Risk
<5%	4080	87.55	0.018	528	11.33	0.047	48	1.03	0.219	4	0.09	0.5	4660	46.12	0.023
5 to <10%	919	33.33	0.036	1360	49.33	0.062	433	15.71	0.144	45	1.63	0.362	2757	27.29	0.07
10 to <20%	134	7.9	0.074	528	31.11	0.081	744	43.84	0.146	291	17.15	0.287	1697	16.8	0.142
20% or more	8	0.81	0.25	44	4.45	0.165	231	23.36	0.184	706	71.39	0.424	989	9.79	0.35
Overall	5141	50.89	0.023	2460	24.35	0.065	1456	14.41	0.154	1046	10.35	0.383	10103	100	0.085
NRI	13.5%														

ARIC = Atherosclerosis Risk In Communities Study, NRI = Net Reclassification Improvement ; basic refers to model without biomarkers but only variables routinely and available to primary care physicians i.e., age, race, gender, blood pressure, heart rate, history of diabetes and CHD.